High-Throughput Genetic Analysis of Single Cells Using Microfluidically Generated Hydrogels

High-Throughput Genetic Analysis of Single Cells Using Microfluidically Generated Hydrogels

A dissertation presented
by

Mira T. Guo

to
The School of Engineering and Applied Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Applied Physics

Harvard University
Cambridge, Massachusetts

November 2017

Dissertation Advisor: Professor David A. Weitz Mira T. Guo

High-Throughput Genetic Analysis of Single Cells Using Microfluidically Generated Hydrogels

Abstract

Typical metagenomic datasets characterize populations as a whole by sequencing all the DNA fragments present in a sample, and are therefore unable to identify different genes as having originated from the same cell or distinct cells. We address this problem by microfluidically encapsulating single cells in hydrogels, amplifying the genomic DNA inside water-in-oil emulsion droplets, sorting out those cells carrying a gene of interest, and sequencing the gene at the single-cell level. Our method is broadly applicable to standard DNA amplification techniques, including polymerase chain reaction (PCR) and multiple displacement amplification (MDA). Since our method does not require cell culture, it is particularly useful for studying unculturable bacteria populations such as those found in the human gut and environmental water sources, and could be extended to determine which species or strains carry a rare metabolic gene.

Table of Contents

References to Previously Published Work

The following chapters are based largely on previously published work.

Chapter 1 **Guo, M.T.**, Rotem, A., Heyman, J.A. & Weitz, D.A. Droplet microfluidics for high-throughput biological assays. *Lab Chip* **12**, 2146-2155 (2012).

Chapter 2 **Guo, M.**, Mazutis, L., Agresti, J., Sommer, M., Dantas, G., Church, G., Turnbaugh, P. & Weitz, D. High-throughput single-cell PCR using microfluidic emulsions. *B Am Phys Soc* March Meet **57:2**, Abstract K1.00156 (2012).

Chapter 3 Tamminen, M.V. & Virta, M.P. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Front Microbiol* **6**, 195 (2015).

Chapter 4 Spencer, S.J., Tamminen, M.V., Preheim, S.P., **Guo, M.T.**, Briggs, A.W., Brito, I.L., Weitz, D.A., Pitkanen, L.K., Vigneault, F., Juhani, M.P., & Alm, E.J. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J* **10**, 427-436 (2016).

The following chapters describe work performed in collaboration with the respective individuals, in addition to the authors listed above.

Chapter 2 Corinne Maurice[1], Henry Haiser[1], Peter Turnbaugh[1,2], David Weitz[3]

Chapter 3 Manu Tamminen[4,5], Linas Mazutis[3], Eric Alm[4], David Weitz[3]

Affiliations [1]FAS Center for Systems Biology, Harvard University, 52 Oxford Street, Cambridge MA 02138, USA
[2]Department of Microbiology & Immunology, University of California San Francisco, 513 Parnassus Avenue, San Francisco CA 94143, USA
[3]School of Engineering & Applied Sciences, Harvard University, 29 Oxford Street, Cambridge MA 02138, USA
[4]Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA 02139, USA
[5]Department of Food and Environmental Sciences, University of Helsinki, P.O. Box 66, FI-00014 Helsinki, Finland

Acknowledgments

I would like to thank the legion members of the Weitz lab and associated non-members for helpful and thought-provoking discussions, as well as scientific and career advice, and generally making the lab a great place to collaborate. As anyone who has been part of the lab knows, it is a highly interconnected network of many researchers with wide-ranging expertise from a vast array of different fields. In addition to those people officially in the lab, I constantly found that other people's collaborators from outside the lab greatly enriched the spirit of scientific inquiry that I benefited from on a daily basis. It is rare among research groups to be able to say that within the lab, I have had the pleasure of interacting with hundreds of former and current group members, and have enjoyed exposure to innumerable research topics far afield from my own areas of focus, which I would otherwise have never heard of. Because of this, I realize that I am indebted not only to the Biofluideros for their biomicrofluidics expertise, but also to all the other subgroup members, for asking unusual questions that only people completely unfamiliar with biology, and with pristine minds free from traditional biological assumptions and dogma, would think to ask. I am truly grateful for what has been a wonderful and intellectually stimulating experience.

There are a number of people that I would like to thank in particular for their invaluable contributions to my graduate experience. I would not have been able to complete the work described in this dissertation without all of my coauthors, and I deeply appreciate the guidance and mentorship I received from Jeremy Agresti and Peter Turnbaugh. I would additionally like to thank all my collaborators on not only projects included in this dissertation, but also many pilot experiments on other topics. I have been very fortunate to have had the opportunity to explore questions in all of your diverse research areas. I have also benefited greatly from scientific and

technical advice, coaching on writing and speaking, and much more, and for that I especially thank Ilke Åkartuna, Assaf Rotem, Linas Mažutis, Allon Klein, Manu Tamminen, Peter Lu, John Heyman, Christoph Muus, Édouard Duliège, Tom Kodger, Yizhe Zhang, Adam Abate, Sorell Massenburg, Ming Guo, Scott Tsai, Don Aubrecht, Amy Rowat, Lloyd Ung, Lolo Jawerth, Eliza Morris, Huidan Zhang, Jimin Guo, Lawrence David, Sarah Köster, Katie Humphry, Julie Brouchon, Anderson Shum, Sujit Datta, Kate Jensen, Tina Lin, Rodrigo Guerra, Jesse Collins, Hee-Sun Han, Mark Romanowsky, Wynter Duncanson, Jim & Connie Wilking, Andy Utada, Adrian Pegoraro, Pascaline Mary, Shmuel Rubinstein, Allen Ehrlicher, Elad Stolovicki, Ralph Sperling, Emily Russell, Rhutesh Shah, Shin-Hyun Kim, Sebastian Seiffert, Oni Basu, Tony Hung, Kirk Mutafopulos, Jon Didier, Pascal Spink, Ruihua Ding, Jesse Collins, KC Hung, Jerome Fung, Tom Dimiduk, the rest of the Manoharan lab, the Turdutton labs, Naveen Sinha, Steffi Utech, Tom de Greef, Roy Ziblat, Graham Rockwell, Tina Huang, Hao Pei, Nick Carroll, Melaku Muluneh, Laura Adams, Tommy Angelini, Benjamin Huang, Bingjie Sun, Thomas Franke, Tuomas Knowles, Julian Thiele, Esther Amstad, Yaniv Edery, Zsolt Terdik, Dorota Koziej, Vasily Zaburdaev, Max Eggersdorfer, Kisun Yoon, Chang-Hyung Choi, Joe McDermott, Dong Chen, Naiwen Cui, Wenshan Zheng, Ye Tao, Yong Guo, Jules Thiery, Yulong Han, Max Zieringer, Peter Yunker, and many others whom I am sure I am forgetting to mention. Thank you for all your distractions during benchwork, which were likely unhelpful from a scientific standpoint, but always enjoyable from a personal perspective; thank you for making me appreciate just how miraculous 4-way stop signs are; and thank you for putting up with my eclectic music taste in LISE. You have all been excellent colleagues, and I will always treasure my memories of time spent in the lab with you.

Most of all, I want to thank my advisor, Dave Weitz, for bringing all of these wonderful people together under the same roof. Dave: thank you for connecting all of us with so many people working on such interesting problems. I am sure there is no other group where I would have gained nearly as much experience on how to collaborate, or communicate. Thank you not only for asking the right questions, and teaching me how to do so, but also for unfailingly knowing who else within a 5 mile radius would ask even better questions on any topic. Moreover, thank you for being so encouraging and supportive of my pursuing topics that drew my interest, while guiding me to narrow down the pivotal question that really mattered in each. Any time I came to you with a new intriguing problem or an idea for a solution, your excitement about it acted as constructive interference for my own – and even when I was confused or frustrated about a scientific problem, or had some dilemma about my own academic or career path, I always left your office with greater clarity than when I entered it, by at least an order of magnitude. It is an experience unique to the Weitz lab to sit in your office for hours after the end of group meeting, discussing data and experimental techniques with 10 other people, and I will miss it very much.

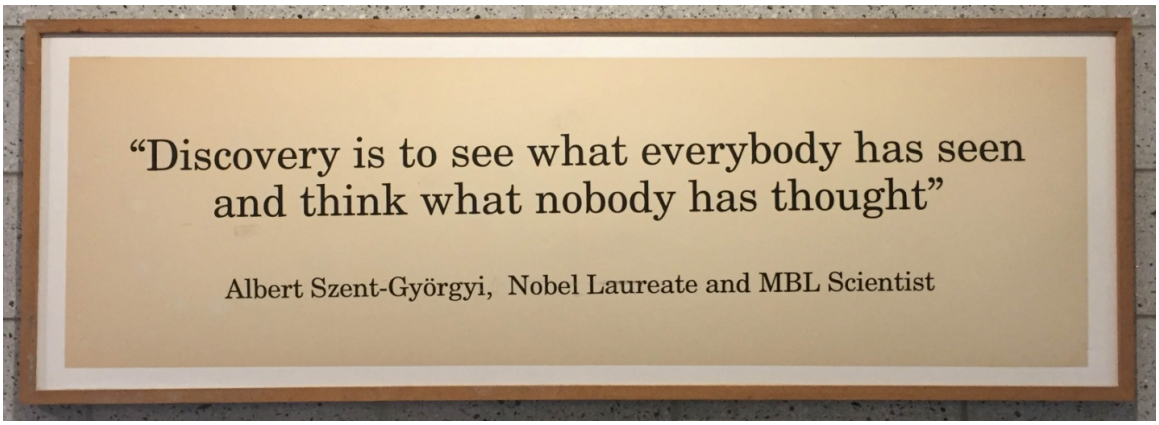
Outside the lab, I would like to thank Vinny Manoharan, Michael Brenner, Dan Needleman, Dave Mooney, Ron DePinho, Marcia Haigis, and Philippe Cluzel for their mentorship and advice, as well as their time sitting on my qualifying and dissertation committees. I thank the FluidicMEMS organizers and the greater Boston microfluidics community, for opening my horizons and giving me a regular taste of the breadth of microfluidics applications in both academia and industry. I thank the excellent staff at the Bauer Core Facility, especially Brian Tilton, Tricia Rogers, Mandy Tam, Christian Daly, and Claire Reardon; at the Center for Nanoscale Sciences, especially Fettah Kosar and Greg Lin; and in the SEAS Facilities Office,

especially Don Claflin, Tom Tribble, Stuart McNeil, Kenneth Collins, Michele Waters, and Jason Ortega, for doing an excellent job behind the scenes keeping our lab facilities running.

Finally, I thank you for your interest in reading this dissertation, and welcome any inquiries or comments on this work at mtguo@alumni.harvard.edu.

Dedication

This dissertation is dedicated to my parents, who have always been an inspiration for inquiry about the natural world and beyond, and who were the first people I could look up to as model scientists. Mom, thank you for always pushing me to be a tough judge of my own work. Dad, thank you for always reminding me that no matter what I end up doing, what's most important is for me to do it well. Thank you both for always supporting me in all my endeavors and interests.



List of Abbreviations

DNA	deoxyribonucleic acid
gDNA	genomic DNA
FACS	fluorescence-activated cell sorting
FCM	flow cytometry
FISH	fluorescence in situ hybridization
MDA	multiple displacement amplification
eMDA	emulsion MDA
OTU	operational taxonomic unit
PCR	polymerase chain reaction
epicPCR	emulsion, paired isolation, and concatenation PCR
gePCR	gel emulsion PCR
SNP	single-nucleotide polymorphism
WGS	whole-genome sequencing

Chapter 1: Droplet microfluidics for high-throughput biological assays

Droplet microfluidics offers significant advantages for performing high-throughput screens and sensitive assays. Droplets allow sample volumes to be significantly reduced, leading to concomitant reductions in cost. Manipulation and measurement at kilohertz speeds enable up to 10^8 samples to be screened in one day. Compartmentalization in droplets increases assay sensitivity by increasing the effective concentration of rare species and decreasing the time required to reach detection thresholds. Droplet microfluidics combines these powerful features to enable currently inaccessible high-throughput screening applications, including single-cell and single-molecule assays.

1.1 Introduction

Many technologies and resources developed over the past several decades have greatly impacted medical research, therapeutics, and diagnostics. Some applications of these technologies and tools would greatly benefit from increased throughput and sensitivity. For example, small compound libraries are powerful resources that can be screened to discover new drugs, but drug discovery can require screening of as many as one million variants.[2] This large number makes compound screens extremely expensive and time-consuming, and thus less accessible. As another example, diagnostic assays that detect pathogens in the bloodstream can easily identify species that typically infect at high concentrations, but are severely hampered or completely obstructed by the long turnaround times required to detect species that occur at concentrations as low as 5 cells/mL.[3] Such low signal-to-noise ratios restrict the possible targets of medical diagnostics, and thus limit their medical utility. These applications, and many others, require screening of small volumes at high rates and with high fidelity.

Droplet microfluidics addresses the need for lower costs, shorter times, and higher sensitivities by using water-in-oil emulsion droplets to compartmentalize reactants into picoliter volumes,[4-8] instead of the microliter volumes commonly used with standard methods. These droplets increase throughput by reducing the volume and increasing the rate at which assays can be performed. Other microfluidic approaches that handle similar volumes use chambers for compartmentalization.[9-12] While these approaches are very powerful, using chambers can introduce the risks of fouling and cross-contamination between samples. Applying fluidic control directly to the reactants may also complicate sample manipulation and retrieval. In contrast, the oil that carries droplets prevents undesirable interactions between reactants and solid surfaces, and facilitates rapid manipulation by separating the reactants from the fluidics.

We will describe those aspects of droplet-microfluidic technology that enable high-throughput screens and sensitive assays. We will discuss the unique advantages that the droplet-microfluidic approach provides, as well as its limitations. To highlight these advantages, we will describe representative examples of applications that are challenging to perform with conventional high-throughput screening methods, and are facilitated by droplet-microfluidic techniques.

1.2 Droplet microfluidics

Assays generally require multiple steps, such as compartmentalization, manipulation, and measurement. Executing these steps using droplet microfluidics can maximize throughput.

Microfluidic devices can be used to compartmentalize reactants by using an inert carrier fluid, usually oil, to encapsulate small volumes of aqueous reagents in droplets and separate the fluidics from the droplet contents. Such devices can produce monodisperse droplets,[6, 13] ranging

in volume from 0.05 pL to 1 nL, or 5 μm to 120 μm in diameter. Droplets can encapsulate cells, DNA, and other particles or molecules that are in the inner aqueous phase (Figure 1.1a).[14] If desired, molecules and particles can be singly encapsulated in individual droplets.[15]

Following an initial encapsulation step, water-in-oil droplets can be manipulated in several ways. Existing droplets can be collected off-chip in a microtube (Figure 1.1b), then reinjected into another microfluidic device (Figure 1.1c).[16] Picoliter volumes of additional reagents can be injected into existing droplets at rates of ~1 kHz.[17] As aqueous droplets flow in carrier oil past a channel containing the aqueous reagents to be added, the interface between each passing droplet and the picoinjection channel can be controllably destabilized, so that some volume is added to the droplet during the brief transit time (Figure 1.1d). Reagents can also be combined by coalescing droplets and allowing their contents to mix.[18] Conversely, other geometries can subtract volume by splitting droplets, with volume ratios controlled by differential channel resistance (Figure 1.1e).[19] This allows controllable quantities of a single droplet to be split off, enabling multiple assays of the same droplet contents, even if the assays are mutually incompatible.[20] Finally, drops can be detected on-chip and efficiently sorted, allowing selection of subpopulations based on a variety of readouts (Figure 1.1f).[21-23]

These microfluidic techniques can be combined to perform nearly any biological reaction or assay that can be performed in a conventional microtiter well plate. Indeed, many of the most commonly used assays in biological research have already been demonstrated in droplet microfluidics,[7, 14, 24-28] and even more are in development. There are several key reasons to use droplets rather than well plates: small volumes, high speeds, low noise, and isolation of droplet contents from solid surfaces and fluidic control. Small volumes conserve expensive reagents, and high speeds significantly reduce the time required to assay extremely large libraries. Droplets

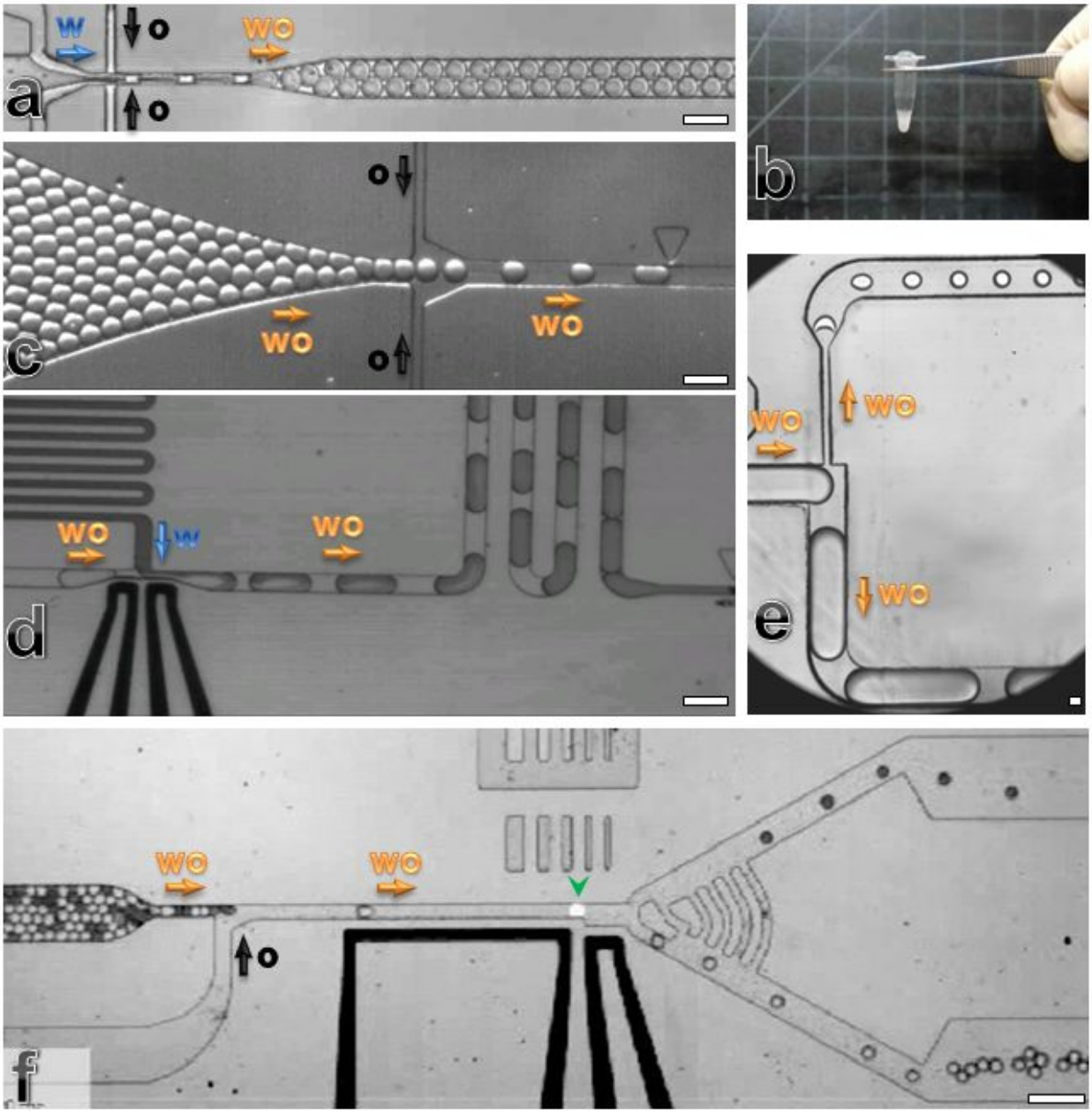


Figure 1.1: Manipulations with droplet microfluidics. (a) A droplet-producing device. An aqueous solution is coflowed with carrier oil to produce water-in-oil emulsion droplets that encapsulate any cells, particles, or molecules present in the aqueous solution. (b) One million droplets, each 25 μm in diameter, are visible as ≈20 μL of emulsion stored in a 200 μL microtube. (c) A reinjection device. Previously formed droplets are taken from off-chip storage and reinjected into a microfluidic device. Oil is added as a spacer between the closely packed droplets. (d) A picoinjection device. Droplets (wo) flow through a T-junction, past a side channel containing a second, blue-dyed aqueous solution (w) to be injected into the droplets. Electrodes (thick black lines) are charged to produce an electric field that locally destabilizes the interface of the droplet passing by the side channel, so that the aqueous solution briefly fuses with the

Figure 1.1 (Continued): droplet, injecting additional fluid. The dye quickly diffuses throughout the droplet. (e) A splitting device. Droplets flowing through a T-junction split into two droplets, with the size of each daughter droplet determined by the geometry of the device. (f) A sorting device. Dark droplets containing blue dye and light droplets containing a fluorescent dye are pooled together and reinjected into a device. A laser spot (bright spot below arrowhead) interrogates the droplets. Electrodes (thick black lines) are charged according to the emitted fluorescence, to deflect only the light droplets into the lower channel. Scale bars denote 100 μm. Arrows indicate the direction of flow for different phases: aqueous fluid (w), carrier oil (o), and water-in-oil emulsion droplets (wo).

achieve low noise by reducing reaction volumes; any background noise present in a solution will be decreased proportionally to the volume. Various materials can be used to form picoliter-sized compartments, but the inert oil-water interface of a droplet has the advantage that it shields droplet contents from the solid walls of a microfluidic device. Droplets thus minimize fouling and cross-contamination. These advantages are exemplified by many applications.

High-throughput screens can especially benefit from small volumes. A typical screen using conventional robotics involves pipetting volumes as small as ~10 μL of reagents into each well of a 384-well microtiter plate, and adding ~10 μL of each variant from a library into different wells. This results in a total reagent volume on the order of milliliters for one microtiter plate, and significantly higher as the number of plates increases. In contrast, the same screen using droplet microfluidics would involve emulsifying volumes as small as ~1 pL of each variant from a library, and adding ~1 pL of reagents to each variant droplet.[28] This represents a reduction in volume by up to 7 orders of magnitude. Since reagent volume is often the limiting factor for how many reactions can be screened, this volume reduction is a critical factor that helps enable droplet microfluidics to achieve extremely high throughput.

Droplet microfluidics can often also achieve higher speeds by enabling reaction detection after shorter time courses. If a single molecule or cell is present in a 1 μL volume, its concentration will be extremely low. If that 1 μL volume is emulsified into one million droplets,

each 1 pL in volume, one of those droplets will contain the single molecule or cell, and its concentration in that droplet will be effectively increased one million-fold. Since reaction rates increase with effective concentration, reaction times that are normally on the order of hours in bulk may decrease to seconds or minutes in droplets.[29, 30] In addition, the minimum concentration of reaction products required for detection will be reached more quickly in smaller volumes. The detection step itself can also be performed rapidly, at rates of ~1000 reactions/second. This rate ultimately limits the number of samples that can be screened, as well as the time resolution of any measurements. If there are ~10^5 samples to be assayed, then time series measurements may be taken every few minutes. For a single endpoint measurement, it is realistic to screen ~10^8 samples/day.

An additional benefit for cell-based assays is the automatic association of genotype with functional output through compartmentalization rather than through a physical bond. This is in contrast to screens that involve fluorescence-activated cell sorting, and could replace membrane-binding systems such as yeast surface display or phage display.[14, 22, 31] Droplets thus offer a particular advantage for secretion assays that may be difficult to adapt to such membrane-binding systems.

Droplet microfluidics does however introduce a significant disadvantage: heterogeneous assays are difficult to perform. For example, many secretion assays normally incorporate washing steps; adapting these to droplets is very difficult. Encapsulating homogeneous liquid reactants and taking a single measurement is straightforward, but removing a buffer entirely from a cell and exchanging it for a different buffer requires more complicated techniques, as discussed in Section 2.2.

1.3 Droplet libraries and compound screens

Conventional high-throughput screens use microtiter plate wells to store a large number of unique elements; each element is then tested for some chemical or biological property. However, high-throughput screens in droplet microfluidics require that the unique elements be stored, manipulated, and tested in droplets. Thus, droplet-based screens require the encapsulation of each element in droplets prior to performing the screen itself. Although it is possible to emulsify and screen each element individually, this would take more time than performing a conventional screen; instead, after encapsulation, droplets of different elements can be pooled into a "droplet library," ready for subsequent use in a single screening assay that includes all library elements. Since each element is emulsified into many droplets, the same library can be used for multiple screens, where each assay uses an aliquot, or small portion, of the library, and each aliquot includes many droplets of each library element.[28]

As an example, small compounds dispensed in microtiter plates, totaling tens of microliters of each element, may be encapsulated to form a droplet library of billions of droplets at a rate of several minutes per plate. This is achieved by parallelizing microfluidic devices and interfacing them with microtiter plates, as shown in Figure 1.2a. An aliquot of one thousandth of this library contains thousands of copies of each library element and this constitutes sufficient statistics for one assay. To perform a screen, the aliquot of droplets is reinjected into a microfluidic device, a target such as a microbial organism is added to each drop, and after some reaction occurs during an incubation step, the droplets are detected and sorted at a typical rate of 1000 Hz (Figure 1.2b). A library of one million compounds can be screened in a day as opposed to months using conventional screens, while using one thousandth of the amount of reagents that is conventionally used.

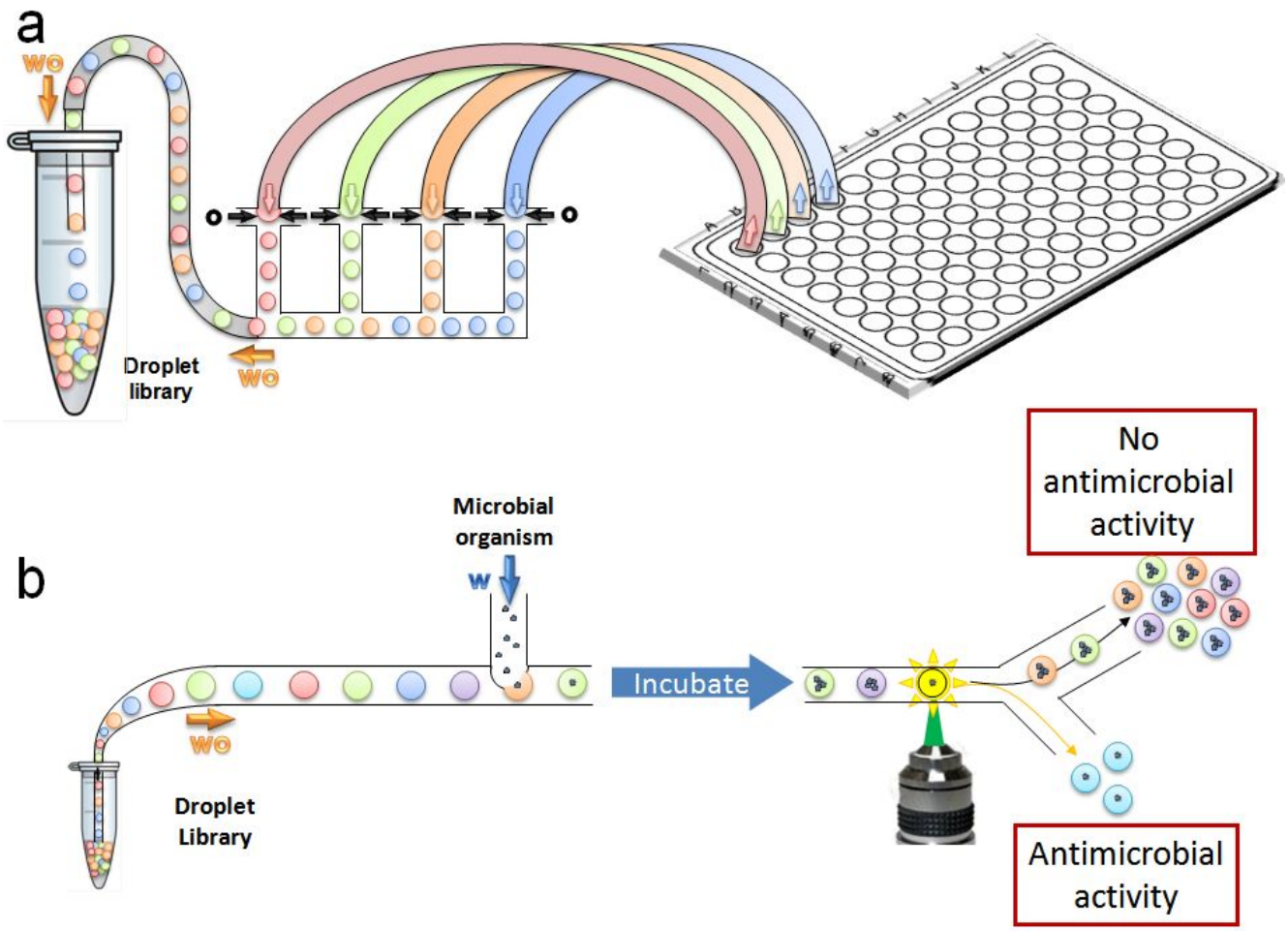


Figure 1.2: Droplet library production and small compound screen. (a) The wells of a microtiter plate are interfaced with the aqueous inlets of multiplexed droplet-producing devices (colored arrows) to simultaneously emulsify the contents of each well. The droplets are pooled together to form a droplet library. Each well contains a unique barcode, represented in the figure as a different color, that is carried by the droplets from that well and is used to decode the contents of each droplet. (b) A droplet library of small compounds is screened for antimicrobial activity. An aliquot of the library is reinjected into a microfluidic device and a single microbial cell is added to each droplet. After incubation, the droplets are screened for growth arrest of the microbe. The small compounds encapsulated in the selected droplets are identified by their barcodes and noted as potential antimicrobial drugs. Arrows indicate the direction of flow for different phases: aqueous fluid (w), carrier oil (o), and water-in-oil emulsion droplets (wo).

When the contents of a microtiter plate are emulsified and pooled to form the droplet library, sample labeling encoded by spatial positioning in the original plate is lost; thus, some other means must be developed to track the library contents. One solution is to associate a unique

barcode with each element and include it with the element in the plate prior to encapsulation. Each droplet then contains a barcode that indicates which element is encapsulated in the droplet. Potential barcodes include nucleic acid sequences and fluorophore combinations.[32, 33]

The advantage of using fluorophore combinations as barcodes for droplet libraries is that they can be read in real-time concurrently with the result of the assay. However, the number of distinguishable fluorophore combinations is limited by the dynamic range of the optical setup that detects the fluorophores. In contrast, nucleic acid sequences cannot be read in real-time but can accommodate arbitrarily large library sizes. Many screens do not require real-time decoding of the barcodes: since the goal of most screens is to isolate a rare element out of a large library, only a small pool of elements are selected and these can be decoded after the screen is completed. Thus, barcodes such as nucleic acid sequences are often a good fit for large screens.

The diverse chemistry of library elements may affect whether chemicals leak out of the water-in-oil emulsion droplets and this could lead to crosstalk between droplets. This occurs if the inert oil and surfactant are permeable to the compound. Chemicals may also affect the physical stability of the droplets encapsulating them. This can happen for example if the chemicals counteract the stabilizing effect of the surfactant, resulting in coalescence during fluidic manipulation, thermocycling, or long-term storage. Appropriately engineered devices and surfactants can mitigate these issues of instability, leakage, and crosstalk, but these issues must be evaluated and minimized in droplets whenever new reagents are used in a library.[16]

Many high-throughput assays screen genetic elements. These may be DNA in cells or viruses, or nucleic acids that encode proteins. In the case of genetic elements, each cell, particle, or molecule constitutes a single element of the library; thus, if a suspension of all elements mixed together is sufficiently diluted before encapsulation, then each droplet will contain just

one element, and the droplet library can be formed using a single microfluidic device. Moreover, the unique genetic sequence that is the object of the screen also serves as a barcode, so that no additional barcoding is required when encapsulating the library. Genetic screens have wide-ranging applications, including directed evolution,[34] SNP measurement and identification,[12] in vitro translation,[7] and metagenomic analyses.[11]

1.4 Cell growth assays and cell-cell interactions

Cell growth assays are widely used in biological and pharmacological screens because growth is often a valuable indication of the effect of a tested condition on a targeted cell. Cell growth is quantified by measuring cell density, and these measurements are usually limited by a density detection threshold, under which the existence of cells cannot be detected. A major advantage of using droplets for cell growth assays is therefore the drastic reduction of the volume in which the cells are grown, which proportionally increases the density of a given population of cells. Thus, the minimum number of cells that can be detected in droplets is lower than in bulk. As a numeric example, the density of a single cell in a 10 μm drop is equivalent to that of a billion cells in a 1 μL well. For many assays, this means that in droplets, the presence and growth of single cells is easily detected.[14]

Assaying cell growth at the single-cell level is crucial for example in diagnostics, where a sample suspected of containing a pathogen is cultured until the population density of the pathogen is high enough to detect, at which point the source of the infection can be identified.[3] Because the density is so much larger, performing this test in droplets can substantially decrease the time required to detect an infection, in some cases even down to the doubling time.

Monitoring single-cell growth in droplets is also important for detecting unusual growth rates.[35-37] Currently, since growth is measured for populations of many cells, the fastest-growing cells dominate the population, and cells with lower growth rates cannot be easily monitored. However, cells with intermediate or fluctuating growth rates may prove to be extremely important once they are isolated and studied using droplet-based screens. For example, detecting rare growth events could improve our understanding of emerging antibiotic resistance, or emerging viral epidemics.

Culturing conditions inside droplets may differ from those in bulk, even with the same medium and external environment. For example, the amount of nutrients available to each cell is initially much smaller than in bulk, the permeability of the emulsion to gases is different from that of bulk aqueous media, and the boundaries of the droplet are generally less adherent than the plastic in a culture flask.[14, 38] Correspondingly, culturing cells in such environments may require optimization per cell type: droplet size must be tuned to adjust nutrient depletion, carrier oil and storage conditions must be chosen to adjust oxygen availability, and beads inside droplets may be necessary as growth substrates for certain cell types.

1.5 Bacterial persistence screens

One example of an application where single cell growth is important is screening for bacterial persistence.[35] A population of bacteria may contain some bacteria with weak antibiotic resistance, and some with high antibiotic resistance. The latter will be more fit and thus highly persistent, while the former will have intermediate persistence. In large volumes, individuals with intermediate persistence will be outgrown by the more fit bacteria before their persistence can be detected and studied. To isolate intermediate persistent bacteria using conventional

methods, cultures are diluted and spread on agar plates so that newly formed colonies are isolated from each other.[39] Several colonies can be harvested from each plate after a time course of up to several days. Colony harvesting is a laborious and poorly scalable process. Moreover, if the persistence is not stable over several generations, colonies may not be detected at all. In contrast, if single bacteria are encapsulated in droplets, even bacteria with the weakest and most transient persistence can be detected and selected at rates of 1000 Hz.

1.6 Plaque assays and virus-host interactions in drops

Another example of a cell growth assay that can be improved by using droplets is the virus plaque assay.[40] Plaques are macroscopically visible patches on a culture plate where cells have died due to a spreading viral infection. In well plates, these plaques expand for as long as 48 hours before the virus infects enough cells to be observed,[41] while in droplets, infections of single cells can be detected after much shorter incubation times. Moreover, single virus particles that are able to infect new host cells but are deficient in some other aspect of propagation would not be detected with conventional plaque assays, yet such viruses may be crucial for understanding the emergence of viral epidemics. To assess host-virus interactions in droplets, host cells can be encapsulated and incubated to establish a baseline before viruses are added to the droplets. After the droplets are incubated through one replication cycle of the virus, they are screened for changes in biological function, such as host cell death or virus copy number (Figure 1.3).

1.7 Antibody screens

Antibodies are used in many fields ranging from biomedical and pharmaceutical research to medical diagnostics and therapeutics.[42] They are able to specifically target different molecular

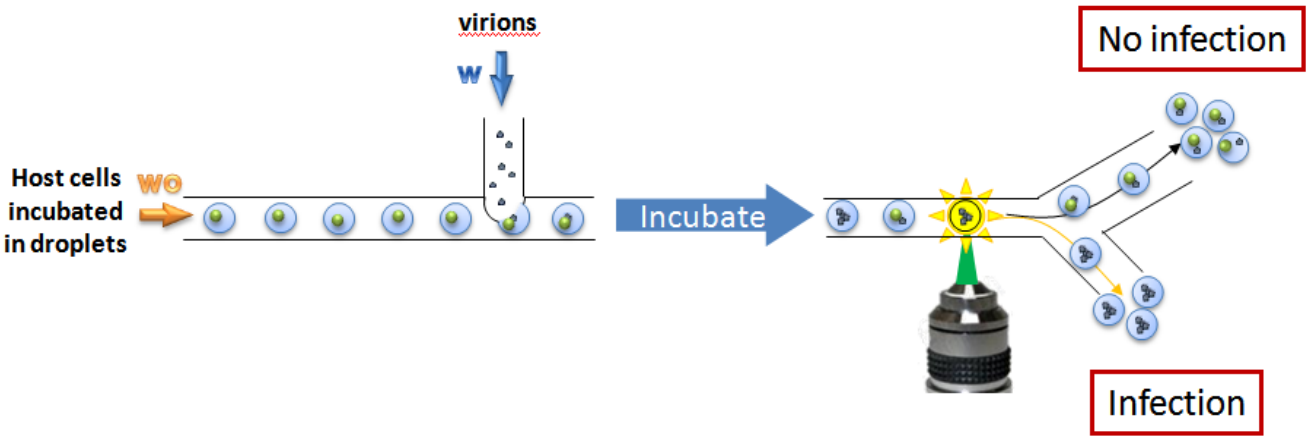


Figure 1.3: Droplet-based plaque assay. Media containing virions is picoinjected into droplets containing host cells. After incubation, droplets are screened for successful virus replication and release, and infectious virions are selected. Arrows indicate the direction of flow for different phases: aqueous fluid (w), carrier oil (o), and water-in-oil emulsion droplets (wo).

domains, which is essential for applications such as quantifying specific proteins and identifying pathogens. In the standard method to develop a monoclonal antibody,[43] antibody-secreting B cells are first isolated from the spleen of a mouse immunized against the desired target. Each B cell produces a single species of antibody, and only a relatively small fraction of species will bind to the desired target. Because a single cell would die before it could fill a standard plate well with a detectable concentration of antibodies, B cells must first be fused with immortal cells to form hybridomas, which can be cultured to generate cell lines. To generate pure antibodies with the desired target specificity, these cell lines must be derived from single cells. Such monoclonal cultures are obtained through several steps. Cultures are initially polyclonal, containing a mixture of several types of hybridomas. These cultures are screened for the desired antibodies, and if the antibodies are found in a given culture, the specific cells responsible for producing them must be singly isolated and cultured again to produce monoclonal cultures, with each clone derived from the divisions of a single hybridoma cell. All cells within each homogeneous clone will produce identical antibodies that have the same binding affinity to the

desired target. Obtaining and validating monoclonal cultures typically takes 10-12 weeks. Droplet microfluidics can bypass the repeated culturing and screening steps by encapsulating and screening single antibody-producing cells directly. A single cell in a 0.05 – 1 nL droplet can produce a detectable quantity of antibodies in hours or even minutes, compared to days or longer in a 0.3 – 1 mL well.[14, 30] In addition, droplet microfluidics can perform the screening step at a much higher rate. At typical sorting rates of ~150 droplets per second, ~10^6 individual cells can be analyzed and sorted in one day. In a typical plate-based screen, ~200 different cells can be analyzed in 10-12 weeks.[43] Thus, relative to standard screening methods, direct droplet screening of single cells enables much higher throughput.

An antibody screen must distinguish between secreted antibodies that bind to the target and those that do not. In plates, this is done by washing: a solid surface displaying the targets is provided, the desired antibodies bind to the surface, and the nonspecific, undesired antibodies are washed away. The specifically bound antibodies that remain can then be detected optically.[43] This type of washing-based assay is a heterogeneous assay. Such assays are typically difficult to accomplish in droplet-based microfluidic systems, in which washing steps are difficult, and must therefore be modified for use in droplets. To overcome this limitation, a droplet-based solution replaces the washing-based assay with an assay based on signal concentration. This method encapsulates cells together with a bead loaded with target molecules. As the cells secrete antibodies, only those antibodies that are specific to the target bind to the co-encapsulated bead. Thus, specific antibodies produce a tightly localized signal inside a droplet, while nonspecific antibodies remain distributed throughout a droplet and produce a diffuse signal (Figure 1.4). A microfluidic sorting device can then select those droplets that contain the desired antibodies, along with the cells that produced those antibodies. Droplets that encapsulate cells with beads

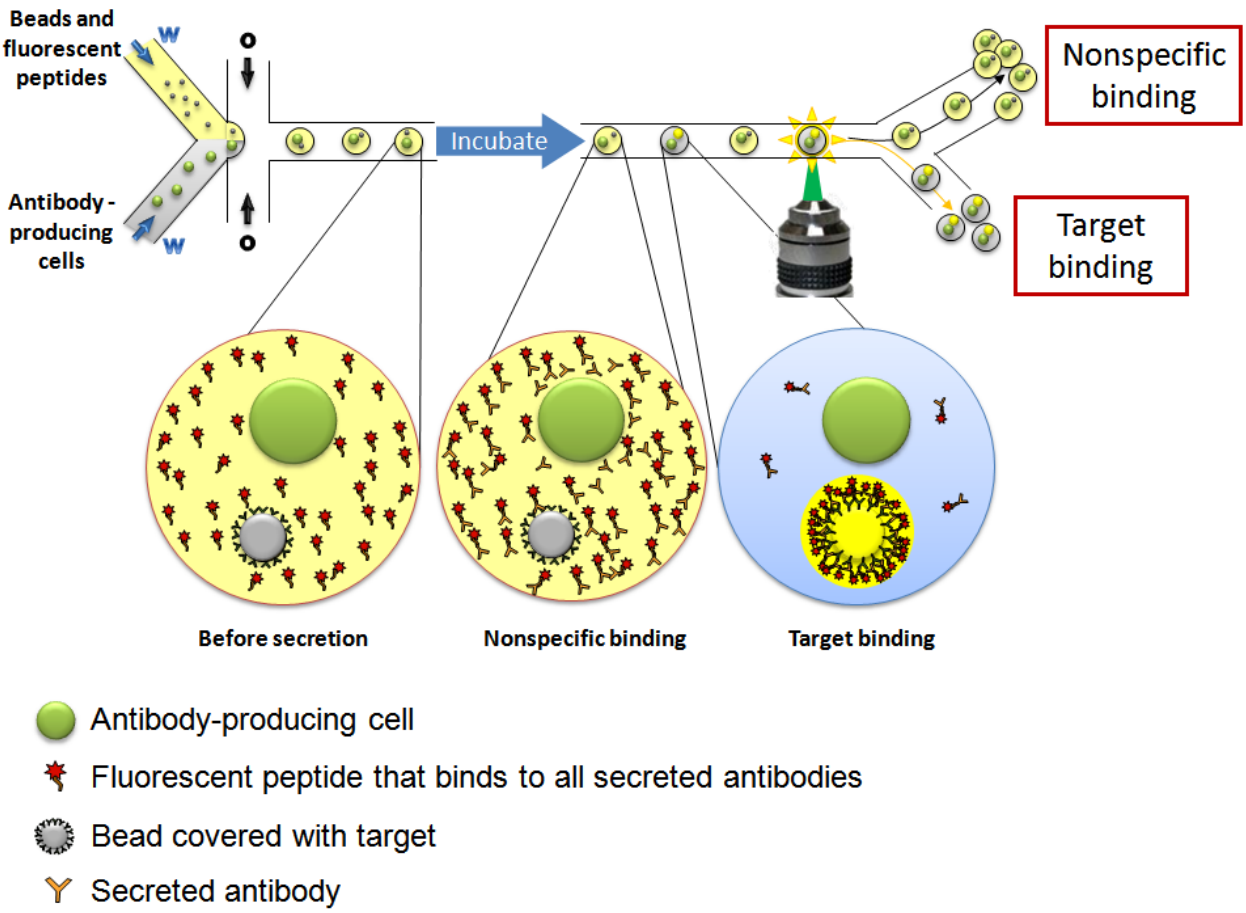


Figure 1.4: Droplet-based antibody screen. Immediately before encapsulation, antibody-producing cells are coflowed with a mix of target-covered beads and fluorescent peptides that bind to all secreted antibodies. Before incubation, the peptides fill all droplets with homogeneous fluorescence. After incubation, in droplets containing antibodies that bind specifically to the target, the antibodies will primarily be attached to the bead, so the peptides bound to the antibodies will produce fluorescence concentrated in a small, bright spot within the droplet. In droplets containing nonspecifically binding antibodies, the fluorescence will remain homogeneously distributed throughout the droplet. Droplets are then screened for antibody binding specificity. Arrows indicate the direction of flow for different phases: aqueous fluid (w), carrier oil (o), and water-in-oil emulsion droplets (wo).

thus represent a heterogeneous system that can yet be manipulated fluidically. This system demonstrates that assays commonly performed in microtiter plates can be adapted for use in microfluidic droplets.

This droplet-microfluidic approach can potentially address a second inefficiency, that of hybridoma production. Fewer than 1 in 100,000 B cells that are subjected to fusion become

viable hybridomas.[43] This dramatically limits the number of potential antibodies that can be isolated. Moreover, this may in fact not be a randomly distributed success rate. For example, some desired B cells, such as those that secrete at a very high rate, or those that secrete antibodies against a particular target, might be resistant to fusion. Hybridoma screening methods are inherently unable to access any B cells that do not stably undergo fusion. Since droplet microfluidics increases the screening rate so dramatically, it becomes feasible to screen single cells, and therefore becomes feasible to screen large numbers of B cells directly, without first immortalizing them through hybridoma generation. This would allow extremely deep mining of the antibody space, and increase the chances of isolating useful antibodies from a single screen. Droplet microfluidics can also screen cells for which there are no robust fusion partners, such as human B cells.[44] Cells from a patient recovering from an infection could be especially valuable.[45] Once the desired B cells are selected, their antibodies can be produced on an industrial scale by retrieving the antibody-encoding gene sequences from the B cells and cloning these sequences into an expression system.

1.8 Directed evolution of enzymes

Directed evolution is a powerful method for developing new variants of enzymes and proteins in general.[34, 46] This method generates a DNA library of mutations of a known enzyme, clones the library into bacteria, and screens the enzymes produced by the bacteria for some desired property, such as increased catalytic efficiency, solubility, or specificity for particular substrates. If the library contains a very large number of mutations, a few variants may have the desired activity. PCR can then be used to retrieve the sequences of those variants from the selected bacteria. To facilitate detection of enzymatic activity, the mutation library is often

cloned into a special strain of bacteria that has been engineered with a reporter, which produces detectable changes proportional to the level of enzymatic activity. These changes may be based on fluorescence or growth.[34]

For large libraries of putative novel enzymes, encapsulating the bacteria library in droplets can increase screening throughput by reducing the reaction volume.[14, 22, 31] This increases the effective concentration of reporter molecules, and thus the incubation time required before detection. Droplets also facilitate handling of large sample numbers. It is particularly significant that through compartmentalization, droplets can maintain the association between observed enzymatic activity and its genetic source. When screening enzymes, it is generally desirable to recover the DNA sequences that code for any selected enzymes. With droplet microfluidics, it is straightforward to simultaneously sort out both phenotype and genotype.

In addition to libraries generated by mutagenesis, environmental samples such as soil, water, and insect guts are a promising source for novel enzymes.[47] Environmental sources harbor a diverse reservoir of uncharacterized bacteria and archaea that have been shown to carry potentially useful DNA polymerases,[48] restriction endonucleases,[49] and metabolic enzymes.[50] To systematically screen environmental samples, extracted DNA could be sheared into short fragments, then cloned into the same reporter-engineered bacteria used for mutation libraries. However, extremely large libraries would be required to fully probe the high diversity found in environmental samples. The sheared DNA fragments should be approximately the size of a typical gene, about 1 kilobase. Since bacteria genomes are usually on the order of 1 megabase, a few million randomly sheared 1 kb fragments would be necessary to cover most of a single genome. One gram of soil contains on the order of 10^3 different genomes,[51] so a library would

have to contain ~10^9 fragments to cover all the genetic diversity in 1 g of soil. Droplet microfluidics can access such large numbers.

1.9 PCR-based analysis of single, rare templates

The polymerase chain reaction (PCR) has wide-ranging applications, especially in medical diagnostics. PCR produces many copies of a DNA template, and is therefore crucial for analyzing small amounts of molecules or cells that must be amplified to higher levels before downstream detection or sequencing steps.[52] For standard volumes and preparation methods, target templates are amplified from an initial concentration of ~10^6 molecules/mL. This fairly high template concentration is necessary because PCR may amplify nonspecific DNA that does not match the desired sequence, especially under suboptimal conditions. The concentration of specific template, or signal, must be sufficiently high to overcome this noise.[52] However, clinical samples such as blood from sepsis patients may carry pathogenic targets at concentrations as low as 1-10 microbes/mL, against a background of 10^7 white blood cells/mL.[3] Even if the white blood cells are somehow removed, the reaction will easily be contaminated by aerosols, dust, and other incidental sources of noise. Careful and specialized sample preparation is therefore necessary to eliminate extraneous DNA and minimize noise.[53, 54]

Instead of increasing the signal-to-noise ratio by taking extra steps to reduce noise, an alternative approach is to increase the effective concentration of target molecules, which constitute the signal (Figure 1.5a).[25, 31] Droplet microfluidics enables single-template PCR by using this approach.[25, 26, 55, 56] For a 1 mL blood sample containing 10 microbes and 10^7 white blood cells, compartmentalization into 1 pL droplets will produce a few droplets containing ~1 microbe at a concentration equivalent to 10^6 cells/mL, and ~10 white blood cells at 10^7 cells/mL.

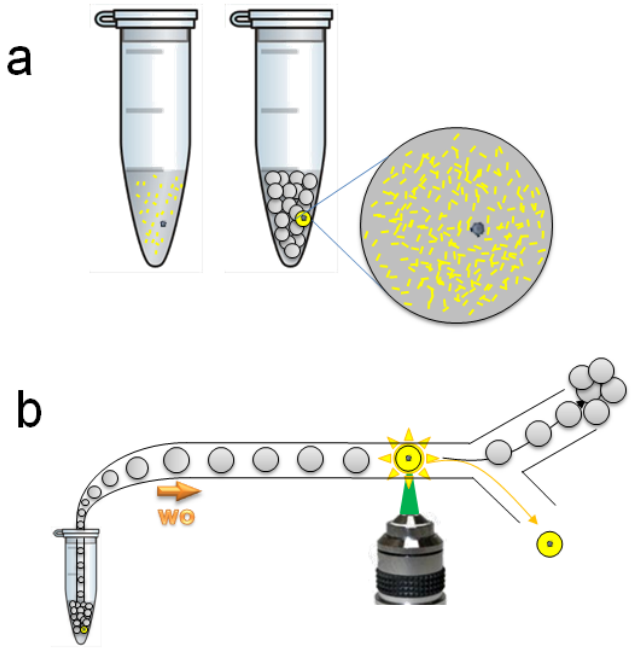


Figure 1.5: PCR-based selection of rare templates. (a) A single, rare template (dark spot) is amplified, producing a fixed number of PCR products (bright fragments) that may be fluorescently stained. In bulk (left), the PCR product concentration is low, while in droplets (right), the effective concentration is much higher, since all of the products are confined to the single droplet that contains the single template. (b) After the PCR thermocycle is performed in a standard microtube, droplets are reinjected into a microfluidic device and screened for successful PCR product formation. The single droplet containing the rare template is retrieved for downstream assays such as sequencing. The arrow indicates the direction of flow for water-in-oil emulsion droplets (wo).

These droplets will have a signal-to-noise ratio of ~1:10, and thus will be easily distinguished from the vast majority of the droplets that contain no microbes and thus much lower signal-to-noise. The sample as a whole will be easily distinguished from a healthy blood sample that contains no microbes and produces only low signal-to-noise droplets. The same sample containing 10 microbes would have a signal-to-noise ratio in bulk of ~1:10^6, which would be very difficult to distinguish from a sample containing no microbes.

In addition to noise, another challenge facing single-template PCR is that the enzymatic activity of the polymerase can be inhibited by certain molecules, such as heme, EDTA, and humic acid.[57] Increasing the effective concentration of the signal will improve not only the

signal-to-noise ratio, but also the signal-to-PCR inhibitor ratio. Droplets are thus less susceptible to inhibitor concentrations that would obstruct PCR in bulk. This is especially relevant for blood and soil samples, which often have high concentrations of PCR inhibitors.

Following compartmentalization, emulsion PCR can be performed to amplify some identifying part of the microbial genome to a detectable level.[24, 55] Droplets containing the amplified pathogen can then be detected to diagnose bacterial or fungal infection. This approach is generally applicable to diagnosing any disease where abnormal DNA species circulate at very low levels in the bloodstream. Another example is certain types of cancer, where circulating tumor cells may be present in the blood at 1 cell/mL.[58]

In the same way that compartmentalization enables detection of rare DNA sequences, it also enables higher resolution measurement of sequence distributions within a sample. PCR amplification is biased against sequences with high complexity, such as G/C-rich sequences, and amplifies these sequences with lower efficiency.[59] Thus, the relative concentrations of different sequences cannot be accurately determined from the final concentrations of their PCR products. With emulsion PCR, since each DNA molecule is singly confined within its own droplet, the amplification of each molecule can run to completion without competition from other sequences, regardless of how slowly each reaction runs.

Single-cell PCR would also be highly informative in metagenomic studies.[50, 60-62] Metagenomics analyzes the genomic sequences in mixed populations of microbes to determine what genes are present and how they affect the function of the population as a whole.[47] Interesting populations include collections of diverse but poorly characterized bacteria or virus species from biomedically relevant environmental samples such as soil,[63] sewage wastewater,[64] or the human gut.[65-67] Typical metagenomic datasets determine all the genes present in such a

population by fragmenting the microbial DNA in a sample and sequencing each fragment. This method yields insight into the distribution of species in a population, the metabolic functions that the population can perform, and other biologically relevant information.[47] However, because short DNA fragments from multiple cells are mixed together, all information about which fragment originated from which cell is lost, and there is no way to associate sequences coming from the same cell. Whole-genome assembly from single cells in an environmental sample is therefore impossible.[68] Even if several identical cells in a complex population allow for assembly from overlapping fragments, assembly is still an enormous computational challenge.[69, 70]

Droplets can be used to associate distant portions of the genome from a single cell. Encapsulating single cells or viruses while their genomes are still intact ensures that all the DNA within a given droplet must have originated from the same cell, no matter how fragmented it becomes in later steps, as long as the droplet remains intact. The droplet effectively replaces the cell membrane or virion capsule as a container for the whole genome.

Once single cells or viruses from a population are encapsulated in droplets, single-cell PCR can target an interesting gene previously detected in the population and amplify that gene in every droplet where it is present. Droplet microfluidics can then sort out those droplets containing the PCR products (Figure 1.5b). Each of these droplets must also contain the original genome that carried the gene that produced those PCR products. The selected droplets can then be coalesced and subjected to whole genome sequencing using conventional methods in bulk. Because the selected genomes of interest will constitute a population of much lower complexity than the original environmental sample, whole genome assembly will be much more feasible. In fact, if the selected genomes were very rarely occurring variants within the original population, whole genome assembly would be impossible without screening in droplets. Droplet

microfluidics thus enables whole genome sequencing of rare single cells of interest from complex environmental samples.

1.10 Limitations compared to bulk assays

While droplet-based microfluidics provides enormous opportunities for high-throughput biological assays, it also has some constraints that limit its applications. Droplet microfluidics is especially well-suited for ultra-high-throughput assays, which can process as many as ~10^8 samples/day. However, not all biological assays require such high throughput. It is significantly more efficient to use droplet microfluidics for assays involving at least 10^5 samples. For fewer samples, tools such as 1536-well microtiter plates are generally sufficient. Indeed, droplets suffer from some limitations compared to bulk assays in microtiter plates. For example, while droplets do produce very small volumes, they also produce very high surface area-to-volume ratios. Thus, the oil-water interface that forms each droplet must be both stable and inert. Finding the right surfactant that will maintain such an interface is often a challenging chemistry problem. Similarly, while a library of a billion droplets can easily be handled in a single microtube, droplets cannot take advantage of the spatial barcoding that is available in the two-dimensional arrays of microtiter plates. Thus, droplet libraries require a suitable barcoding system, which can be difficult to develop.

There are also some specific cases where other methods may present advantages over droplet microfluidics. An array of microfluidic chambers and valves arranged in an n-by-n matrix can efficiently map out the full combinatorial space of pairwise interactions within a library,[71] but performing the same exploration in droplets would likely require an impractical number of parallel dropmakers and picoinjectors. Heterogeneous assays involving washing steps are

routinely performed in bulk, but are difficult to adapt to droplet microfluidics. Flow cytometry machines are often used to sort single cells into individual microtiter plate wells, but the technology has not yet been developed to steer single droplets exiting a microfluidic device into individual wells. More generally, interfaces between droplet microfluidic devices and the macroscopic world remain to be developed.

Despite these limitations, droplet microfluidics nevertheless has great potential for many biological assays. Indeed, these limitations do remain challenges, and are the subject of further development.

1.11 Conclusions

Droplet microfluidics offers several distinct advantages that can be leveraged in many ways to improve a wide range of important biological applications. All of the applications we have discussed here can benefit from the ultra-high throughput attainable by droplet microfluidics through the use of picoliter volumes and kilohertz rates for sample manipulation and detection. Ultra-high throughput has particularly significant implications for cost when screening large libraries of precious small compounds. Small volumes lead to an enormous increase in effective concentrations and signal-to-noise ratios, which is critical for performing single-cell assays, such as those used to study bacterial persistence, virus-host interactions, and cell growth in general. High sensitivity is also critical for detecting and analyzing small amounts of biological molecules, such as antibodies, enzymes, and single genomes. Compartmentalization in droplets can be useful as a method for linking phenotype to genotype, in place of genetic techniques such as surface display. Finally, droplet microfluidics not only magnifies the scale of existing high-

throughput screens, but also enables new types of experiments. The single-cell population studies that we have discussed are only one example of many novel studies that should become feasible.

Chapter 2: Selective sequencing of single gut microbes by gel emulsion PCR

We rapidly enrich a population of human gut bacteria for cells carrying variations of a gene of interest, and then obtain the various sequences of that gene found in each individual cell. Emulsion PCR and microfluidically generated hydrogels enable sorting and sequencing of single cells by creating single-cell compartments that persist in both oil and water continuous phases.

2.1 Introduction

The microbiota found in the human gut constitute a complex population consisting of several hundred phylotypes[72] and carrying genes from thousands of functional groups.[73] Individual human hosts carry unique ensembles of microbial phylotypes and functional genes, which often correlate with many aspects of host physiology. Each cell in the ensemble has its own distinct impact on the function of the total community. Sequencing genes from single cells would yield insights into these individual impacts, such as which ones produce specific physiological effects, and which ones are coupled together due to originating from the same cell. For example, the obese human microbiome contains a higher proportion of genes involved in nutrient metabolism than the lean microbiome; obese humans also carry a higher ratio of Firmicutes to Bacteroidetes phylum members than lean humans. Single-cell genetic analysis would reveal how these metabolic genes differ between cells, which nutrient-metabolizing cells are found in the obese gut, and which phylotypes carry which genes. These questions are difficult to address with bulk sequencing techniques.

Single-cell techniques have successfully isolated single cells from complex populations and probed their genetics; nevertheless, existing methods have not yet analyzed large populations of environmental microbiota. Nanoliter-scale PDMS chambers can isolate single cells and offer

convenient spatial addressing,[11] but this method requires extensive manual manipulation, which limits its throughput to fewer than one hundred cells per hour. Moreover, sequencing of unculturable cells isolated from environmental samples has only been demonstrated on 201[74] or fewer[60, 62, 75, 76] cells at a time. In contrast, emulsion droplets offer higher throughput at kilohertz rates, but PCR of whole single cells in emulsion droplets has thus far only been applied to cell cultures or mock populations of culturable cells,[77, 78] and without sequencing the PCR products.

In this chapter, we rapidly enrich populations of unculturable human gut bacteria for cells carrying variations of a gene we target, and sequence the variants of that gene found in different cells. We isolate ~10^7 cells in 1 hour by using high-throughput droplet microfluidics, and sequence the *BT_4738* gene[79] carried by 67 individual cells. Our technique improves upon manual manipulation techniques by sorting through cells at kilohertz rates. Our high-throughput approach, applied here to unculturable bacteria from the human gut, realizes the first step necessary to achieve complete genome sequencing for single cells of interest.

2.2 Methods

2.2.1 Workflow in brief

To obtain sequences of targeted genes from single cells, we first use a microfluidic device to encapsulate single cells together with acrylamide in water-in-oil emulsion droplets. We then polymerize the droplets to trap the encapsulated cellular genomic DNA, so that the resulting microgels can be transferred from the continuous oil phase to a continuous aqueous phase, yet without allowing gDNA from different cells to mix together. Once the microgels are suspended in the aqueous phase, we wash PCR reagents into the microgels, with the exception of reverse primers, which are already covalently incorporated into the polyacrylamide gel network during

the encapsulation step. In each experiment, we also include a negative control of drops containing only reverse primer with no cells; < 0.1% of these microgels are fluorescent after PCR. We reemulsify the microgels before PCR thermocycling, to prevent PCR products from diffusing between microgels. After thermocycling is complete, we cool the emulsion down to room temperature, so that free PCR product strands will anneal to covalently bound reverse strands and will no longer diffuse. We then transfer the microgels once more to a continuous aqueous phase, and wash DNA staining reagents into the microgels. Finally, we use FCM to sort brightly stained single microgels containing PCR products into individual microtiter plate wells, and reamplify and sequence the PCR products from each microgel. We thus determine the various sequences of our gene of interest found in individual bacteria cells (Figure 2.1).

2.2.2 Single-cell encapsulation

We use a previously described flow-focusing geometry similar to Figure 1.1a to encapsulate cells in monodisperse water-in-oil emulsion drops[14] We use soft lithography to fabricate PDMS-glass devices with a junction width and height of 10 μm. We produce 10^8 drops/hour using syringe pump-controlled flow rates of Q_{inner} = 100 μL/hr and Q_{outer} = 150 μL/hr. These are of diameter ≈ 15 um, with CV < 0.05. The outer phase is Novec fluid 7500 (3M) with 1% w/w fluorinated surfactant[16] (in-house synthesis by Ralph Sperling, or purchased from RAN Biotechnologies, EA surfactant). By including an appropriate dilution of suspended cells in the inner aqueous phase, we can produce drops that contain single whole cells. The fill number of encapsulated cells is m ≈ 0.3 cells/drop and follows Poisson statistics, with a slight bias against drops containing exactly 1 cell, likely due to clumping of cells before encapsulation (Figure 2.2).

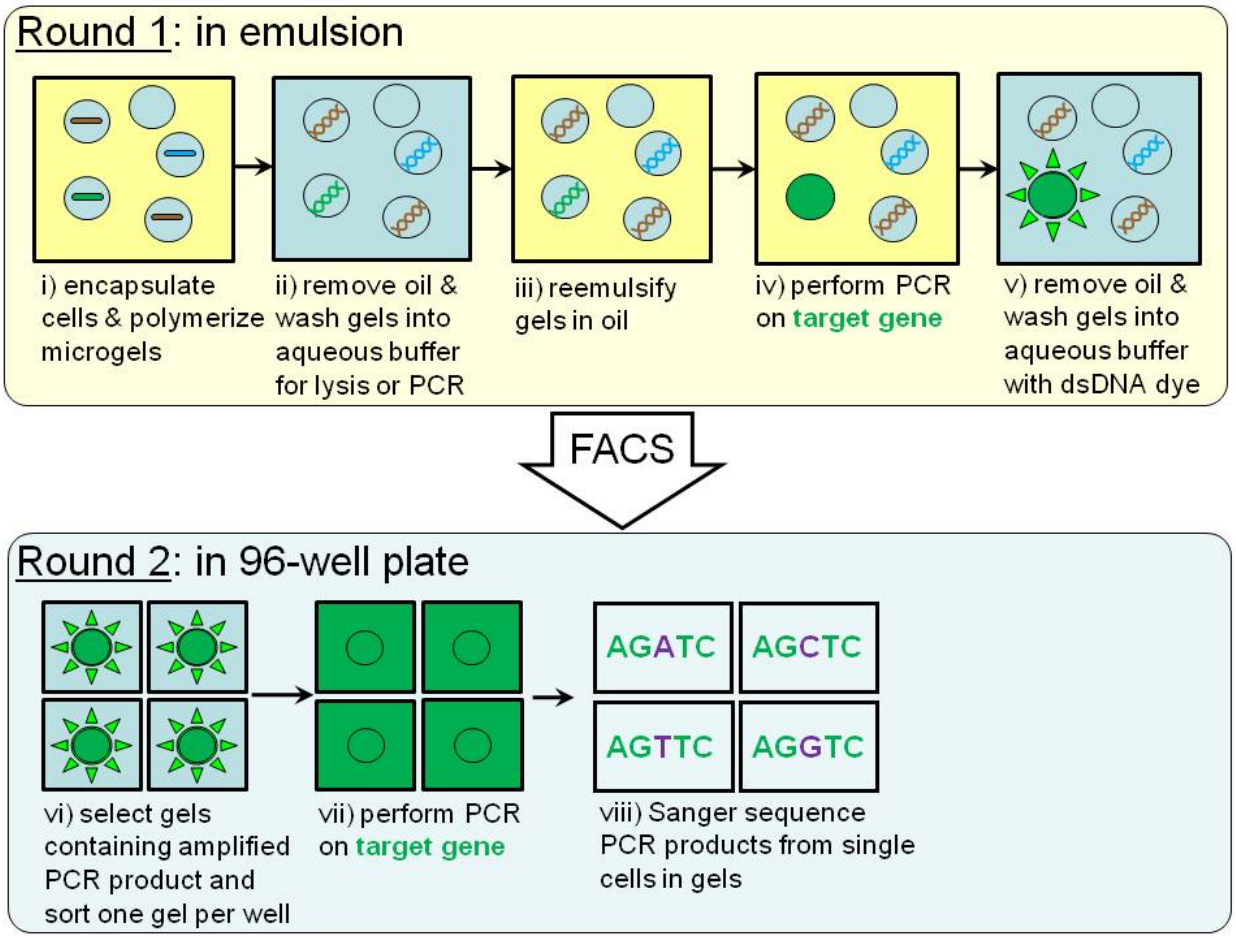


Figure 2.1: Single-cell gePCR workflow. Round 1: (i) We first use a microfluidic device to encapsulate single cells together with acrylamide in water-in-oil emulsion droplets. We then polymerize the droplets, so that the encapsulated cellular genomic DNA is trapped inside a microgel. (ii) We transfer the resulting microgels from the continuous oil phase to a continuous aqueous phase, to enable washing of water-soluble PCR reagents into the microgels. (iii) We reemulsify the microgels, so that they are again dispersed in a continuous oil phase, and then (iv) thermocycle. (v) After thermocycling is complete, we transfer the microgels once more to a continuous aqueous phase, to wash in DNA staining reagents. Round 2: (vi) We use flow cytometry (FCM) to perform fluorescence-activated cell sorting (FACS), and sort brightly stained single microgels containing PCR products into individual microtiter plate wells. (vii) We perform a second round of PCR to reamplify the target gene, producing a higher amount of product in the larger volume of the well. (viii) Finally, we use Sanger sequencing to analyze the PCR products from each microgel.

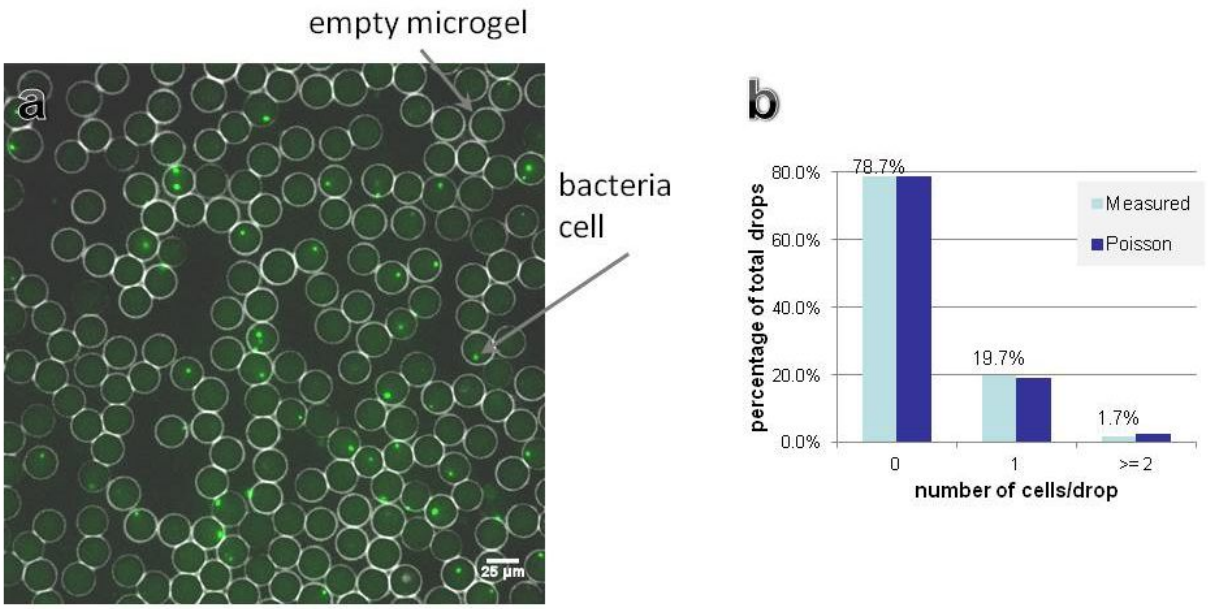


Figure 2.2: Workflow step (ii). Microgels remain monodisperse and retain encapsulated cells after polymerization and washing into aqueous staining buffer. Each cell is visible as a punctate dot (bright green) within an encapsulating gel (white outlines). Microgels that do not contain any cells are visible as dimly green circles inside white outlines; the faint fluorescence is due to nonspecific staining of polyacrylamide by SybrGreen. The cell encapsulation rate distribution of drops containing 0, 1, or at least 2 cells matches the Poisson distribution.

2.2.3 Gel polymerization

Polymerizing drops into microgels allows us to remove the continuous oil phase and wash the drop contents into any aqueous buffer, while still being able to keep the drops separated and resuspend them in oil. To produce microgels, we include monomer (4% acrylamide, Sigma-Aldrich #A4058), cross-linker (3% bis-acrylamide, Sigma-Aldrich #A9926), and catalyst (0.3% w/v ammonium persulfate, Sigma-Aldrich #A3678) in the inner aqueous phase, and accelerant (0.53% v/v tetramethylethylenediamine, Sigma-Aldrich #T7024) in the outer oil phase. Drops can then be polymerized after collection by incubating for 12 hours at 65°C. Drops remain stable throughout the polymerization process.

After drops polymerize to become microgels, we break the emulsion and wash the microgels into a continuous aqueous phase to enable diffusion of biological reagents into the microgels. Microgels retain their spherical shape and captured cells after the several steps required for washing: vortexing, centrifugation at 5-8 krcf for 30 s, and resuspension. We wash with a series of solvents: 1 wash in 20% (v/v) perfluorooctanol (VWR #AAB20156-18) in Novec^{TM} 7500 (3M), 2 washes in 1% (v/v) Span-80 (Sigma-Aldrich #85548) in hexane (VWR #BDH1129), and finally 3 washes in aqueous TET buffer (10 mM Tris pH 8.0; 10 mM ethylenediaminetetraacetic acid; 0.1% Triton X-100, NEB #T9284). Once the microgels are dispersed in the final aqueous phase, they swell within seconds, increasing in diameter by ≈40%. Tris facilitates imaging by limiting swelling. Microgels swell further when washed into pure water containing no salts, and shrink when washed into PCR mix, which demonstrates they are permeable to water. Cells encapsulated in microgels can be stored for at least 6 months at 4°C in TET buffer without significant DNA degradation, and for several years without any visible change in physical properties, although we have not measured pore size after long-term storage. They can also be stored in 10% (v/v) ethanol in TET buffer to ensure sterility.

See Appendix A: for a detailed protocol on cell encapsulation in polyacrylamide microgels.

2.2.4 Polymerase chain reaction on single cells inside emulsified microgels

To perform PCR on encapsulated cells, we additionally include 1 μM acrydited reverse primer for a gene of interest during the encapsulation step. The acrydite is incorporated into the polyacrylamide gel network during polymerization. The reverse primer is thus covalently bound to the gel, so that it will not diffuse out during washing in aqueous buffer.

Immediately before the PCR thermocycle, we wash 1 volume of microgels from TET storage buffer 2 times in 20 volumes of pure deionized water (VWR #45001-044). We then wash the

microgels 3 times in 1 volume of PCR reagents (1 μM forward primer; 1x standard Taq buffer with $MgCl_2$, NEB #M0273; 0.5% BSA, NEB #B9000; 250 μM dNTPs, NEB #N0447; 0.2% Tween-20, Sigma-Aldrich #P9416), and 1 time in ½ volume of PCR reagents with Taq enzyme (0.175 U/μL, NEB #M0273). Finally, we resuspend the microgels in ePCR oil[80] (mineral oil, Sigma-Aldrich #M5904; ABIL EM90, Evonik; 0.05% Triton X-100, Sigma-Aldrich #T9284) by centrifuging the microgels, removing the buffer, pipetting up and down 10 times in a 200 μL pipette tip with 4 volumes of ePCR oil, and agitating in a beadbeater for 3 minutes with 2 mm glass beads at 1 bead per volume. We aliquot 50 μL of the resulting suspension into several thin-walled PCR tubes, which can be thermocycled in a standard PCR machine.

The ePCR oil separates the microgels from each other, so that the forward primer, polymerase, dNTPs, and salts are able to diffuse freely within each microgel, but not between microgels. During the first thermocycle, the forward primer diffuses to the genomic target sequence, anneals to it, and generates a forward strand of PCR product (Figure 2.3a). The reverse primer is unable to diffuse due to its covalent bond to the polyacrylamide network. During the second thermocycle, the forward strand diffuses to the reverse primer, and generates a reverse strand of PCR product. From this point, PCR proceeds as it normally would in bulk solution; the only difference is that only the forward strands and primers diffuse, while the reverse ones are covalently bound and stationary.

Once PCR is complete, each microgel should contain either only primers, without any template or amplification products, or else a template along with products amplified from only that template. Compartmentalization during thermocycling is therefore critical, since the oil prevents forward strands of one sequence from diffusing into another microgel that contains a different sequence.

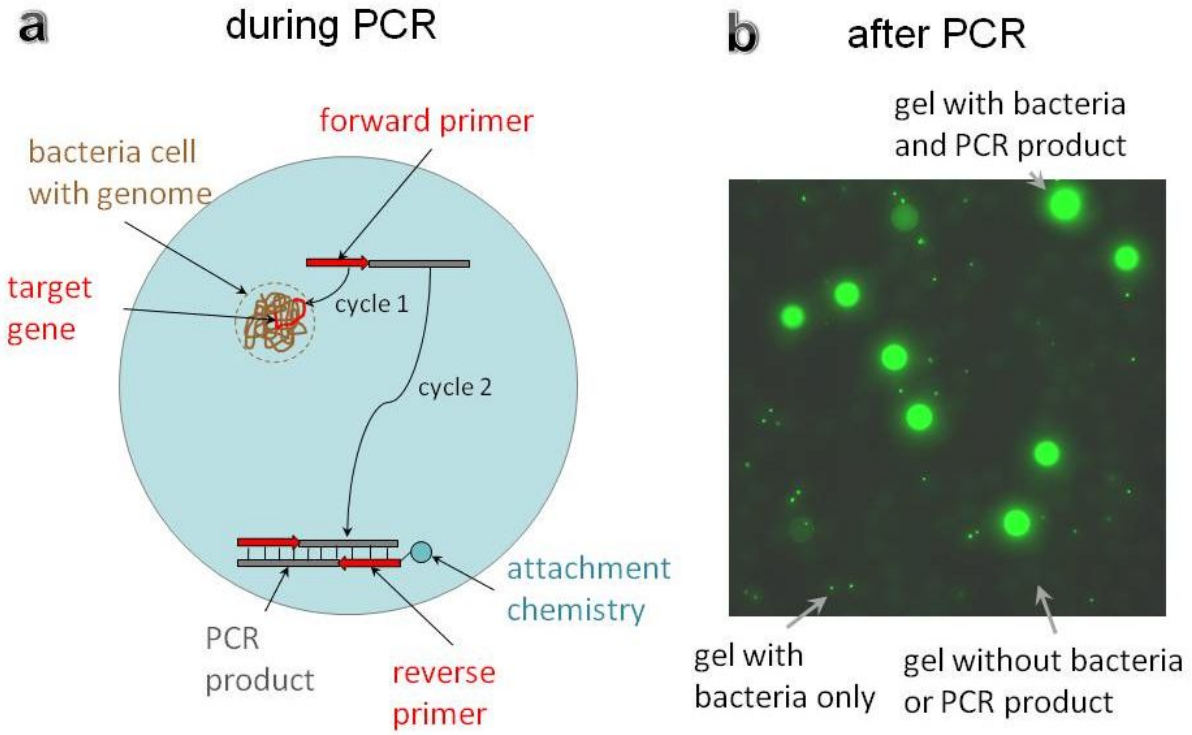


Figure 2.3: Workflow steps (iv) and (v). (a) PCR involves several molecular agents: a trapped bacterial genome, covalently bound reverse primer, freely diffusing forward primer, polymerase, dNTPs, and salts. The bacterial genome is physically trapped within the hydrogel network, due to the relatively small polyacrylamide pore size. The reverse primer is covalently bound to the polyacrylamide network by acrydite attachment chemistry, which occurs during acrylamide polymerization. During the first thermocycle, the forward primer diffuses to the genomic target sequence, anneals to it, and generates a forward strand of PCR product. The reverse primer is unable to diffuse due to its covalent bond to the polyacrylamide network via the acrydite molecule (small blue circle). During subsequent thermocycles, the forward strand diffuses to the reverse primer, and generates a reverse strand of PCR product. (b) After PCR, there are three populations of microgels: completely empty, containing unamplified bacteria only, and containing both bacteria and PCR product. In the last population, the original bacterial genome is not visible since it is overwhelmed by the strong signal from the PCR product.

Compartmentalization after thermocycling is also critical, so that we may sort out each microgel of desired amplification products, along with their source template contained in the same microgel. Since sorting requires further washing in aqueous buffer, the DNA in each

microgel must remain separated even after the oil is removed. The covalently bound reverse strands achieve this by annealing to the forward strands once the emulsion is cooled to 25°C.

Following the thermocycle, we wash the microgels back into aqueous staining buffer to determine how many and which gels contain amplified PCR product for the gene of interest. The washing process into aqueous buffer from mineral oil-based ePCR oil is the same as from fluorinated Novec 7500 oil, except that we omit the 1 wash in 20% (v/v) perfluorooctanol in Novec 7500. The microgels can now be stained with any desired intercalating dye or fluorescent *in situ* hybridization (FISH) probe.

See Appendix B: for a detailed protocol on performing PCR in emulsified microgels. See Appendix C for a detailed protocol on performing FISH in microgels.

Encapsulating individual cells in isolated compartments solves the problem of PCR bias. Since each compartment contains only a single template, there is no competition between templates. Detecting rare variants requires high throughput as well as eliminating PCR bias – if 0.001% of cells carry a gene, then to obtain 10 cells that carry the gene, we must assay 10^6 cells. Microfluidics enables us to encapsulate 10^6 cells in ~2 minutes, and we can easily manipulate 3 x 10^6 compartmentalized cells in each PCR reaction. We should thus be able to detect events with an occurrence of 0.001%.

PCR produces amplified DNA inside a fraction of microgels, which we detect by uniform fluorescence within individual microgels after washing them into staining buffer (Figure 2.3b). After a successful PCR, some microgels will be filled with PCR products covalently anchored by the reverse primer portion of the sequence, and these products will fluoresce when stained by an intercalating double-stranded DNA dye such as SybrGreen (Lonza #50513), or by a FISH probe. To confirm that the PCR products result from our deliberately encapsulated cells rather than

some contaminating DNA or cells, we also produce a negative control of microgels containing no cells, and measure the fraction of those microgels that are fluorescent.

The fraction of microgels that contain PCR products is determined by the cell encapsulation rate and the PCR efficiency. In the case of 100% efficiency, we would expect the final fraction of microgels containing PCR product to be equal to the initial fraction containing cells, multiplied by the fraction of cells carrying the PCR target. PCR efficiency is one limitation on throughput, so it is desirable to increase it.

2.2.5 Sorting out single gels containing single-cell amplicons

The first round of PCR in emulsified microgels does not produce enough DNA for Sanger sequencing, so another round of amplification is necessary. For 500 bp amplicons, the first round produces at most 1 μM PCR product in each microgel, assuming that the reaction runs to completion. For microgels that are ~2 pL in volume, this translates to ~10^6 molecules per microgel. Sanger sequencing generally requires 7 fmol or 10^9 molecules of DNA per sample, so we must amplify the DNA by another factor of 1000. We do this by sorting out single microgels containing PCR products into 96-well plates, and running separate PCR reactions in each well. Microgels that contain PCR products are those that contain a single cell carrying the gene of interest targeted by the PCR primers, and are therefore those that contain the PCR products that we wish to sequence.

We use flow cytometry (FCM) to sort out single microgels, by staining microgels and then gating on the population of bright microgels. Since SybrGreen has a high binding affinity for double-stranded DNA, and fluoresces brightly when bound, microgel fluorescence is an indicator of PCR product concentration.[81] To confirm that we are sorting out the correct population, we compare the percentage of bright and dark microgels detected by FCM to that detected by

fluorescence microscopy. We also sort several thousand microgels into a single imaging plate well and compare the FCM and fluorescence microscopy counts.

2.2.6 Polymerase chain reaction on pre-amplified templates from single cells

After the second round of PCR, each microplate well should contain DNA amplified from a single cell inside a single microgel, but some wells may contain no product or incorrectly sized products due to inaccuracy in the FCM sorting step. For example, a correctly gated fluorescent microgel may stick to the microplate wall or splash out of the well. To minimize this problem, we prefill microplate wells with PCR reagents before sorting, then immediately seal and thermocycle the plates after sorting is complete. Occasionally, microgels containing only primer-dimers or nonspecific PCR products from the first round of PCR may also fall above the fluorescence threshold for microgels containing the desired products; these false positives will produce only low molecular weight or nonspecific products after the second round of PCR. We use SybrGreen and plate spectroscopy to distinguish wells containing amplification products from empty wells. We distinguish nonspecific from correctly sized products by gel electrophoresis and then sequence correctly sized products.

2.2.7 Sequencing genes from single cells

We use Sanger sequencing on products that range in length from 120 bp to 1.3 kb, through services provided by Genewiz and Beckman Coulter.

2.3 Results

2.3.1 Validation of method using cultured cells

2.3.1.1 16S *from* Escherichia coli *cells*

We culture *E. coli* (strain K12 MG1655) cells and encapsulate them at a concentration of 0.23 cells/drop. During encapsulation, we include an acrydited reverse primer for *E. coli 16S* (Eco 16S-1387r, 5' /5Acryd/ GGG CGG TGT GTA CAA GGC 3', IDT). In the PCR reagent wash, we include a forward primer for *E. coli 16S* (Eco 16S-63f, 5' CAG GCC TAA CAC ATG CAA GTC 3', IDT). These primers target the 8f-1387r region of the 16S ribosomal gene, which all cells in the culture should carry.

PCR efficiency for amplifying *16S* from *E. coli* ranges from approximately 20% to 60%; for every 100 microgels that contain encapsulated cells, 20-60 microgels contain PCR products. We describe one particular experiment as an example: when stained with SybrGreen, 17% of *E. coli* microgels (62 out of 355) are fluorescent, while 8% of empty microgels (17 out of 223) are fluorescent. When stained with a FISH probe complementary to the covalently bound reverse strand of PCR product (EcoHyb62C TEX, 5' /5TEX615/ CAT ACA AAG AGA AGC GAC CTC G 3', IDT), 13% of *E. coli* microgels (39 out of 311) are fluorescent, while < 0.1% of empty microgels (0 out of 1000) are fluorescent. With an encapsulation rate of 0.23 cells/drop, 21% of microgels should contain 1 or more *E. coli* cells, so the PCR efficiency is 13% ÷ 21% = 61%.

Our efficiency of only 61% is likely due to the genomic DNA being partially inaccessible. We do not perform any special lysis steps prior to PCR, aside from the initial denaturation step at 95°C at the start of the thermocycle, so some residual cell walls, cell membranes, or DNA-binding proteins may restrict the accessibility of the DNA to the polymerase. Insufficient lysis is supported by two observations. First, our efficiency for the 16S gene in encapsulated *Propionibacterium acnes* cells is 0%. This indicates some dependence of the efficiency on the

species, and possibly peptidoglycan thickness, of the targeted cell. *E. coli* may be relatively responsive to lysis by heating, compared to *P. acnes*; this would produce the observed difference in efficiencies. Second, when we encapsulate naked *P. acnes* genomic DNA, following a spin column-based extraction step, our efficiency rises to 10%. Again, this indicates that the PCR efficiency increases when the DNA is in a more accessible form. It is therefore likely that inefficiency in the PCR reaction is primarily due to insufficient cell lysis, rather than the polyacrylamide microgel, emulsion oil, or surfactant. The efficiency could potentially be increased by pretreating the microgels with alkaline chemicals or enzymes to degrade macromolecules other than the genomic DNA.

The affinity of the sequence-specific probe for the PCR products contained by the microgels indicates that the PCR products match the *E. coli* 16S sequence. We thus confirm that we successfully target and amplify the *16S* gene from individual encapsulated *E. coli* cells inside emulsified microgels.

2.3.1.2 Elen_2529 *from* Eggerthella lenta *cells*

We culture *E. lenta* cells and encapsulate them at a concentration of 0.058 cells/drop. During encapsulation, we include an acrydited reverse primer for *E. lenta Elen_2529* (Ele cyt2-R658, 5' /5Acryd/ CGG CGC GCT TTT TCA GCG TT 3', IDT). In the PCR reagent wash, we include a forward primer for *E. lenta Elen_2529* (Ele cyt2-F425, 5' TGC GCT GGT CGC AAG GTC TG 3', IDT).

PCR efficiency for amplifying *Elen_2529* from *E. lenta* is 43%. We describe one particular experiment as an example: when stained with SybrGreen, 2.5% of *E. lenta* microgels (4 out of 161) are fluorescent, while < 0.1% of empty microgels (0 out of 1000) are fluorescent. With an

encapsulation rate of 0.058 cells/drop, 5.8% of microgels should contain 1 or more *E. lenta* cells, so the PCR efficiency is 2.5% ÷ 5.8% = 43%.

We sort SybrGreen-stained single microgels into a microtiter plate for a second round of PCR before sequencing. Gel electrophoresis shows that 11 out of 16 wells contain PCR product. Wells lacking PCR product may result from inaccurate microgel sorting into those wells, or from PCR failure despite accurate sorting. We are able to obtain high quality Sanger sequencing reads from all 11 wells that contain PCR product, while the other 5 wells produce poor quality reads. The efficiency for the three steps of sorting, PCR, and sequencing is 52%: in total, we attempt to sequence 60 wells, and obtain 31 high quality sequences, all of which align with *Elen_2529*. The successful alignment confirms that we are able to target and amplify a functional gene from individual encapsulated *E. lenta* cells inside emulsified microgels.

In subsequent experiments, to minimize Sanger sequencing runs that are wasted on wells lacking PCR product, we use plate spectrophotometry to select only wells that do contain amplified DNA. We choose this method over gel electrophoresis because it is more efficient and gives few (~10%) false positives, which are generally due to high levels of primer-dimers. Since we only target a single gene, there is no need to cut out bands for gel extraction, and PCR purification is sufficient cleanup for Sanger sequencing. Within these wells that contain amplified DNA from the second round of PCR, the rate of high-quality reads is 85%.

2.3.2 16S *from gut bacteria*

We isolate gut bacteria cells from two fecal samples, F4 and F5, and encapsulate them at a concentration of 0.13 and 0.04 cells/drop, respectively. During encapsulation, we include an acrydited reverse universal primer for bacterial *16S* (16S-907r universal, 5’ /5Acryd/ CCG TCA

ATT CCT TTR AGT TT 3', IDT). In the PCR reagent wash, we include a forward universal primer for bacterial *16S* (16S-8f universal, 5' AGA GTT TGA TCC TGG CTC AG 3', IDT).

PCR efficiency for amplifying *16S* from gut bacteria is 2% for F4 and < 9% for F5. When stained with SybrGreen, 0.2% of F4 microgels (1 out of 500) are fluorescent, while < 0.2% of empty microgels (0 out of 500) are fluorescent. With an encapsulation rate of 0.13 cells/drop, 12% of microgels should contain 1 or more F4 cells, so the PCR efficiency is 0.2% ÷ 12% = 2%.

We stain microgels with SybrGreen before sorting, a second round of PCR, and sequencing. The percentage of bright versus dark microgels detected by FCM closely matches the percentage that we observe by fluorescence microscopy (Figure 2.4, top). After sorting out the population of bright microgels, we again use both techniques to confirm that our sort is successful (Figure 2.4, bottom).

After the second round of PCR, each microplate well should contain DNA amplified from a single cell inside a single microgel, but in general, 30-90% of wells may contain no product or incorrectly sized products due to inaccuracy in the PCR or sorting steps. Nonspecific amplification during PCR can produce bright gels that contain DNA but lack the desired amplicon of the correct size. This source of error can be addressed by optimizing PCR conditions. During sorting, the flow cytometer may incorrectly sort autofluorescent dust or other particles that fall within the forward scatter and side scatter gates, or particularly bright cells or clumps of cells in microgels lacking any PCR product. These represent false positives during sorting, but would produce a well with no product after the second round of PCR. This source of error can be addressed by gating more tightly within the fluorescent population. However, if the objective of a particular application is not quantitation of gene distribution within a diverse cell

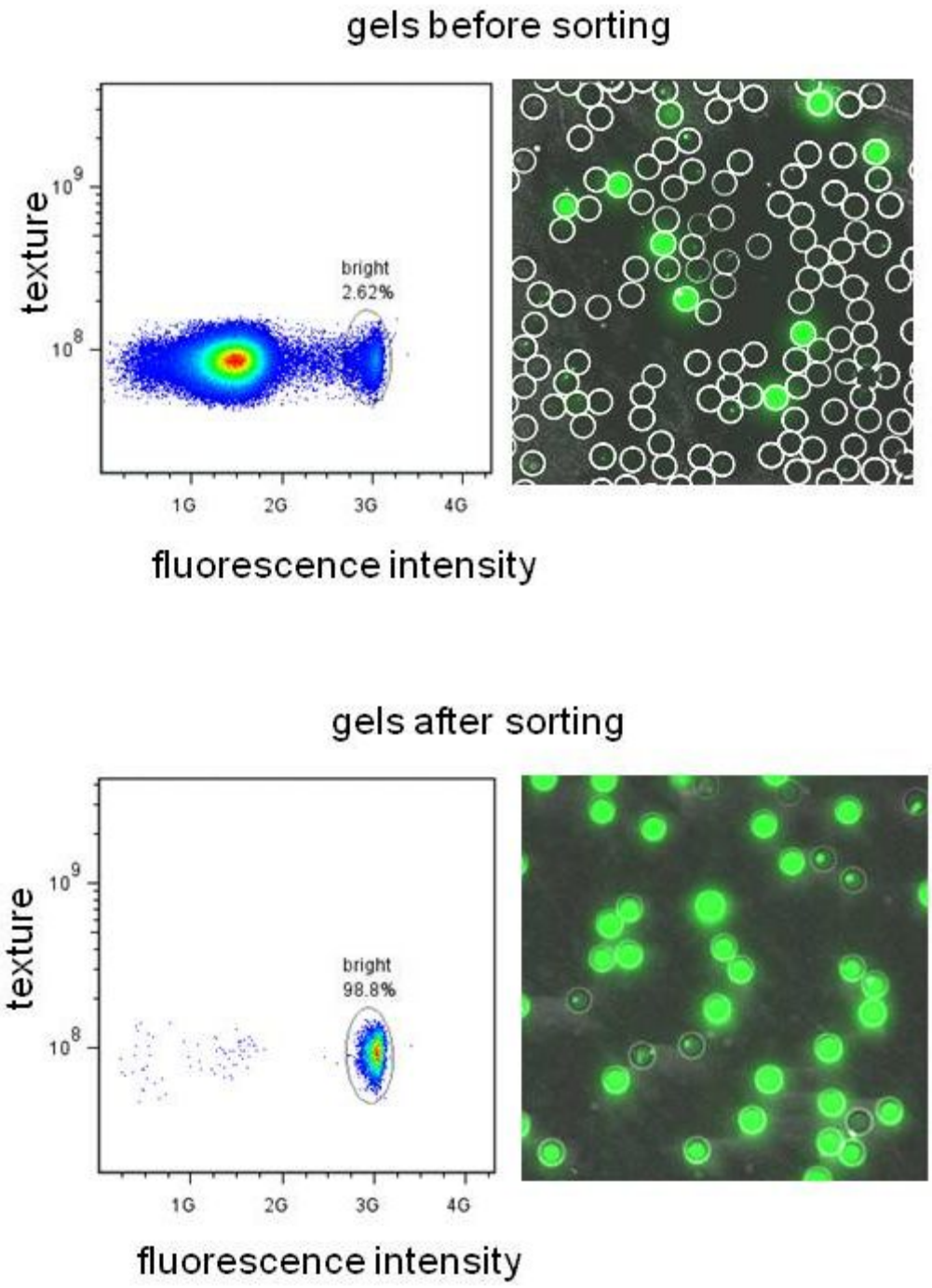


Figure 2.4: Workflow step (vi). We start from a mixed population of microgels containing different amounts of PCR product. Using both flow cytometry and fluorescence microscopy, we confirm successful selection of bright microgels filled with PCR product.

population, but simply detection and sequencing of a rare gene of interest, such false positives would not negatively impact the final result.

For F4 microgels, gel electrophoresis shows that 4 out of 5 wells contain PCR product, while spectrophotometer fluorescence measurement shows 53 out of 92 wells. We obtain high quality Sanger sequencing reads with QV20+ > 500 from 51 of these 53 wells. These reads align to a diverse range of operational taxonomic units (OTUs) including members of the Firmicutes, Bacteroidetes, and Verrucomicrobia phyla (Figure 2.5, top right). For F5 microgels, gel electrophoresis shows that 3 out of 5 wells contain PCR product, while spectrophotometer fluorescence measurement shows 45 out of 92 wells. We obtain high quality reads from 20 of these 45 wells, which align to members of the Firmicutes, Bacteroidetes, and Proteobacteria phyla (Figure 2.5, bottom right). An additional 5 wells align successfully despite having lower quality reads. Our efficiency for the three steps of sorting, PCR, and sequencing is 55% for F4 microgels, and 30% for F5 microgels. The successful alignment for these fractions of microgels confirms that we are able to target and amplify the *16S* gene from individual encapsulated gut bacteria inside emulsified microgels.

The phylogenetic distribution of the single-cell reads we obtain using gel emulsion PCR is roughly consistent with what we obtain through conventional bulk methods (Figure 2.5, left). Next-generation sequencing on bulk extracted DNA from the F4 sample shows a phylogenetic distribution very similar to the results from 93 single-cell microgels. A notable difference is that the "Other" category is composed primarily of Actinobacteria for the bulk analysis, and Proteobacteria and Chlorobi for the single-cell gels. We know that our single-cell method is biased against at least one Actinobacteria species, *P. acnes* (see section 2.3.1.1), so these results are consistent. However, although this bias limits the accuracy of our phylogenetic distribution, it

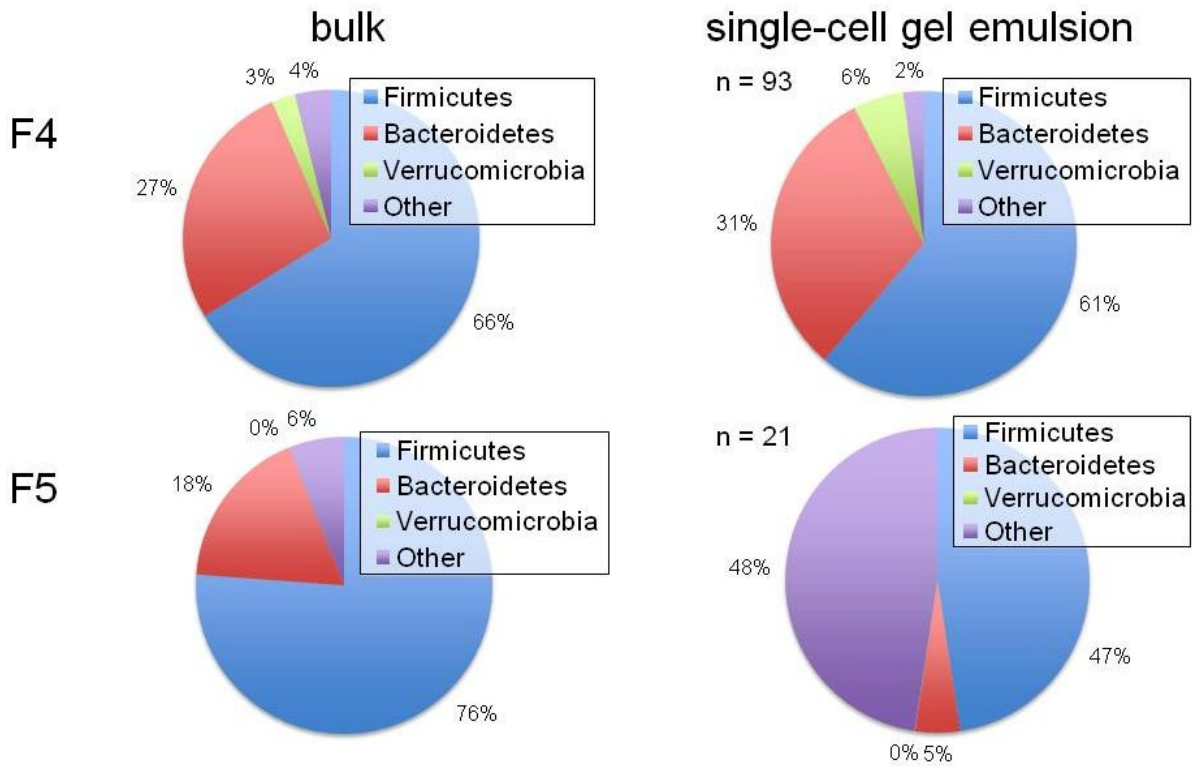


Figure 2.5: Single-cell reads (right) from 2 fecal samples, F4 (top) and F5 (bottom) align to a diverse range of OTUs, including members of the Firmicutes, Bacteroidetes, and Verrucomicrobia phyla, as well as Proteobacteria and Chlorobi ("Other" category). The phyla represented in single-cell results are roughly consistent with those from conventional bulk methods (left), with the exception that the "Other" category primarily represents Actinobacteria for bulk sequences, but is composed of Proteobacteria and Chlorobi for the single cells. The two methods match more closely for sample F4, from which we obtained the much higher number of 93 single-cell reads; the low count of n = 21 may be partly responsible for the discrepancy in sample F5.

could also be a way to selectively exclude certain abundant phyla that would normally overwhelm the signal from rare species. The F5 sample shows a greater difference between the bulk and single-cell phylogenetic distributions, which may be partly due to the lower number of 21 single-cell microgels that yielded sequences. The "Other" category for the single-cell method shows the same bias against Actinobacteria and toward Proteobacteria.

Our method shows promise for accessing the genetics of novel, uncultured bacteria, which may represent rare cells within the gut population that would produce undetectably low signal in a bulk analysis. Twelve of the 76 sequences we obtain align to novel, uncultured bacteria. One F5 sequence aligns to unclassified *Lachnospiraceae*, 2 F5 sequences and 6 F4 sequences align to unclassified *Ruminococcaceae*, and 3 F4 sequences align to unclassified bacteria.

2.3.3 BT_4738 from gut bacteria

Using the same two fecal samples, F4 and F5, we isolate gut bacteria cells and encapsulate them with an acrydited reverse universal primer for *Bacteroides thetaiotaomicron BT_4738* (BT4738.R, 5' /5Acryd/ GCC GAC GAT CAG TTC GTT TTC 3', IDT). In the PCR reagent wash, we include a forward universal primer for *B. theta BT_4738* (BT4738.F, 5' TCG AAG CGT TGG CAG AAG AA 3', IDT). This gene encodes a pyruvate-formate lyase, which is relevant for carbohydrate metabolism[79, 82].

We again obtain high quality Sanger sequencing reads for single-cell microgels (Figure 2.6). In particular, our read quality demonstrates that we successfully identify genetic variation at the single-cell level within a diverse gut microbe population, through a comparison with bulk extracted gDNA from sample F5. The bulk gDNA carries a synonymous single-nucleotide polymorphism (SNP) of CAA → CAG, when compared to the reference sequence (NCBI Gene ID 1074904). However, the gDNA represents an average over all cells, and as is visible from the chromatogram, there is a secondary peak for an A nucleotide at this position. Our single-cell reads deconvolute this average and clearly show that the SNP visible in the gDNA in fact represents a combination of single cells that do (5 cells) or do not (12 cells) carry the SNP. Furthermore, other SNPs that do not even appear in the basecalls for the bulk gDNA are

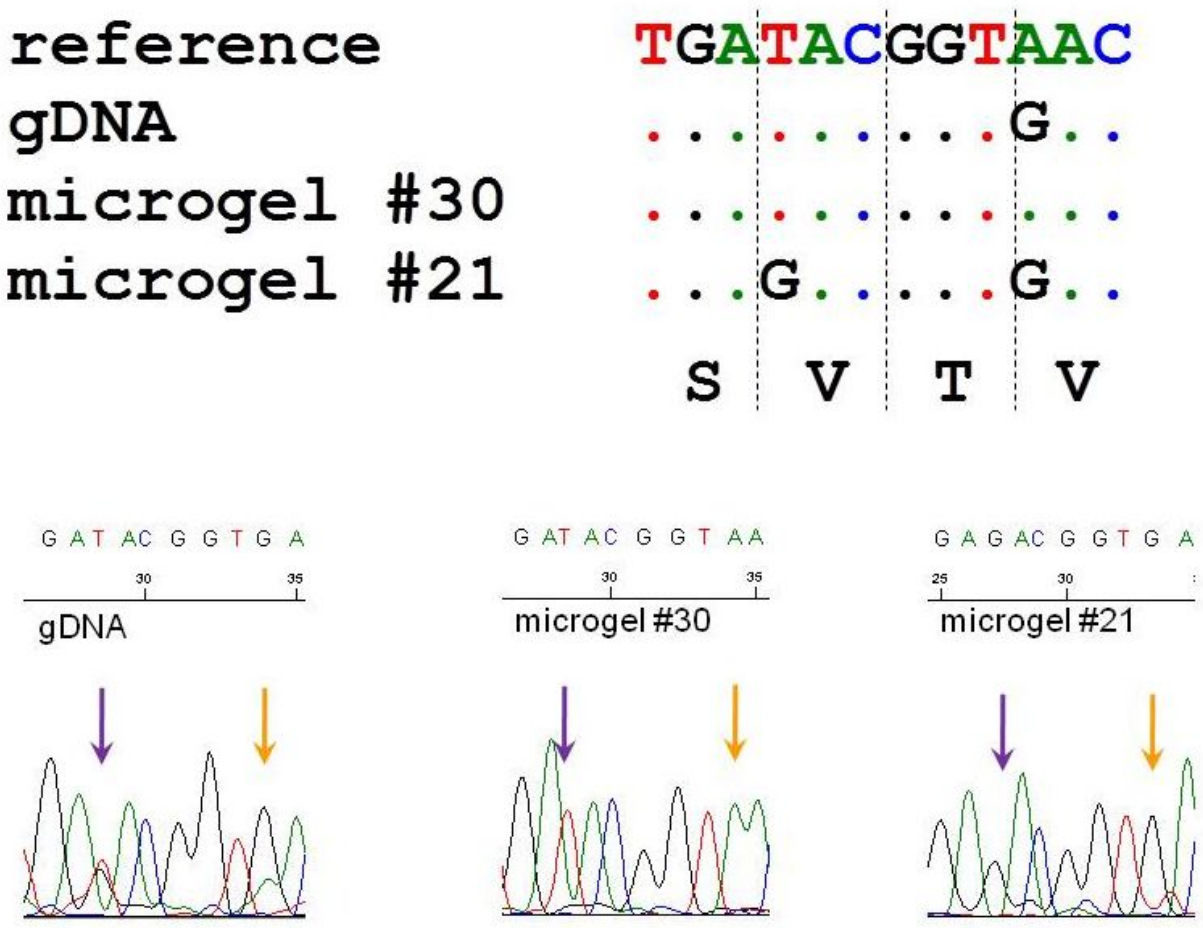


Figure 2.6: Representative sequences from single-cell microgels after PCR targeting *BT_4738* . Line 1: Reference sequence from NCBI Gene database (NC_004663.1), line 2: bulk extracted gDNA, line 3: sorted microgel #30, line 4: sorted microgel #21. The bulk gDNA carries a SNP of CAA → CAG (left chromatogram, orange arrow), with a secondary peak for the wild-type A nucleotide. Our single-cell reads deconvolute this average into clear single peaks for either A (middle chromatogram, orange arrow) or G (right chromatogram, orange arrow) in different individual cells. This effect is also visible for a second SNP (purple arrow). Both SNPs are synonymous, and their reverse complements code for valine.

detectable at the single-cell level. Thus, our method can be used to explore the diversity of particular genes within a complex population such as the gut microbiome.

2.3.4 Lysis in microgels

In contrast to liquid emulsion droplets, microgels are compatible with a wide range of lysis protocols upstream of PCR. Chemical or enzymatic lysis reagents can easily be added to

homogeneous liquid droplets using picoinjection[17] or electrocoalescence[18], but such reagents are often incompatible with PCR, and must be neutralized or diluted before thermocycling. Our microgels facilitate this step; following encapsulation and polymerization, we simply diffuse lysis reagents into the microgels, and then easily dilute them to negligible concentrations by washing in pure water, yet without losing the gDNA trapped in the hydrogel matrix. This allows us to use harsh reagents such as strong acids and bases, as well as enzymes such as lysozyme, proteinase K, and mutanolysin, without searching for a chemical neutralizing agent or increasing the size of each droplet through dilution.

Microgels also allow for greater flexibility in choosing mechanical lysis methods. For example, microgels are amenable to freeze/thaw cycles, which makes them particularly useful for encapsulating environmental bacteria that are often resistant to enzymatic lysis. Microgels thus facilitate investigation of novel bacteria that do not belong to the handful of tractable culture lines.

Microgels are additionally compatible with heat lysis. Microgels can be incubated in boiling water for up to 15 minutes while still maintaining structural integrity; liquid droplets will coalesce under these conditions. Alternatively, microgels can be used in a PCR thermocycle with an extended initial denaturation step of 3 minutes at 94°C, similarly to liquid droplets. We choose this method for encapsulated cells from cultures and from human gut samples. PCR efficiency varies with cell species and source (Fig. x), which is likely because brief heating is insufficient for lysing bacteria cells of all types. Nevertheless, we are able to use emulsified microgels to sufficiently amplify even single gut microbes for downstream sequencing, which has not previously been achieved.

2.4 Conclusion

We first microfluidically encapsulate cells in emulsified hydrogels, then wash in PCR reagents and perform single-cell emulsion PCR. The hydrogel network allows us to separate genomes and PCR products originating from different cells, even while diffusing reagents for a series of different reactions in and out of microgels suspended in a continuous aqueous phase. We next sort out cells of interest along with the PCR products generated from each of those cells, and finally sequence the PCR products. We thus use emulsion PCR and microfluidically generated hydrogels to enable sorting and sequencing of single human gut bacteria cells.

Single-cell genetic analysis is useful for analyzing biological samples that contain mixed populations of genetically distinct individuals. Such samples may be drawn from a broad variety of medically relevant sources, including the microbiota of the human gut[83-85] and other tissues,[86] sewage wastewater,[64, 87] soil,[88] and mutation-laden tumors.[89] Characterizing heterogeneous populations at the single-cell level can reveal details such as the frequency distribution of different species, or which variants of a particular gene are present.

Conventional metagenomic techniques are powerful but poorly suited for obtaining information about single cells. For example, targeted amplicon studies use gene-specific primers to target a specific gene for amplification.[73, 85] These studies can determine which different sequences of that gene are present in the population, and how many are present of each unique sequence, but PCR bias limits how accurately the distribution of different sequences can be measured.[59] Additionally, very rare sequences will be overwhelmed by more common ones, and will thus be undetected.[24] As a second example, shotgun sequencing uses random primers to sequence short fragments of all DNA present in a sample.[90] The resulting collection of gene fragments can in principle reveal all variants of all genes present, but not which cell each

fragment originated from, since fragments from multiple cells are mixed together. Furthermore, in practice, whole-genome assembly from environmental samples presents a significant computational challenge,[91, 92] and very rare variants are again difficult to detect. To maximize information obtained from a population, it is necessary to isolate individual cells before any DNA shearing or fragmentation.

Chapter 3: Two-stage gel encapsulation for single-cell MDA

Beyond sequencing specific genes from single cells as described in the previous chapter, whole-genome sequencing would yield even greater insight into the components of a diverse environmental population. While gePCR could enable the identification of gut bacteria species carrying *B. theta BT_4738*, discovering all genes carried by those bacteria of interest would not be possible. Instead, a similar approach using whole-genome sequencing (WGS) is necessary. We demonstrate that microfluidically encapsulated *E. coli* can be used as templates for multiple displacement amplification (MDA). Since MDA requires physical space on a microscopic level that is unavailable within a typical gel network, we use a material with a temperature-dependent gel-liquid phase transition to reversibly switch between gel-phase and liquid-phase emulsion droplets. Finally, we propose a workflow for highly selective WGS using this method.

3.1 Polyacrylamide hydrogel network inhibits MDA

MDA uses the enzyme phi29 to copy DNA, with random hexamers as primers. Phi29 is suitable for whole-genome amplification due to its high processivity and fidelity. Since it is able to synthesize a complementary strand of DNA from a template while simultaneously displacing any previously hybridized complementary strand, it does not require a three-step thermocycle for denaturation, hybridization, and extension. As long as nucleotides and primers are available and the reaction is held at 30°C, phi29 will continuously synthesize DNA that is complementary to whatever template is present in the reaction chamber. This significantly simplifies the amplification process.

However, the strand displacement activity of phi29 poses a major challenge to substituting it for PCR polymerases in our gePCR system. MDA produces amplicons in the form of long (tens

of kilobases), highly branched DNA polymers, which form a dense network. This is in contrast to PCR, which produces short (hundreds of bases up to a few kilobases) DNA fragments. These do not exhibit tangling because the persistence length of DNA is roughly 1700 bp. In our PCR experiments described in the previous chapter, amplicons range from 300 – 1400 bp, and are thus linear polymers whose diffusion and synthesis are not inhibited by the polyacrylamide network. Hyperbranched MDA amplicons, on the other hand, are obstructed by a hydrogel with ~150 nm pore size.

To confirm our prediction that MDA cannot proceed within a hydrogel, we use a protocol similar to gePCR. We first encapsulate *E. lenta* cells in polyacrylamide microgels (Appendix A:), then wash MDA reagents into the microgels (according to kit protocol from Epicentre RepliPHI Phi29 reagent set, catalog #RH031110), before suspending them in emulsion oil and incubating them for 6 hours at 30°C to perform emulsion MDA (eMDA). After incubation, we wash the microgels back into aqueous buffer for staining with SybrGreen.

A successful MDA reaction should produce DNA uniformly distributed throughout the reaction compartment, and thus uniformly bright microgels. Instead, we observe patchy fluorescence on the microgel surfaces, which is consistent with MDA occurring primarily in the thin liquid film between the microgel surface and the interface with the surrounding emulsion oil, and producing patches of hyperbranched DNA stuck to the microgels (Figure 3.1). It may seem surprising that MDA could occur there, since the genomic template should be unable to diffuse to the hydrogel surface, but there may be free DNA fragments or early-stage amplicons with less branching that can diffuse through the hydrogel. Also, some microgels contain bacteria that already happened to be at the surface when the acrylamide was originally polymerized. We

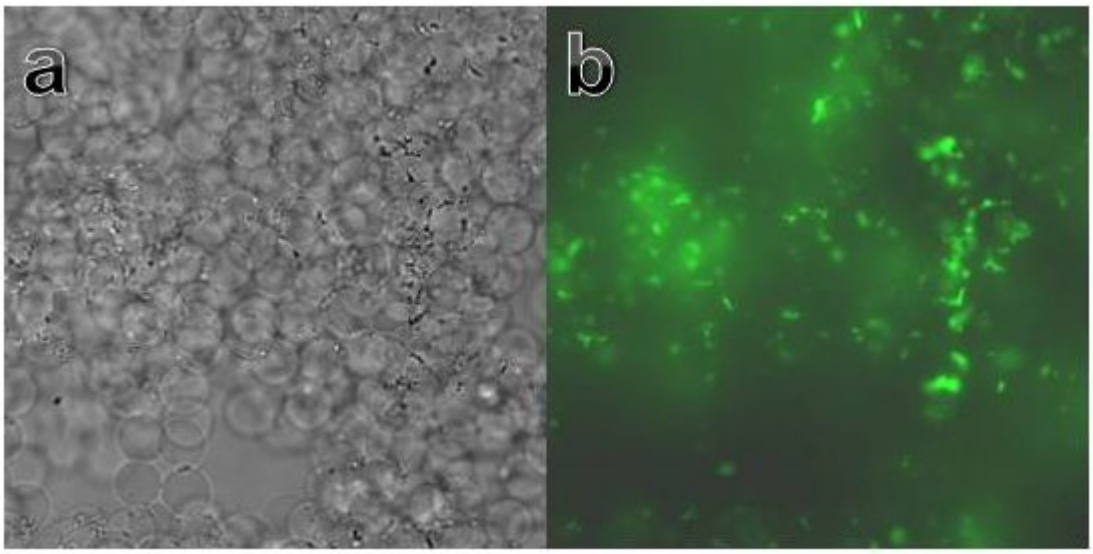


Figure 3.1: MDA produces clumps of DNA on microgels. (a) Bright-field, (b) fluorescent microscopy. Our established workflow for gePCR cannot use MDA as an alternative amplification method. While PCR produces small linear amplicons that are evenly distributed within individually separated and sortable microgels (see **Figure 2.3**), MDA produces tangled clumps of DNA on the microgel surfaces. This may lead to crosstalk between amplicons derived from different individual genomes, and also causes the microgels to stick together, so that they cannot be individually sorted. Images courtesy of Linas Mažutis.

also observe extensive clumping between microgels, which could be caused by interactions between hyperbranched DNA patches on different microgels. Both patchy fluorescence and clumping would interfere with downstream sorting; patchy fluorescence is difficult to distinguish from noise caused by autofluorescent debris, and clumps can produce crosstalk, exclude size-based sorting, and clog both conventional and microfluidic sorters. We therefore need to solve the problem of how to allow enzymatic reactions like MDA that require a liquid phase, while still enabling downstream heterogeneous assays like washing.

3.2 Method: Agarose shell with dissolvable polyacrylamide core

We use two stages of encapsulation with different gel materials: first polyacrylamide for its durability and easy handling, and second agarose for its thermo-reversibility. In the second stage,

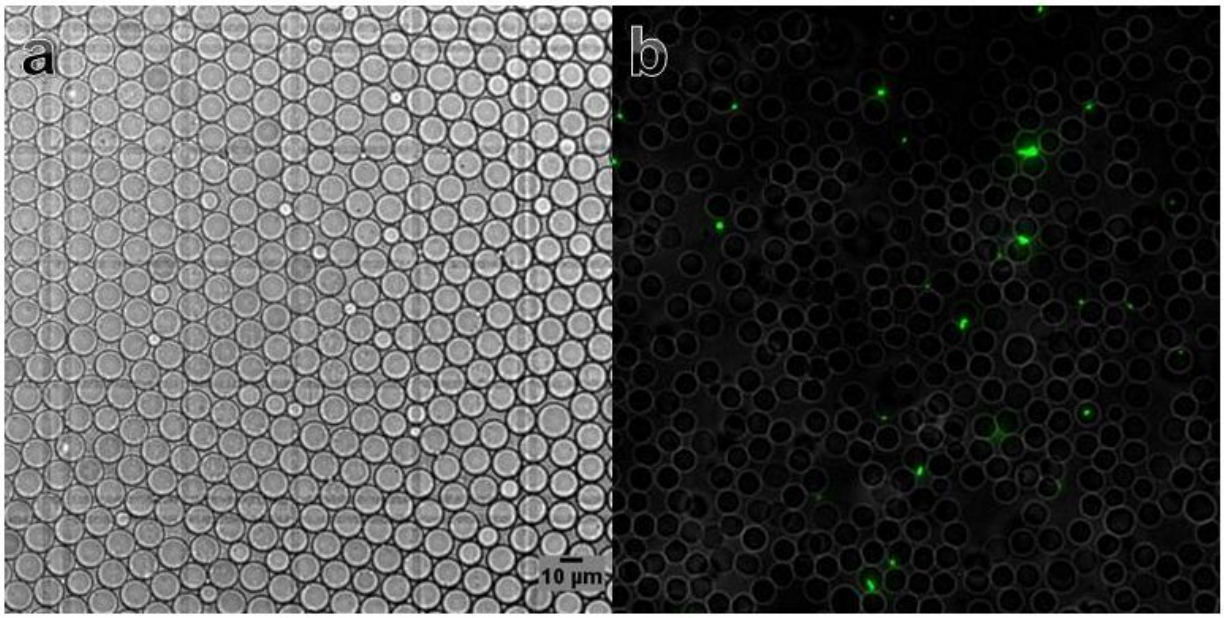


Figure 3.2: Successful encapsulation of E. coli XL1 cells in polyacrylamide microgels. (a) Liquid acrylamide-in-oil droplets, immediately after collection. (b) Microgels after polymerization, washing into TET aqueous buffer, and SybrGreen staining.

we dissolve the polyacrylamide crosslinker while adding agarose, to convert the polyacrylamide microgels to agarose microgels. Since agarose is a gel at low temperatures and a liquid at high temperatures, this allows us to reversibly convert our microfluidic compartments between liquid and gel phases by simply altering the temperature of our sample.

We first encapsulate *E. coli* XL1 bacteria in polyacrylamide microgels using a protocol similar to that for gePCR, with the key difference that we include bisacryloylcystamine (BAC, Sigma) as a crosslinker instead of bisacrylamide (Figure 3.2). We use soft lithography to fabricate PDMS-glass devices with a junction width of 10 µm and height of 5 µm, and use flow rates of Q_{inner} = 100 µL/hr and Q_{outer} = 150 µL/hr. We choose these dimensions to produce microgels that are sufficiently small for use in downstream steps. After polymerization, we break the emulsion using PFO and wash the microgels into a continuous aqueous phase of TET buffer, then lyse the encapsulated cells by heating at 95°C for 10 minutes. This lysis step must be

performed before the second stage of encapsulation in agarose, since agarose will be liquid at high temperatures and thus unusable for compartmentalizing individual cells in aqueous buffer. Indeed, we enable heat lysis by using polyacrylamide for the first stage of encapsulation.

During the second encapsulation step, which we perform immediately before the MDA reaction, we coflow the polyacrylamide microgels with a heated 1.6% w/v agarose solution containing MDA reagents ("positive sample", Figure 3.3), including dithiothreitol (DTT, from Epicentre RepliPHI Phi29 reagent set, catalog #RH031110). Since DTT breaks the disulfide bonds between the BAC molecules that crosslink the polyacrylamide, it dissolves the gel network, thus converting the microgels to a liquid phase and releasing the lysed bacteria. This eliminates physical interference with the highly processive MDA reaction. There is also no chemical interference or free radical formation that could inhibit enzymatic activity, since the polyacrylamide is still in polymer form and not monomer form, despite no longer being structured as a network. As a negative control, we coflow water with the agarose and MDA reagents, producing agarose microgels that do not contain any polyacrylamide or DNA.

To maintain the separation between DNA originating from different cells, we flow oil through a second coflow junction, forming liquid droplets of agarose and MDA reagents-in-oil loaded with polyacrylamide microgels. The kinetics of the polyacrylamide dissolution are on the order of seconds and much slower than the microfluidic encapsulation, which takes place on the order of milliseconds. Thus, by the time the individual bacteria cells are no longer compartmentalized by the gel network, they are already compartmentalized by the oil.

The double coflow device has a junction width and height of 17 μm, with a 14 μm constriction at the second junction (adapted from a design provided by Assaf Rotem). It produces 33 μm agarose droplets using flow rates of Q_{inner} = 111 μL/hr, Q_{middle} = 155 μL/hr, and Q_{outer} =

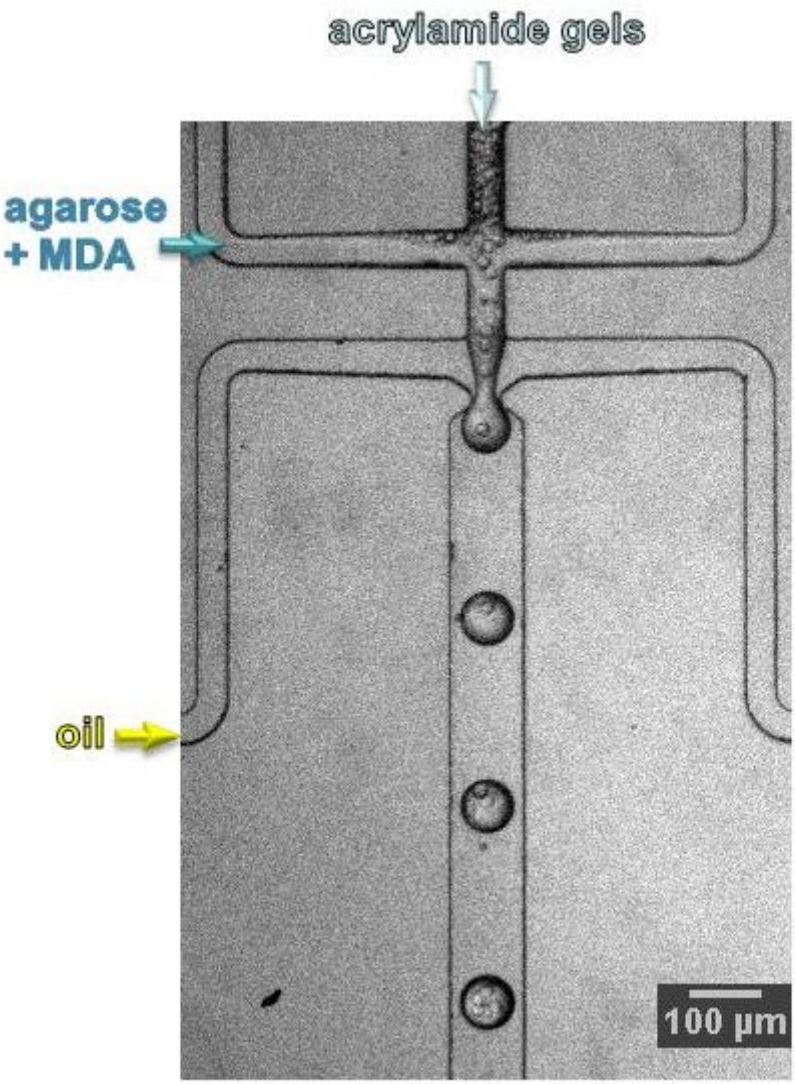


Figure 3.3: Double coflow dropmaker. At the first junction, polyacrylamide microgels coflow with a mixture of agarose and MDA reagents. Milliseconds later, oil enters from the side channels at the second junction, creating liquid agarose drops with encapsulated polyacrylamide microgels. As the DTT included in the MDA reagents diffuses through the polyacrylamide microgels, it dissolves them. Some dissolution is visible at the first junction, due to a small amount of microgels backflowing into the agarose channel and sticking to the channel walls. This introduces a negligible amount of crosstalk between drops. For illustration purposes, this device is larger (50 um coflow, 40 um constriction) than the ones used in later experiments, and contains larger acrylamide gels.

1555 µL/hr, with a final concentration of 0.9% w/v agarose. We choose these junction dimensions to produce the smallest agarose droplets possible without experiencing any

perturbation from passage of the polyacrylamide microgels through the junction. As discussed in Section 1.2, smaller droplet size enables higher throughput and lower reagent volumes; furthermore, microgels smaller than 30 μm offer the convenience that they can be sorted using conventional flow cytometry in addition to microfluidics.

Following collection of the emulsion, we incubate it at 30°C for 6-18 hours to allow the MDA reaction to proceed, then leave it on ice for a few minutes to stop the MDA reaction, and store it at 4°C. Lowering the temperature to 0°C also gels the agarose, which minimizes coalescence. We assess DNA amplification in the post-MDA agarose microgels by using fluorescent microscopy. Since we include SybrGreen with the MDA reagents during the second encapsulation, microgels containing successfully amplified product are bright, while unamplified microgels are dark. Finally, we break the emulsion using PFO, and wash the microgels into a continuous aqueous phase of TET buffer, to enable diffusion of biological reagents into the microgels during downstream assays.

As an example of a heterogeneous aqueous-phase assay, we perform FISH on the microgels, using a fluorophore-labeled DNA probe (5' GCG CCT ATT AAT GAC AAC AA Cy5 3', IDT) that targets the tetracycline resistance tetB gene carried by *E. coli* XL1. We use a protocol similar to that for gePCR, except without any heating above room temperature, to prevent melting the agarose. To partially compensate for any decrease in annealing efficiency between the probe and MDA product caused by the lower temperature, we include the probe in the polony annealing buffer washes, before incubating for 15 minutes at room temperature and then washing away unbound probe with 1E buffer.

3.3 Results

3.3.1 Two-stage polyacrylamide-agarose encapsulation

We successfully encapsulate *E. coli* XL1 cells in polyacrylamide microgels with a loading rate of 0.06 cells/drop (Figure 3.2). The droplets are initially 13 μm in diameter as collected, and range from 9-14 μm in diameter after polymerization and washing into TET aqueous buffer. To our knowledge, these are smaller than any previously reported polyacrylamide microgels microfluidically produced with this degree of monodispersity.

We observe an increase in both polydispersity and anisotropy for these microgels, relative to the larger microgels (15-20 μm) used for gePCR. This could be due to the smaller microgel diameter. We generally find that acrylamide polymerization becomes less reliable and requires higher concentrations of catalyst and accelerant, as gel size decreases from the 1 cm typically used for conventional sequencing to the 63 μm used for inDrop and smaller for other applications. Inhomogeneous polymerization in these very small microgels could lead to different degrees of swelling in aqueous buffer, and thus irregularities in size and shape. Alternatively, this may instead be due to some effect from the BAC crosslinker compared to bisacrylamide, which we did not investigate. Regardless, since the utility of microfluidics in this application is not to enable precise quantitation, but rather high-throughput selection of rare cells of interest, followed by WGS of those few selected cells, the degree of droplet size irregularity that we observe does not significantly impact our current results.

We successfully produce 20-30 μm agarose microgels. We choose the polyacrylamide-in-agarose loading rate based on the bacteria-in-polyacrylamide loading rate, to achieve a bacteria-in-agarose loading rate of no more than 0.3 cells/drop; it is only important to avoid loading multiple bacteria per agarose microgel, as loading multiple empty polyacrylamide microgels inside the same agarose microgel does not lead to any crosstalk.

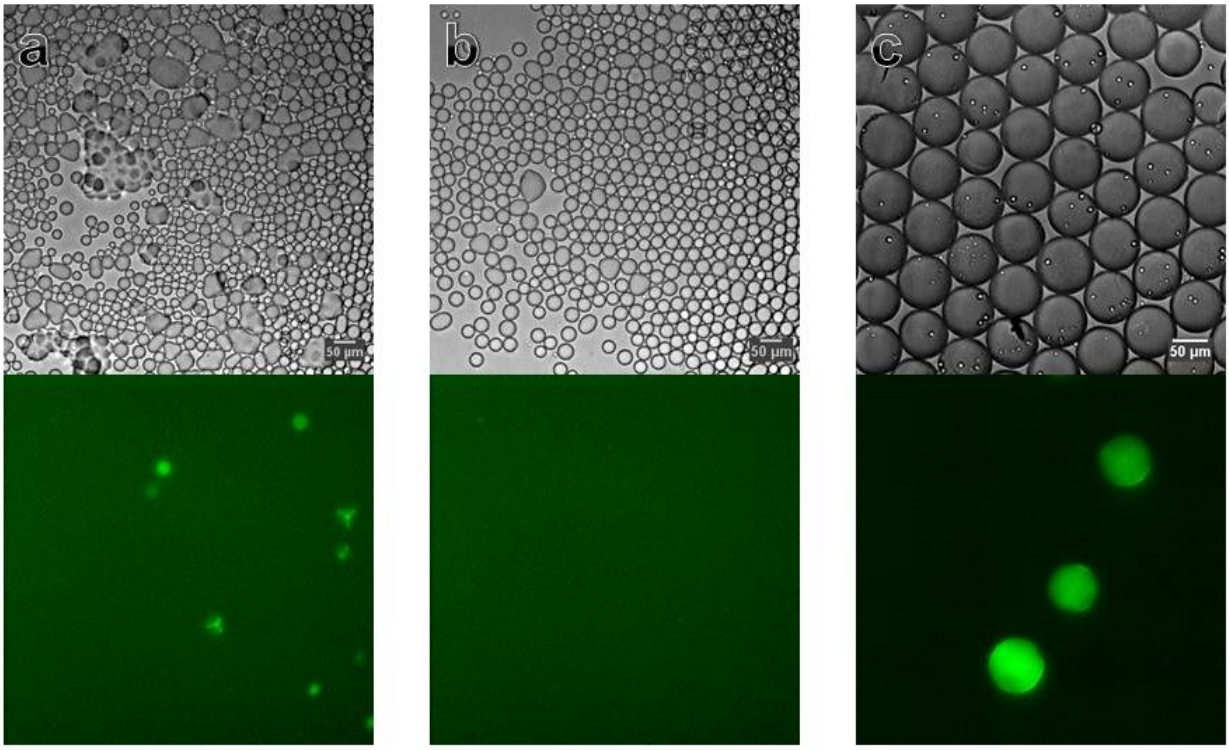


Figure 3.4: Fluorescent microscopy shows successful DNA amplification. (a) *E. coli* XL1 in 0.9% w/v agarose, (b) negative in 0.9% w/v agarose, (c) *E. coli* XL1 in 0.7% w/v agarose. Microgels are in oil after 30°C incubation, and contain SybrGreen included during second stage of encapsulation in agarose. 0.9% positive microgels exhibit polydispersity, while 0.7% positive microgels are monodisperse.

Following incubation at 30°C, we observe a high degree of polydispersity in the positive sample (Figure 3.4a, top). Polydispersity is far less prevalent in the negative control (Figure 3.4b, top), and is likely due to unstable flow during encapsulation. We do observe that even polyacrylamide microgels as small as ≈13 µm in diameter still perturb agarose droplet formation as they pass through the 14 µm constriction. Flow stability is also dependent on agarose viscosity and gelation state, and thus agarose concentration. Higher agarose concentrations not only increase viscosity, but also accelerate gelation at room temperature, so that it may occur before dropmaking is complete, and cause clogging in the microfluidic channels, tubing, or even

syringes. Furthermore, polyacrylamide microgel size, agarose viscosity, and agarose gelation state may synergistically destabilize agarose droplet formation.

We can eliminate polydispersity by using a coflow device with a larger 50 μm junction and 40 μm constriction, and reducing the agarose final concentration from 0.9% w/v to 0.7% w/v (Figure 3.4c, top). This supports our conjecture that the polydispersity is due to unstable flow rather than coalescence after collection. However, the resulting agarose droplets are much larger than desired at 62 μm in diameter, and furthermore do not gel reliably, even when left at 0°C for several hours. The issue thus remains to be resolved of how to minimize polydispersity for 30 μm agarose droplets, while maintaining a sufficiently high agarose concentration to easily gel after droplet production. Performing encapsulation in a heated environment at 30°C or slightly higher could be a solution, along with further reducing polyacrylamide gel size.

We also observe numerous aggregates of agarose microgels in the 0.9% sample, which again are more prevalent in the positive sample than the negative control. The cause of this is less clear, but the appearance of the aggregates is consistent with coalescence after collection, and after partial or complete gelation. Solid microgels are visible within the aggregates as distinct spheres; in a few aggregates, they are clearly separated by a bridge of differentially stained liquid phase. The agarose concentration may affect aggregate formation, as few if any are visible in the 0.7% sample.

3.3.2 MDA reaction

We successfully amplify DNA from single encapsulated bacteria, as detected by fluorescence from SybrGreen, which we include in the MDA reagents during the agarose encapsulation step. No fluorescence is visible in the negative control (Figure 3.4b, bottom), while both the 0.9% (Figure 3.4a, bottom) and 0.7% agarose (Figure 3.4c, bottom) concentrations contain droplets

filled with MDA product. Interestingly, a few of the microgel aggregates in the 0.9% sample exhibit fluorescence in the thin liquid film between microgels. This further supports our theory that hydrogel matrices exclude MDA and force the reaction to instead proceed inside whatever liquid volume is available, even if this is limited to a thin film. Aside from these few instances, bright microgels contain MDA product that is evenly distributed throughout their volume, and should thus be amenable to high-throughput sorting by FACS or microfluidics.

As a simple check to confirm that our MDA product is in fact amplified genomic DNA from our encapsulated *E. coli* XL1, we run a similar liquid eMDA protocol without any gel, followed by PCR and sequencing. For this experiment, we use the same final concentrations of reagents, except that we substitute water for agarose, acrylamide, and polymerization reagents. We use purified gDNA as the template, which we dilute in water to three concentrations equivalent to an average of 100, 10, or 0.3 genomes/drop. We also include a negative control of pure water (0 genomes/drop). Since there is only a single liquid aqueous phase, we use a single-junction dropmaker to encapsulate the gDNA dilutions in 61 μm droplets of MDA reagents.

After we incubate the MDA droplets at 30°C for 6 hours, the percentage of droplets in each sample that show bright fluorescence qualitatively matches our expectations, although it is significantly lower than the gDNA loading rate. We inactivate the phi29 by heating at 80°C for 20 minutes, then break the emulsion and use the coalesced aqueous phase as a template for PCR with primers targeting the tetB gene. As expected, only the 0.3, 10, and 100 gpd samples produce a band matching the tetB product, while the 0 gpd reaction is clean with no band. Finally, we use Sanger sequencing to confirm that the amplified bands all match the known tetB sequence (data not shown). We thus demonstrate that our MDA protocol accurately amplifies purified gDNA, when used with a liquid emulsion.

3.3.3 FISH

We are able to maintain the post-MDA microgels in a gelled state after washing them into aqueous buffer, but our current protocol requires some modifications to enable sequence-specific selection. After incubation at room temperature with the Cy5-labeled FISH probe, the microgels are still clearly visible as gelled spheres, but the staining is patchy and too uneven for automated sorting (Figure 3.5). However, the Cy5 staining does likely indicate the presence of MDA product, based on a comparison between the gels loaded with and without cells; the *E. coli* gels exhibit patchy staining, while the water-only gels exhibit no staining at all.

Two possible causes of the patchy staining are insufficient denaturation of the MDA product, or spatial segregation of the target sequence. It is possible that our alkaline treatment with 0.1 M NaOH does denature the MDA product, but since the DNA remains entangled in the agarose gel network, the two strands remain in close proximity and easily reanneal after we neutralize the NaOH. This could prevent the probe from hybridizing efficiently to the target sequence, since with our current protocol, we do not add the probe to the microgels until after neutralization. Including the probe in the denaturation or neutralization steps could help it to compete with the second strand. Some alternative denaturation agent, such as formamide or guanidine hydrochloride, may also be more effective.

Alternatively, the patchy staining may indicate that the probe has hybridized well with the target sequence, but the target sequence itself is spatially localized within the microgel, due to hyperbranching from a single locus. We could test this using a positive control of agarose microgels loaded only with a low concentration of tetB MDA product, which could be amplified in bulk solution from a tetB plasmid. After eMDA, these microgels should have the target sequence distributed evenly throughout their volume, since it should be the only sequence

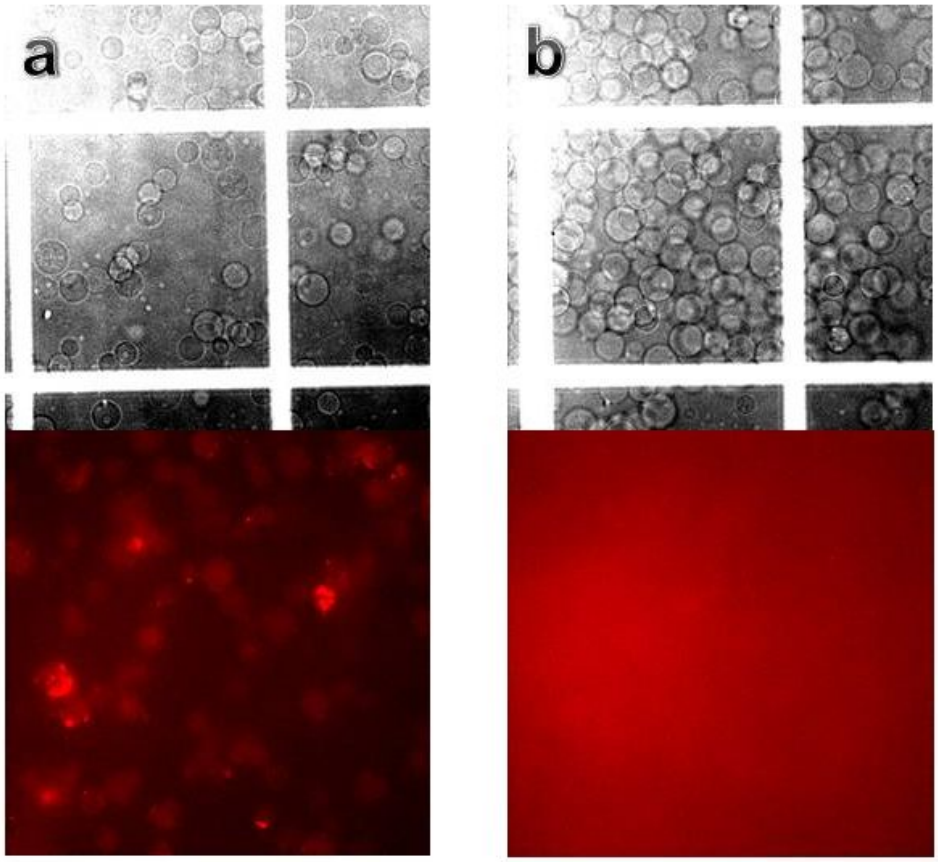


Figure 3.5: After MDA and FISH with Cy5 probe, microgels exhibit patchy staining. Microgels contain (a) *E. coli* XL1 or (b) no cells. Staining is too uneven for downstream automated sorting.

present. If FISH probing is no longer patchy but instead resembles SybrGreen staining, then our current results are likely due to spatial segregation.

Regardless of the cause, patchy staining could be circumvented by a third stage of encapsulation in polyacrylamide, followed by PCR targeting the tetB sequence, as has been previously described.[93] Using the agarose-encapsulated MDA product as a template for polyacrylamide-based gePCR produces microgel shells loaded with evenly distributed and covalently bound PCR amplicons, which can then be labeled with FISH probes and sorted by FACS.

3.4 Discussion

We have demonstrated several microfluidic techniques applied to genetic material:

- encapsulation of *E. coli* cells in polyacrylamide microgels smaller than 15 μm
- heat lysis of cells singly compartmentalized in a heat-resistant gel material
- replacement of one gel material with a second, heat-labile gel material, while maintaining single-cell compartmentalization
- single-cell emulsion MDA within liquid agarose droplets smaller than 30 μm
- gelation of agarose emulsion droplets loaded with MDA product, followed by washing into an aqueous continuous phase for downstream heterogeneous assays
- a heterogeneous assay consisting of a wash with a denaturing alkaline solution, reneutralization, addition of a FISH probe for sequence-specific staining, and a wash with buffer to remove unbound probe

With further optimization and an additional automated sorting step for MDA and next-generation sequencing library preparation in bulk, the combination of these techniques would comprise a WGS workflow for selected single cells (Figure 8).

Two key components of this proposed workflow must be improved for it to be viable. First, during the second stage of encapsulation, the agarose droplets must be monodisperse. Monodispersity is a key advantage of our microfluidic approach, compared to existing single-cell MDA methods that use bulk emulsification. It allows us to consider only a single, uniform droplet volume when tuning the cell concentration to minimize the number of multiply encapsulated cells, rather than having to consider the largest droplet size in a highly polydisperse mixture. Monodispersity thus increases throughput by eliminating the numerous smaller, empty drops that would otherwise be wasted volume. Second, the ability to select only those agarose

microgels containing a sequence of interest is critical for exploring rare cells. FISH is one way to do so, but is not straightforward for MDA product encapsulated in microgels. A second approach that has been successful is to perform a third round of encapsulation followed by PCR, effectively using the post-MDA agarose microgels as a template for gePCR.[93]

Chapter 4: Massively parallel sequencing of single cells by epicPCR

In the previous two chapters, we discussed two approaches to solving the problem of how to genetically characterize rare cells of interest from within a large and diverse population. In chapter 2, we proposed gePCR as a way to selectively generate two sets of amplicons from two different genes, so that when the two genes are carried by the same individual cell, both sets of amplicons will be associated with the same microgel. In chapter 3, we then proposed geMDA followed by FISH as a way to nonspecifically amplify all DNA from the population, yet in a way that keeps amplicons from the same cell spatially localized within the same microgel, and then sort out those droplets containing the amplicons of interest together with all other genetic information originating from the same cell. Both these approaches rely on localization of amplicons from different genes within a particular microgel, and then require a sorting step to isolate each microgel containing the target amplicons within a microtiter plate well. While sorting can be performed rapidly at kilohertz rates, this requirement still poses a fundamental limitation on throughput.

In this chapter, we summarize a previously published[94] third approach, epicPCR. As usual, we first isolate single cells in microgels, then suspend the microgels in emulsion droplets. However, we then selectively amplify two genes of interest from only those cells that carry both target genes, so that selection occurs as part of the amplification step itself. epicPCR furthermore produces fusion amplicons that each consist of a concatenation of the two target genes, so that genes originating from the same cell are colocalized not only within the same droplet, but also within the same molecule, and will thus be part of the same sequencing read. This eliminates the need to perform any sorting step between amplification and sequencing. Our throughput is therefore limited only by the number of sequencing reads we can obtain, and we can easily

sequence tens thousands of single cells in a single sequencing run. We demonstrate the utility of epicPCR for single-cell metagenomics by applying the method to a population of freshwater lake bacteria. By targeting the *16S* rRNA phylotyping gene together with the sulfate reductase gene *dsrB*, we identify the sulfate-reducing bacteria present in the population, including members of novel OTUs.

4.1 Results

The key aspect of epicPCR that enables linkage of two target genes within the same amplicon is the use of fusion primers. A traditional PCR primer pair consists of two oligomers that match different sequences within the same gene. Amplifying two different genes normally requires two separate sets of primer pairs, that is, four oligomers matching four different sequences. For epicPCR, we instead concatenate two of those oligomers targeting different genes into a single fusion primer that partially matches both of the target genes. Thus, amplicons from either gene that incorporate the fusion primer will themselves act as primers for the other gene. By including the fusion primer at a much lower concentration than the remaining two oligomers that act as regular primers, we control the reaction to primarily amplify the complete fusion product of both genes.

To verify that fusion amplicons represent genes from the same single cell, without any crosstalk between emulsion droplets, we perform a control experiment by spiking our lakewater sample with microgels loaded with covalently bound amplicons of a known *16S* sequence. Since these microgels contain only the *16S* gene without any *dsrB* sequences, fusion PCR targeting both genes should not recover any of the spiked-in *16S* sequence. Indeed, the spiked-in sequence is absent from all 372,223 reads we obtain, so our false positive rate is less than 3×10^{-6}.

epicPCR successfully identifies a number of known sulfate reducers in various *Deltaproteobacterial* families. We obtain *16S-dsrB* fusion amplicons matching *Syntrophobacteraceae*, *Syntrophaceae*, and *Desulfobacteraceae*. In addition, 16% of a total of 2 million amplicons represent novel OTUs, that is, bacteria that were not previously known to reduce sulfate. Our technique is thus useful for discovering bacteria in an environmental sample with a desired metabolic capability, so long as the target metabolic gene carries known conserved priming sites.

Since epicPCR first encapsulates single bacteria in microgels before adding PCR reagents, it has the advantage of enabling PCR-incompatible lysis steps before amplification, including enzymatic, detergent, and heat treatment. Independent of *dsrB* selection, we find that treatment with lysozyme, proteinase K, and Triton X-100 improves recovery of OTUs from the phyla *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Planctomycetes*, and *Chloroflexi* from lakewater samples 2 m below the surface. Using the same lysis treatment, we also recover rare candidate phyla, including H-178, which is present in the population at an abundance of 7.8×10^{-4}, according to conventional bulk sequencing results. We thus show that epicPCR can sequence single cells as rare as nearly 10^{-5}, and allows for chemical and heat pretreatment while maintaining single-cell compartmentalization.

4.2 Comparison to gePCR and geMDA

In comparison to gePCR and geMDA, epicPCR is especially suitable for applications where the only desired genetic information is the sequences of particular target genes, and not the rest of the genome. For such cases, epicPCR entails a straightforward workflow that can feed directly into next-generation sequencing library prep, without any sorting required since the selection

step is already included in the amplification step. Colocalization of the target sequences on the same molecule greatly facilitates downstream analysis, and enables sequencing of target cells at much higher throughput than gePCR or geMDA, since the limiting step is sequencing rather than sorting.

If recovery of the original genome is desired, the epicPCR workflow could in principle be modified to allow this, but throughput would then become comparable to that of gePCR or geMDA, since it would be necessary to sort out those droplets containing fusion amplicons along with their microgel-encapsulated original genomes. Some sequence-specific readout of the presence of the correct fusion amplicons would likely also be required; in theory, those droplets containing both target genes should contain orders of magnitude more PCR product than droplets containing only one of the targets, but some nonspecific background will typically be present. Once the positive droplets are sorted out, the microgels containing the original genomes can simply be centrifuged to separate them from the fusion PCR products.

For a detailed report on our epicPCR method and results, see Spencer, S.J. et al. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J* **10**, 427-436 (2016).

References

1. Guo, M.T., Rotem, A., Heyman, J.A. & Weitz, D.A. Droplet microfluidics for high-throughput biological assays. *Lab Chip* **12**, 2146-2155 (2012).

2. Wang, W. et al. Identification of small-molecule inducers of pancreatic beta-cell expansion. *Proc Natl Acad Sci U S A* **106**, 1427-1432 (2009).

3. Maaroufi, Y. et al. Rapid detection of Candida albicans in clinical blood samples by using a TaqMan-based PCR assay. *J Clin Microbiol* **41**, 3293-3298 (2003).

4. Kawakatsu, T., Kikuchi, Y. & Nakajima, M. Regular-sized cell creation in microchannel emulsification by visual microprocessing method. *Journal of the American Oil Chemists Society* **74**, 317-321 (1997).

5. Thorsen, T., Roberts, R.W., Arnold, F.H. & Quake, S.R. Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys Rev Lett* **86**, 4163-4166 (2001).

6. Anna, S.L., Bontoux, N. & Stone, H.A. Formation of dispersions using "flow focusing" in microchannels. *Applied Physics Letters* **82**, 364-366 (2003).

7. Tawfik, D.S. & Griffiths, A.D. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* **16**, 652-656 (1998).

8. Umbanhowar, P.B., Prasad, V. & Weitz, D.A. Monodisperse emulsion generation via drop break off in a coflowing stream. *Langmuir* **16**, 347-351 (2000).

9. Thorsen, T., Maerkl, S.J. & Quake, S.R. Microfluidic large-scale integration. *Science* **298**, 580-584 (2002).

10. Marcy, Y. et al. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* **3**, 1702-1708 (2007).

11. Marcy, Y. et al. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci U S A* **104**, 11889-11894 (2007).

12. Fan, H.C., Wang, J., Potanina, A. & Quake, S.R. Whole-genome molecular haplotyping of single cells. *Nat Biotechnol* **29**, 51-57 (2011).

13. Garstecki, P., Fuerstman, M.J., Stone, H.A. & Whitesides, G.M. Formation of droplets and bubbles in a microfluidic T-junction-scaling and mechanism of break-up. *Lab Chip* **6**, 437-446 (2006).

14. Koster, S. et al. Drop-based microfluidic devices for encapsulation of single cells. *Lab Chip* **8**, 1110-1115 (2008).

15. Edd, J.F. et al. Controlled encapsulation of single-cells into monodisperse picolitre drops. *Lab Chip* **8**, 1262-1264 (2008).

16. Holtze, C. et al. Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab Chip* **8**, 1632-1639 (2008).

17. Abate, A.R., Hung, T., Mary, P., Agresti, J.J. & Weitz, D.A. High-throughput injection with microfluidics using picoinjectors. *Proc Natl Acad Sci U S A* **107**, 19163-19166 (2010).

18. Ahn, K., Agresti, J., Chong, H., Marquez, M. & Weitz, D.A. Electrocoalescence of drops synchronized by size-dependent flow in microfluidic channels. *Applied Physics Letters* **88** (2006).

19. Link, D.R., Anna, S.L., Weitz, D.A. & Stone, H.A. Geometrically mediated breakup of drops in microfluidic devices. *Physical Review Letters* **92** (2004).

20. Chen, D. et al. The chemistrode: a droplet-based microfluidic device for stimulation and recording with high temporal, spatial, and chemical resolution. *Proc Natl Acad Sci U S A* **105**, 16843-16848 (2008).

21. Ahn, K. et al. Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices. *Applied Physics Letters* **88** (2006).

22. Agresti, J.J. et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc Natl Acad Sci U S A* **107**, 4004-4009 (2010).

23. Baret, J.C. et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab Chip* **9**, 1850-1858 (2009).

24. Dressman, D., Yan, H., Traverso, G., Kinzler, K.W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* **100**, 8817-8822 (2003).

25. Kiss, M.M. et al. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal Chem* **80**, 8975-8981 (2008).

26. Beer, N.R. et al. On-chip single-copy real-time reverse-transcription PCR in isolated picoliter droplets. *Anal Chem* **80**, 1854-1858 (2008).

27. Mazutis, L. et al. Droplet-based microfluidic systems for high-throughput single DNA molecule isothermal amplification and analysis. *Anal Chem* **81**, 4813-4821 (2009).

28. Brouzes, E. et al. Droplet microfluidic technology for single-cell high-throughput screening. *Proc Natl Acad Sci U S A* **106**, 14195-14200 (2009).

29. Hong, J.W. & Quake, S.R. Integrated nanoliter systems. *Nat Biotechnol* **21**, 1179-1183 (2003).

30. Love, J.C., Ronan, J.L., Grotenbreg, G.M., van der Veen, A.G. & Ploegh, H.L. A microengraving method for rapid selection of single cells producing antigen-specific antibodies. *Nat Biotechnol* **24**, 703-707 (2006).

31. Kelly, B.T., Baret, J.C., Taly, V. & Griffiths, A.D. Miniaturizing chemistry and biology in microdroplets. *Chem Commun (Camb)*, 1773-1788 (2007).

32. Meyer, M., Stenzel, U., Myles, S., Prufer, K. & Hofreiter, M. Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res* **35**, e97 (2007).

33. Han, M., Gao, X., Su, J.Z. & Nie, S. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat Biotechnol* **19**, 631-635 (2001).

34. Petrounia, I.P. & Arnold, F.H. Designed evolution of enzymatic properties. *Curr Opin Biotechnol* **11**, 325-330 (2000).

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

36. Brehm-Stecher, B.F. & Johnson, E.A. Single-cell microbiology: tools, technologies, and applications. *Microbiol Mol Biol Rev* **68**, 538-559 (2004).

37. Lecault, V. et al. High-throughput analysis of single hematopoietic stem cell proliferation in microfluidic cell culture arrays. *Nat Methods* **8**, 581-586 (2011).

38. Clausell-Tormos, J. et al. Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms. *Chem Biol* **15**, 427-437 (2008).

39. Moyed, H.S. & Bertrand, K.P. hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *J Bacteriol* **155**, 768-775 (1983).

40. Green, M. & Loewenstein, P.M. Human adenoviruses: propagation, purification, quantification, and storage. *Curr Protoc Microbiol* **Chapter 14**, Unit 14C 11 (2006).

41. Wobus, C.E. et al. Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. *PLoS Biol* **2**, e432 (2004).

42. Carter, P.J. Potent antibody therapeutics by design. *Nat Rev Immunol* **6**, 343-357 (2006).

43. Yokoyama, W.M., Christensen, M., Santos, G.D. & Miller, D. Production of monoclonal antibodies. *Curr Protoc Immunol* **Chapter 2**, Unit 2 5 (2006).

44. Yu, X., McGraw, P.A., House, F.S. & Crowe, J.E., Jr. An optimized electrofusion-based protocol for generating virus-specific human monoclonal antibodies. *J Immunol Methods* **336**, 142-151 (2008).

45. Yu, X. et al. Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* **455**, 532-536 (2008).

46. Kaur, J. & Sharma, R. Directed evolution: an approach to engineer enzymes. *Crit Rev Biotechnol* **26**, 165-199 (2006).

47. Sleator, R.D., Shortall, C. & Hill, C. Metagenomics. *Lett Appl Microbiol* **47**, 361-366 (2008).

48. Simon, C., Herath, J., Rockstroh, S. & Daniel, R. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* **75**, 2964-2968 (2009).

49. Morgan, R.D., Calvet, C., Demeter, M., Agra, R. & Kong, H. Characterization of the specific DNA nicking activity of restriction endonuclease N.BstNBI. *Biol Chem* **381**, 1123-1125 (2000).

50. Ottesen, E.A., Hong, J.W., Quake, S.R. & Leadbetter, J.R. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314**, 1464-1467 (2006).

51. Torsvik, V., Goksoyr, J. & Daae, F.L. High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56**, 782-787 (1990).

52. Kramer, M.F. & Coen, D.M. Enzymatic amplification of DNA by PCR: standard procedures and optimization. *Curr Protoc Mol Biol* **Chapter 15**, Unit 15 11 (2001).

53. Rueckert, A. & Morgan, H.W. Removal of contaminating DNA from polymerase chain reaction using ethidium monoazide. *J Microbiol Methods* **68**, 596-600 (2007).

54. Woyke, T. et al. Decontamination of MDA Reagents for Single Cell Whole Genome Amplification. *PLoS One* **6**, e26161 (2011).

55. Zeng, Y., Novak, R., Shuga, J., Smith, M.T. & Mathies, R.A. High-performance single cell genetic analysis using microfluidic emulsion generator arrays. *Anal Chem* **82**, 3183-3190 (2010).

56. Baker, M. Clever PCR: more genotyping, smaller volumes. *Nature Methods* **7**, 351-355 (2010).

57. Radstrom, P., Knutsson, R., Wolffs, P., Lovenklev, M. & Lofstrom, C. Pre-PCR processing: strategies to generate PCR-compatible samples. *Mol Biotechnol* **26**, 133-146 (2004).

58. Pantel, K., Brakenhoff, R.H. & Brandt, B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. *Nat Rev Cancer* **8**, 329-340 (2008).

59. Henke, W., Herdel, K., Jung, K., Schnorr, D. & Loening, S.A. Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res* **25**, 3957-3958 (1997).

60. Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R. & Phillips, R. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* **333**, 58-62 (2011).

61. Rodrigue, S. et al. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS One* **4**, e6864 (2009).

62. Woyke, T. et al. One bacterial cell, one complete genome. *PLoS One* **5**, e10314 (2010).

63. Dantas, G., Sommer, M.O., Oluwasegun, R.D. & Church, G.M. Bacteria subsisting on antibiotics. *Science* **320**, 100-103 (2008).

64. Cantalupo, P.G. et al. Raw sewage harbors diverse viral populations. *MBio* **2** (2011).

65. Turnbaugh, P.J. & Gordon, J.I. An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**, 708-713 (2008).

66. Gill, S.R. et al. Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359 (2006).

67. Sommer, M.O., Dantas, G. & Church, G.M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128-1131 (2009).

68. Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M. & Lasken, R.S. Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol* **11**, 198-204 (2008).

69. Miller, J.R., Koren, S. & Sutton, G. Assembly algorithms for next-generation sequencing data. *Genomics* **95**, 315-327 (2010).

70. Pignatelli, M. & Moya, A. Evaluating the fidelity of de novo short read metagenomic assembly using simulated data. *PLoS One* **6**, e19984 (2011).

71. Gomez-Sjoberg, R., Leyrat, A.A., Pirone, D.M., Chen, C.S. & Quake, S.R. Versatile, fully automated, microfluidic cell culture system. *Anal Chem* **79**, 8557-8563 (2007).

72. Eckburg, P.B. et al. Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638 (2005).

73. Turnbaugh, P.J. et al. A core gut microbiome in obese and lean twins. *Nature* **457**, 480-484 (2009).

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

75. Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A. & Quake, S.R. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* **6**, e16626 (2011).

76. Youssef, N.H., Blainey, P.C., Quake, S.R. & Elshahed, M.S. Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma). *Appl Environ Microbiol* **77**, 7804-7814 (2011).

77. Leng, X., Zhang, W., Wang, C., Cui, L. & Yang, C.J. Agarose droplet microfluidics for highly parallel and efficient single molecule emulsion PCR. *Lab Chip* **10**, 2841-2843 (2010).

78. Kumaresan, P., Yang, C.J., Cronier, S.A., Blazej, R.G. & Mathies, R.A. High-throughput single copy DNA amplification and cell analysis in engineered nanoliter droplets. *Anal Chem* **80**, 3522-3529 (2008).

79. Xu, J. et al. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, 2074-2076 (2003).

80. Williams, R. et al. Amplification of complex gene libraries by emulsion PCR. *Nat Methods* **3**, 545-550 (2006).

81. Schneeberger, C., Speiser, P., Kury, F. & Zeillinger, R. Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain. *PCR Methods Appl* **4**, 234-238 (1995).

82. Turnbaugh, P.J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031 (2006).

83. David, L.A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559-563 (2014).

84. Wu, G.D. et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105-108 (2011).

85. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. *Nature* **486**, 222-227 (2012).

86. Human Microbiome Project, C. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-214 (2012).

87. Munck, C. et al. Limited dissemination of the wastewater treatment plant core resistome. *Nat Commun* **6**, 8452 (2015).

88. Forsberg, K.J. et al. The shared antibiotic resistome of soil bacteria and human pathogens. *Science* **337**, 1107-1111 (2012).

89. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

90. Kurokawa, K. et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**, 169-181 (2007).

91. Yilmaz, S. & Singh, A.K. Single cell genome sequencing. *Curr Opin Biotechnol* **23**, 437-443 (2012).

92. Pop, M. Genome assembly reborn: recent computational challenges. *Brief Bioinform* **10**, 354-366 (2009).

93. Tamminen, M.V. & Virta, M.P. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Front Microbiol* **6**, 195 (2015).

94. Spencer, S.J. et al. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J* **10**, 427-436 (2016).

Appendix A: Making polyacrylamide microgels

Based on a protocol from Jeremy Agresti, 2/26/2009.

WARNING: Be sure to wear personal protective equipment that is appropriate for each step. I have included specific precautions where I felt steps included unusual hazards, but do not interpret the absence of a warning as an indication of low hazard. Be sure to read MSDS and equipment manuals.

Flow-Focusing Device (Drop Maker)

1. Combine 50g polydimethylsiloxane (PDMS) Sylgard elastomer (Dow Corning) + 5 g crosslinker in a single cup.
2. Mix using program 2 on Thinky mixer-defoamer.
3. Pour into 10 µm drop maker molds and degas.
4. Bake in big oven for at least 1 hour, preferably 4-6 hours.
   Note: Baking too short makes punched holes too small because the PDMS will deform from being too soft. Baking too long makes punched holes too big because the PDMS will not be elastic enough. Baking too long also dries out the PDMS, which can lead to cracking and leaking around the inlets/outlet, while baking too short can also lead to cracking and leaking because the PDMS cannot deform far enough around the inserted tubing.
   I have used 7 months-old pre-poured masters to make devices with no problem.
5. Punch inlet and outlet holes with biopsy punch.
6. Wash glass slide with isopropanol (IPA), then wipe firmly with cleanroom wipe. Wash PDMS with IPA, then blow dry with compressed air.
7. Plasma treat glass slide 2-3 times at maximum power (~125) for 120 s in old plasma oven. The last treatment should be immediately before the next step.
8. Plasma treat the glass slide together with the PDMS in the new plasma oven for 10 s at 30 power, then bond and hold together for 30 s. Bake at 65° for 15 min to overnight.
9. Aquapel the oil channels. Just touching the inlet should be sufficient. Device will visibly fill. Blow out Aquapel with a 1 mL BD plastic syringe and quickly dry any liquid that comes out to minimize PDMS swelling. Leave overnight at RT or 15 minutes at 65°C, then cover inlets/outlets with tape.

4x AB Stock

Option #1: Pore size ≈100 nm, total acrylamide 24%T, 3%C

Bis-acrylamide	3.6 mL
Acrylamide	2.58 mL
H_2O	3.82 mL

Option #2: Pore size ≈150 nm, total acrylamide 16%T, 3%C

Bis-acrylamide	2.4 mL
Acrylamide	1.72 mL
H_2O	5.88 mL

- Using a syringe and a 0.45 µm or smaller filter, filter the AB mix into a 15 mL Falcon tube.
- "Bis-acrylamide" refers to a mix of 40% total 19:1 acrylamide:bis-acrylamide aqueous solution (Sigma #A9926). "Acrylamide" refers to 40% acrylamide aqueous solution (Sigma #A4058).
- I have stored AB stock for 7 months @ 4°C without any problem.
- Option #2 works better for gel emulsion PCR.

Gel Mix

WARNING!! Acrylamide is toxic and carcinogenic. It is less dangerous in liquid solution than in crystal/powder form, simply because it is easier to contain and does not produce dust, but remember that if you allow any liquid to dry out, it will turn into acrylamide dust. Clean up any spills immediately and keep the liquid contained. Seal any significant amounts of waste liquid (> 50 µL) in a closed container in such a way that when it evaporates, acrylamide dust will not escape. The crystal/powder form should always be handled in a fume hood, but the liquid form may be handled on the bench as long as it is not allowed to evaporate completely.

1. Prepare fresh 10% APS. If the APS crackles audibly when you add water, then it is still good.
   - 0.0200 g APS → 200 µL PCR-clean dH_2O (VWR # 45001-044)
   - *APS Storage:* Powdered APS stock containers should be closed tightly, wrapped in parafilm, and stored in a dessicator at 4°C. APS solution can be aliquoted and frozen at -20°C immediately after preparation, and stored for several months. **Throw away any leftover solution** from each aliquot at the completion of an experiment – **do not refreeze.** When making fresh solution, minimize exposing the main stock of powdered APS to ambient humidity. Allow the main stock to warm to room temperature before opening the container.
2. Prepare the disperse phase aqueous gel mix (500 µL total volume). Do not add APS until immediately before making the drops.

Mix for 100 nm pore size

Final Conc	*Stock*	*Volume*
6%T, 3% C	4x AB, 24%T, 3%C	125 µL
0.2%	30% BSA	3.33 µL
1 µM	100 µM rev Ac primer	5 µL
0.2%	10% APS	10 µL
10%	TBSETw	50 µL
	Water	306.67 µL
		500 µL

Mix for 150 nm pore size

Final Conc	*Stock*	*Volume*
4%T, 3% C	4x AB, 16%T, 3%C	125 µL
1 µM	100 µM rev Ac primer	5 µL

0.3%	10% APS	15 µL
10%	TBSETw	50 µL
	Water	305 µL
		500 µL

- TBSE = 1x Tris-Buffered Saline + 10 mM EDTA
    - The TBS is to keep bacteria intact and prevent lysis, if encapsulating cells. The EDTA is to stabilize DNA and prevent nuclease activity, if encapsulating primers or gDNA. If you are not encapsulating cells or DNA, you may substitute water instead of TBSE.
- TBSETw = TBSE + 0.05% Tween-80
    - The Tween-80 is to dissociate any clumps of bacteria, if encapsulating bacteria. If you are not encapsulating bacteria, the Tween-80 is not necessary.

If encapsulating bacteria:

- Depending on the bacteria species, use OD of 0.2-1.2 for filling at 0.3 bacteria/drop.
- If bacteria tend to clump, wash them 1 time in TBSE + 0.05% Tween-80.
- If encapsulating gram-positive bacteria that may be difficult to lyse, culture them to a maximum of OD 1.0 before encapsulation, and increase the concentration if necessary by spinning and resuspending in a smaller volume. This works for *B. subtilis*: when grown to OD ~6.0, Bsu does not amplify by PCR, but when grown to OD 0.8 (about 6 hours from reinoculation with 5 µL saturated culture), it does. This is likely due to sporulation at high culture densities.

3. Prepare the continuous phase oil.

WARNING!! TEMED is toxic and corrosive, and should always be handled in a fume hood. It has a fishy smell and you should not be able to smell it during handling.

Final Conc	*Stock*	*Volume*
0.4%	TEMED	4 µL
	Oil (2% fluorosurfactant w/w in HFE-7500)	1 mL (750 mL if using 150 nm pore size gel mix)

Fluorosurfactant is a Krytox-PEG-Krytox triblock surfactant synthesized by RanBioTech, based on PMID 18813384.

4. Load into 1 mL clear plastic, disposable syringes with 27-gauge ½ inch needles (VWR #BD305109) and PE/2 tubing (Scientific Commodities #BB31695-PE/2). Alternatively, blunt needles (VWR #89134-170) are much safer, especially when handling Biosafety Level 2 or higher encapsulants, but are also much more expensive. Run pumps until liquid comes out of tubing end. Insert into flow focusing device inlets. For channel priming, run both phases at 300 µL/hour until one phase just reaches the junction. Then decrease that phase to 100 µL/hour until the other phase reaches the junction.
5. Run in device at a 1:1.5 inner:outer flow rate ratio. Devices can tolerate flow rates up to

Junction width	Q_{inner}	Q_{outer}	
10 µm	100 µL/hour	150 µL/hour	
15 µm	200 µL/hour	300 µL/hour	possibly higher, not tested

25 µm 600 µL/hour 900 µL/hour
10 µm device will make ≈15 µm drops at 10,000 drops/sec, which will swell to 20 µm in water. Outlet is 40 µm across at narrow part.

6. Collect initial polydisperse drops in a 200 µL gel-loading pipette tip (VWR #37001-150). Collect final monodisperse drops through PE/2 tubing in a microtube under 50-200 µL mineral oil to keep them from drying out.
7. Bake at 65°C overnight. For 100 nm pore size mix (6%T, 3%C final AB concentration), gels seem polymerized after 1 hour, and completely polymerized after 3 hours; 13-hour bake does not seem to affect gels' performance. For 150 nm pore size mix (4%T, 3%C final AB concentration), gels are completely polymerized after 12 hours; 24-hour bake does not seem to affect gels' performance.
8. (Optional) Test polymerization by adding 10 volumes IPA to 1 volume post-bake white gel cream. Vortex well, then centrifuge at 5krcf for 30 s. You should see a very solid, small opaque white pellet. Remove supernatant (IPA), add 5 volumes dH_2O, pipette up and down and vortex very well, and observe under a light microscope. The IPA/dH_2O wash will shrink and then reswell polymerized gels so that they will look like nice spheres. The IPA wash will dissolve unpolymerized acrylamide and nothing will be visible except possibly some debris.

For reusable glass syringes:

- Syringes should be washed with IPA, then dH_2O. If contamination is a concern, follow the dH_2O wash with a 0.1 M HCl wash, then wash very well with dH_2O. Blow dry with compressed air.

<u>Washing into Buffer</u>

1. Remove most of the mineral oil from the top of the gels.
2. Add 1 mL of 20% perfluorooctanol in HFE-7500. Vortex and centrifuge at 5krcf for 30 s.
   Note: This gets rid of the fluorosurfactant, displacing it from the interfaces with fluorinated octanol.
3. Remove lower oil phase. The gel particles will look like a transparent snotty continuous layer on top of the oil at this point.
   Note: If the particles are incompletely polymerized, you may see at this step that the snotty layer has a smaller volume than it should.
4. Add 1 mL of 1% v/v Span-80 in hexane, vortex and centrifuge at 5krcf for 30 s.
   WARNING!! Hexane is a neurotoxin and should always be handled in a fume hood. It smells similar to gasoline and you should not be able to smell it during handling.
   Note: Hexane is miscible with fluorocarbon oil, and immiscible with H_2O. Span-80 makes the drops disperse evenly.
5. Remove top hexane layer. The gels will still look like one solid mass, this time at the bottom of the tube.
   Note: If gels are incompletely polymerized, they will disappear at this point, or else begin disappearing at this point and continue disappearing in subsequent Span-80+hexane washes, possibly leaving a small white residual pellet. Not sure what this is. You may also see three layers (top: clear hexane, middle: cloudy/transparent gel pellet, bottom: clear liquid) instead of the correct two (top: clear hexane, bottom: cloudy/transparent gel pellet).
6. Add 1 ml of 1% v/v Span-80 in hexane, vortex and centrifuge at 5krcf for 30 s.
7. Remove top hexane.

8. Add 1 ml of TET (10 mM Tris + 10 mM EDTA + 0.1% triton X-100), vortex or pipette up and down to disperse gel particles.
   Note: For gels that contain DNA, it is important for the buffer to contain EDTA, which chelates Mg to prevent nuclease activity. PET (phosphate as a buffer instead of Tris) will likely also work. The Triton X-100 keeps the gels dispersed and prevents them from sticking to each other.
9. Centrifuge at 5krcf for 30 s.
   Note: For lower-concentration acrylamide or crosslinker gels, e.g. 4% Total acrylamide with 3% relative Crosslinker (4%T/3%C), a higher spin at 8krcf for 60 s may be necessary.
10. Carefully aspirate buffer from the top, making sure to remove oil in water emulsion (cloudy white layer) from the top of the water. Handle tubes gently as slight disturbances will cause the emulsion to mix downwards and be more difficult to remove. When aspirating, move the pipette tip quickly in circular motions to remove as much as possible from the walls.
11. Resuspend in TET and transfer to another tube. This will prevent the residual emulsion from mixing with the sample, keeping it cleaner.
12. Spin down, resuspend, and spin down again with TET (total of 3 washes in TET).
13. Resuspend in 1 mL TET.
14. Store gels at 4°C. Do not freeze.

Appendix B: Performing PCR in emulsified microgels

Based on a protocol from Jeremy Agresti, 4/24/2009

PCR Master Mix
For 1 mL:

Final Conc	*Stock*	*Volume*
1 μM	100 μM primer (forward, no Ac)	10 μL
1x	10x standard Taq buffer (w/ $MgCl_2$)	100 μL
0.5%	100% BSA	5 μL
250 μM	10 mM each dNTP	25 μL
0.2%	10% Tween-20	20 μL
	Water	840 μL
		1000 μL

Emulsion oil: Griffiths 3 formulation

Final Conc	*Stock*	*Volume*
4%	ABIL EM90	2 mL
0.05%	Triton X-100	25 μL
	Mineral oil	47.975 mL
		50 mL

Preparing Gels for PCR

1. Vortex gels, take 100 μL of suspension.
   Note: I have used gels up to several months old with no apparent drawback, except that if the DNA has an r_g < gel pore size, the concentration of DNA in the gels will be less than it originally was immediately after they were made. This happens even after a short storage time too (overnight or less) though.
2. Centrifuge at 5krcf for 30 s. Wash 2x in 1 mL water (VWR # 45001-044). This must be PCR-clean!
   Note: Gels swell significantly (≈2x volume) and are very difficult to see.
3. Wash 3x in 100 μL PCR mix (equal volume) with 5 min RT incubation between washes.
   Note: Less time may be sufficient, but at least some incubation time is necessary, as PCR fails if washes are done immediately after each other. Gels shrink significantly after 1st wash.
4. Add Taq polymerase (NEB #M0273) to final concentration of 0.175 U/μL of gel suspension in ½ volume PCR mix. For example, for 5 U/μL Taq polymerase stock, mix 2.8 μL Taq + 50 μL PCR mix, then add that to 30 μL gel pellet (which should result from 100 μL suspension). Incubate 10 min at RT.
   Note: Again, less time may be sufficient.
5. Centrifuge at 5krcf for 30 s and remove supernatant.
6. Add 13.3 volumes of emulsion oil to 1 volume gel pellet (e.g., 400 μL emulsion oil to 30 μL gel pellet). Pipet U/D 10x with P200 tip. Add 1 glass bead (2 mm diameter, flame-sterilized) for every ≈50 μL emulsion (e.g., 8 beads for 430 μL emulsion) and homogenize in beadbeater at low speed for 3 minutes.
   Note: Beadbeating speed and time will need to be optimized depending on the machine. If homogenization is not vigorous enough, many clumps of gels will be visible by microscopy. If it is too vigorous, fragments

of broken gels in half-moon or crescent shapes will be visible. See end of appendix for alternative reemulsification methods.

7. Aliquot 50 µL emulsion to each PCR tube.
8. Cycle in PCR machine without final low temperature hold.
   Note: Final low temperature hold contributes to emulsion breaking. Not sure whether final hold at RT or no final hold at all (gradual cooling from final T_{ext} to RT) is better.

Note: I usually use 50 µL gel suspension (half the example volumes above), but 100 µL will likely be easier to work with if you are learning the protocol and using it for the first time.

PCR Thermocycles

94°	3:00
65 cycles of	
94°	0:05
60°	0:45
72°	2:00
72°	6:00

total run time ≈4h30m

Note: Final t_{ext} of 6:00 may not be necessary. This is for gels encapsulating cells, with PCR targeting 16S universal 8f-907r.

Breaking Emulsion

1. Combine emulsions with 1 mL hexane. Use it to wash out the PCR tube as well, to minimize any lost emulsion volume that may remain on the tube surface. Hexane with Span-80 also works, if you want to conserve chemical storage space.
2. Spin down at 5krcf for 30 s and remove supernatant.
3. Add 1 mL hexane, vortex, and spin down again at 5krcf for 30 s.
4. Wash 3x with 1 mL TET buffer.
5. Resuspend in 4 volumes buffer for each volume of emulsion (e.g., 800 µL buffer for 200 µL emulsion).

Alternative Reemulsification Methods

These do not seem to work as well.

Syringe Pumping

1. Add 1.25 volumes oil to gel pack from 1 volume gel suspension (≈2/3 volume gel pack).
2. Vortex 10 s, pipette up and down quickly 25 times with a P200 tip.
3. Vortex briefly.
4. Sonicate 2 minutes.
5. Vortex 10 s to 2 hours.
6. Pump up and down 10 times through a 1 mL Normject syringe (or other oil-free syringe with minimal dead space in luer) with a 27½ gauge needle.
7. Aliquot into PCR tubes.

Tube Pumping

1. Same as syringe pumping, but for step 6., instead pump back and forth 10 times through PEEK or other small diameter tubing using two syringes.

- This is not very effective.
- This results in very small, shrunken (d < 5 μm) gels, possibly compacted due to high shear. The shear stress of being forced through the tubing may squeeze water out of the gels, making them smaller. They may recover and re-expand after a longer time period.

Pipetting

- Pipetting up and down with a 200 μL tip does not break up medium clumps.

Sonication

- Sonication seems to increase clumping.

Vortexing

- Vortexing reduces clumping, but only after a long time, and does not break up small clumps.
- Vortexing with several 2 mm diameter glass beads quickly breaks up clumps, but still does not break up small clumps.

Filtering

- Filtering through a silica/porcelain disc with 30 μm pores reduces clumping but with great sample loss, and some clumps still either get through or squeeze around the sides of the disc filter.

Appendix C: Performing FISH in microgels

Based on a protocol from Jeremy Agresti, 6/25/2009.

Materials

Denaturation reagents

0.1 M NaOH
0.1 M HCl

Polony annealing buffer (PAB)

50 mM NaCl
1.5 mM $MgCl_2$
10 mM Tris HCl, pH 8.0

1E buffer

10 mM Tris pH 7.5
50 mM KCl
2 mM EDTA
0.01% Triton X-100

Gels with Bacteria

Use microgels containing Bsu and Eco at 10^9 microgels/mL. For each sample, use at least 40 µL gel suspension (≈4×10^7 microgels; this is to facilitate microgel pellet visibility for ease of handling).
1 volume = 1 volume gel pellet, after spinning down

Preheat bath and 150 volumes 1E buffer for each volume of gel pellet to 9°C below the probe melting temperature (T_m).

Denature and Remove Free Strand

1. Spin down gel suspension at 5krcf for 30 s and remove supernatant.
2. Add 100 volumes 0.1 M NaOH to 1 volume gel pellet and vortex.
3. Incubate at room temperature for 5 min to allow free strand of product to diffuse out.
4. Spin down at 5krcf for 30 s, remove supernatant (NaOH), and add 5 volumes 0.1 M NaOH. Vortex.
5. Reneutralize with 5 volumes 0.1 M HCl.

Prepare and Probe Beads

1. Spin down at 5krcf for 30 s and remove supernatant.
2. Wash 3x in PAB (spin down at 5krcf for 30 s). Resuspend in 2 volumes PAB and transfer to PCR tube.
3. Add hybridization probe to ≈3x the concentration of target. (For 1 µM template, add probe to 3 µM final concentration. If using 100 µM probe stock, add 0.1 volumes probe.)
4. Run PCR cycle of 95°C 5:00, (T_m – 5)°C 10:00, (T_m – 9)°C final hold.
5. Add 150 volumes preheated 1E buffer to 1 volume gel pellet.

6. Spin down at 5krcf for 30 s.
7. Remove ≈84% volume.
8. Pipet remaining volume up and down to mix, and transfer to PCR tubes.
9. Spin down at 5krcf for 30 s and remove supernatant.
10. Resuspend in 1 volume 1E buffer.
11. Spin down at 5krcf for 30 s and remove supernatant.
12. Resuspend in 2 volumes 1E buffer.
13. Spin down and resuspend in 30 volumes 1E buffer.
14. Keep on ice until imaged.

Appendix D: Performing MDA in emulsified microgels

Based on a protocol from Manu Tamminen, 2/14/2012.

Cell Prep

Use cell stock frozen in 50% ethanol. Wash once in TBSETw (1x Tris-Buffered Saline + 10 mM EDTA + 0.05% Tween-80).

.

Acrylamide gels

WARNING: See Appendix A: for important safety information, reagent catalog numbers, buffer recipes, and detailed steps for producing gels.

Make gel stock with 22% acrylamide and 0.30% bisacryloylcystamine.

acrylamide + cross-linker stock	487.5 µL
cells	25 µL
10% Triton X-100	10 µL
10% APS	12.5 µL
	535 µL

1. Use 10 x 5 um dropmaking device with Q_i/Q_o of 100/150 uL/hr.
   Outer phase: 2% fluorosurfactant w/w in HFE-7500
2. Collect emulsion under air.
3. Add 0.8% v/v TEMED, relative to the total emulsion volume (oil + water drops).
4. Incubate overnight at room temperature, on a stir plate with a tiny stirbar (1 cm long or smaller) and gentle stirring.
5. Wash into TET storage buffer. Gels should have d ≈ 13 um.
6. Lyse encapsulated cells by heating tube in 95°C water bath for 5-10 minutes. Store at 4°C.

Agarose gels

1. Dissolve agarose to make stock solution: 2% w/v agarose powder in water
   - Use "strong" ultra-low gelling agarose, type IX-A (Sigma-Aldrich #A2576), to avoid premature gelling before drop formation.
2. Sterilize with a 0.2 µm filter. Stock can now be stored at 4°C.
3. Wash acrylamide gels 1x in 1 mL water and adjust to desired concentration.

Middle phase

2% agarose	779 µL
10x MDA buffer	100 µL
hexamers	50 µL
dNTPs	16 µL
DTT	40 µL
1:200 diluted SybrGreen	5 µL
φ29 polymerase	10 µL
	1000 µL

1. Heat agarose in 95°C water bath to melt, and prepare 30°C heat block.
2. Load syringes with all reagents except for the middle phase.
   Inner phase: acrylamide gels in water
   Outer phase: 2% fluorosurfactant w/w in HFE-7500
3. Immediately before starting flow, prepare the middle phase with all reagents except φ29 polymerase. Allow it to cool below 37°C, then add the polymerase and load it into the syringe.
4. Use 17 um coflow device with $Q_i/Q_m/Q_o$ of 111/155/1555 uL/hr.
5. Collect emulsion under air.
6. Image for 0 hour baseline. Drops should have d ≈ 33 um.
7. Incubate at 30°C for 6-18 hours.
8. Put tubes on ice for a few minutes to stop reaction and store at 4°C.

FISH

Do the following steps while **always keeping the tube on ice**.

1. Remove excess oil, or transfer emulsion cream to another tube.
2. Add an equal volume of perfluorooctanol to the emulsion cream, and three volumes of TET buffer.
3. Flick vigorously to mix.
4. Let settle on bench for a few minutes. You should see a shimmery aqueous phase separate.
5. Transfer the aqueous supernatant to a new tube.
6. Probe according to Appendix C: with the following changes:
   a. Omit all heating steps.
   b. Perform annealing at room temperature.
   c. Add FISH probe during PAB wash steps.