TURUN YLIOPISTO

Optimisation of the Split-Pool Barcoding Technique for Microbial Community Analysis

Jasmin Talvitie

Physiology and Genetics

Master's thesis

Credits: 30 op

Supervisors:

Manu Tamminen

Niina Smolander

27.11.2024

Turku

Master's thesis

Subject: Biology, Physiology and Genetics
Author: Jasmin Talvitie
Title: Optimisation of the Split-Pool Barcoding Technique for Microbial Community Analysis
Supervisors: Manu Tamminen, Niina Smolander
Number of pages: 61 pages, 12 appendix pages
Date: 27.11.2024

Understanding microbial community assembly and the role of horizontal gene transfer (HGT) and metabolic cross-feeding in community dynamics is essential for advancing microbial ecology. Traditional methods, including microscopy, culture-dependent techniques, and some culture-independent sequencing approaches, provide valuable insights but fall short in revealing fine-scale interactions and gene flow within complex microbial communities. This study aims to develop an optimised split and pool barcoding protocol for high-throughput, single-cell analysis of microbial interactions, which offers an approach for studying gene dynamics and functional diversity in varying ecosystems.

The split and pool barcoding method combines the modular barcoding approach of SPLiT-seq protocol with polyacrylamide beads and semi-permeable capsules to tag and sequence individual microbial cells. Key optimisations included reducing PCR cycles to minimise signal crosstalk and refining ligation conditions to prevent unwanted hybridisations. Efforts to use semi-permeable capsules faced fabrication challenges and signal mixing, resulting in their exclusion from the final protocol.

Results demonstrated successful targeting of the 16S rRNA V4 gene region, within synthetic microbial communities, achieving reliable single cell barcoding with minimal signal mix-up. The protocol provides a cost-effective, high-throughput method for microbial community analysis, addressing limitations of existing methods and enabling more precise tracking of transfer of genes and other community interactions. This advancement lays a solid foundation for further applications, such as studying the transfer of other genes, for example antibiotic resistance genes, in various increasingly complex microbial ecosystems.

Key words: protocol optimization, microbial communities, next generation sequencing, single-cell sequencing, polyacrylamide beads, semi-permeable capsules, split-pool barcoding

Contents

1 Introduction

1.1 Microbial Communities

Microbial communities are groups of microorganisms, such as bacteria, archaea, fungi, and viruses, that coexist in close environments and interact with each other and their surroundings (Konopka 2009). The concept of a community extends beyond physical boundaries. Rather than being only defined geographically or spatially, microbial communities are outlined by the intensity and nature of interactions among their members (Young et al. 2008). Ducklow (2008) and Ehrlich (1998) highlight how microbial communities play important roles in ecosystem processes, such as nutrient cycling and decomposition, by catalysing biochemical reactions that are essential for sustaining life. This process involves breaking down organic matter into inorganic compounds, while also altering chemical components between their reduced and oxidised states.

Interactions between microorganisms are essential in ecosystem dynamics and understanding them is crucial for advancing the study field of evolutionary ecology. Although the importance of these interactions is recognised, many interactions are still poorly understood. (Prosser et al. 2007.) According to Konopka (2009), interactions range from momentary interactions to more significant outcomes like horizontal gene transfer (HGT) and trait co-evolution. Investigating emergent properties of microbial communities, for example taxonomic and functional diversity, give insights into ecosystem stability. Stability, including resilience and resistance, together describe ecosystem responses to environmental disturbances. Richer diversity is often associated with better stability due to complementary traits of different species, by which species can compensate for each other's weaknesses and distribute ecological functions across multiple species. (Loreau 2000; Tilman et al. 2006.) Through processes, such as HGT, microorganisms can exchange genetic material, which influences the resilience and adaptation of the whole community. (Konopka, 2009.)

1.1.1 Microbial Heterogeneity

According to Chen et al. (2017), microorganisms are studied by examining their genes, interactions, and physiology, usually by analysing groups of microorganisms. However, these methods leave out heterogeneity at a cellular level, overlooking important variations between

individual cells as a result. This variation allows microorganisms to adapt to changing environments and conditions, therefore being crucial for their survival. Studying and understanding microbial heterogeneity offers unique comprehension of the behaviour and characteristics of microbial communities and overall ecosystem dynamics. Brehm-Stecher & Johnson (2004) explain that microbial community heterogeneity can be categorised into several groups depending on their biochemistry, physiology, genes, and functionality, and they often overlap, as biochemical or functional differences may emerge from genetic factors, and physiological differences can be influenced by external factors. Biochemical heterogeneity refers to differences in cellular activity or make-up, which is caused by processes such as aging of cells, turnover, differentiation events, genetic mutations, or random transcription noise. Physiological heterogeneity explains the different structural features of cells, such as varying shape and size and is influenced by cell cycle progression and microenvironmental conditions.

Genetic heterogeneity refers to differences in genetic sequences, mutations, and genomic arrangements seen within microbial populations. Genetic heterogeneity can be caused by various mechanisms, including mutations, genome rearrangements, and HGT. (Wilmes et al. 2009.) Thomas & Nielsen (2005) explain the three mechanisms of HGT: transformation, which involves the absorption of DNA by capable bacteria, transduction, which occurs when DNA is transferred via bacteriophages, and conjugation, which involves transferring genetic material between neighbouring bacteria through specialised organelles known as pili. Plasmids play a significant role in genetic heterogeneity by acting as vectors for HGT. Genetic heterogeneity mediated by these mechanisms drives microbial evolution, adaptation to changing environments, and the appearance of novel phenotypes within microbial communities.

Ackermann (2015) describes functional heterogeneity as variations in metabolic activity, gene expression patterns, and physiological responses, that are exhibited by microbial cells within a population. It is affected by differences in nutrient availability, spatial distribution, and intercellular communication. Evans et al. (2020) illustrate functional heterogeneity through the concept of metabolic cross-feeding, which occurs when metabolically different subpopulations trade metabolites, enabling them to use resources more efficiently and access metabolic pathways that would be otherwise unattainable. Through cross-feeding, bacterial populations can maximise their utilisation of resources and adapt to diverse environmental conditions, which enhances their fitness and survival. Understanding functional heterogeneity is crucial for deciphering complex biological processes such as development, differentiation, and fitness.

1.1.2 The Study of Microbial Communities

Traditional methods for studying microbial communities rely on microbial culturing, with specific media, for different microbes. These culturing approaches have revealed a diverse range of microbes that are involved in several ecosystem processes. However, culturing approaches only access under 0.1% of the microbial community due to a phenomenon known as cultivation bias, which limits the ability to culture microbes in laboratory conditions. (Hill et al. 2000.) An alternative culture-dependent method is the community-level physiological profiling, which evaluates microbial communities by examining their use of carbon substrates (Konopka et al. 1998). Garland & Mills (1991) explain the BIOLOG® system as a common tool for this approach. BIOLOG® uses 95 separate carbon sources to detect microbial activity through colour changes in wells. These profiles show functional differences among microbial communities, though they require careful standardisation and interpretation to avoid biases. There are also several limitations to these methods. The inoculum density affects the results, and some microbes may overshadow others in the wells, therefore skewing the data. Haack et al. (1995) note that physiological profiles may also be imprecise if some more abundant species in the samples grow efficiently on specific substrates or if there's competition during the microbial growth period.

Methods not relying on microbial cultures have gained popularity in microbial community analysis due to advancements made in molecular technologies and by having less limitations compared to their culture-dependent counterpart (Fukuda et al. 2016). Hill et al. (2000) divide culture-independent methods into two categories. The first category uses fluorescence microscopic methods, specifically fluorescent in situ hybridisation (FISH), which is a widely used technique for prokaryotic taxonomic studies. In FISH, rRNA is bound to fluorescent nucleic acid probes, and the probed cells are examined using a scanning confocal laser microscope. Fukuda et al. (2016) explain, that the advantages of FISH include the phylogenetic detection of microorganisms, high sensitivity, and the ability to show the location and structure of bacteria in a quick time frame. The method is only limited in its requirement of multiple differently labelled probes when targeting a variety of bacteria.

The second category of culture-independent methods rely on extracting molecules associated with particular microorganisms, such as phospholipid fatty acids or nucleic acids, proceeded by

measurement and identification. By comparing the molecular profiles from the samples to known profiles, the types, and abundances of microorganisms can be ascertained. Nucleic acid approaches, particularly the analysis of 16S ribosomal RNA (rRNA) genes, are crucial for understanding microbial diversity. (Hill et al. 2000.) Small ribosomal DNA molecules are present in Eukarya, Bacteria, and Archaea, and consist of conserved and variable regions, which makes them excellent targets for studying phylogenetic relationships and composition of microbial communities (Woese 1987; Woese et al. 1990).

Culture-independent methods, that target the 16S rRNA genes, include methods that use PCR, gene fragmenting, gel electrophoresis, clone libraries and next-generation sequencing (Fukuda et al. 2016). For example, quantitative PCR can be used to measure 16S rRNA genes in bacteria using fluorescent labels or dyes like SYBR Green (Bustin et al. 2005). Terminal restriction fragment length polymorphism can be used to examine differences in 16S rRNA genes by amplifying and digesting the gene region with fluorescently labelled primers (Zhang et al. 2008). Denaturing-gradient gel electrophoresis amplifies the 16S rRNA gene with Guanine-Cytosine-rich primers and separates DNA fragments on a polyacrylamide gel with denaturants (Muyzer et al. 1993). Clone library analysis can be done by extracting DNA, amplifying 16S rRNA genes, inserting the PCR products into plasmids, and sequencing the selected clones to identify microbes (DeSantis et al. 2007).

The development of next-generation sequencing technologies has significantly expanded the understanding of microbial diversity, revealing more bacteria and diverse taxa in various environments (Fukuda et al. 2016). These developments have also enabled the widespread use of 16S rRNA gene sequencing for targeted metagenomic studies (Kamble et al. 2020). Metagenomics explores microbial communities' genomic content in complex microbial ecosystems, such as those in soil, water bodies, and the human microbiome (Streit & Schmitz 2004).

1.2 DNA Sequencing and NGS

DNA sequencing is a technique by which the order of nucleotides is determined in DNA. Nucleotide sequences consist of four bases: adenine, guanine, cytosine, and thymine and they offer valuable information and understanding of genetic function. (Griffiths, A. J.F. 2024.) According to Shendure et al. (2017), early sequencing efforts relied on methods such as primer

extension, where the sequence information was obtained by analysing incorporated nucleotides, and RNA sequencing, where DNA was transcribed into RNA and cleaved at specific nucleotides for identification. However, two groundbreaking methods emerged in the mid-1970s that revolutionised DNA sequencing. One was developed by Sanger and Coulson, known as the chain termination method, while the other was created by Maxam and Gilbert, called the chemical cleavage method. Both approaches tracked the positions of bases across a DNA molecule using radioactive labels, with Sanger's method relying on DNA polymerase and chain-terminating nucleotides, and Maxam and Gilbert's method utilising chemical reactions to selectively break DNA at specific nucleotide positions (Maxam & Gilbert 1977; Sanger et al. 1977).

These two methods paved the way for new applications, such as the shotgun sequencing, proposed by Staden (1979), where randomly broken DNA fragments are sequenced and then assembled through their overlapping regions. The arrival of automated Sanger sequencers further accelerated DNA sequencing capabilities, reaching a thousand bases in a single day by 1987 (Hood et al. 1987; Smith et al. 1986). Further advancements in enhancing the efficiency of sequencing and software development led to decreasing costs, and the generation of millions of bases a day by 2001. The progress in sequencing efficiency also led to the completion of the genomes of *H. influenzae*, *S. cerevisiae*, and *C. elegans* in the mid and late 1990s. (Shendure et al. 2017.) The Human Genome Project's (HGP) human genome sequence followed closely behind and was finished in 2004 ("International Human Genome Sequencing Consortium" 2004).

According to Hu et al. (2021) at the beginning of the 21st century, a new era of DNA sequencing technologies, known as next-generation sequencing, emerged and rapidly replaced Sanger sequencing. NGS is characterised by its high throughput single molecule sequencing capabilities, meaning that it enables the sequencing of hundreds and thousands of genes or genomes in a short time. NGS techniques have significantly improved in precision and integration with data collection and analysis. As a result, they have greatly reduced the reliance on older, laborious sequencing methods.

'Next generation' can misleadingly imply advancements that go beyond what's currently available and to provide clarity, sequencing techniques are often categorised as second and third generation (Slatko et al. 2018). Hu et al. (2021) shed light on the significant milestones in the

evolution of NGS platforms. These milestones include the emergence of second-generation technologies, including Ion Torrent and Illumina, as well as the advancement to third-generation sequencing systems like Oxford Nanopore and Pacific Biosciences. Second-generation approaches rely on short reads where millions of short DNA fragments are parallelly sequenced, allowing sequencing reactions to occur simultaneously. However, these methods struggle with reassembling short fragments to reconstruct the original DNA sequence. This is because long DNA segments contain variations in structure and regions with repetitive sequences that can cause errors in the assembly process. Unlike short-read sequencing, third-generation sequencing methods generate reads that exceed 10 kb in length. Long-read sequencing technologies address the challenges associated with short reads and enables real-time targeting of DNA molecules. Initially, these technologies struggled with accuracy, but advancements in computational analysis have since enhanced their precision.

Kulski (2016) discusses that advancements in NGS technology together with the development of bioinformatics tools have enabled its use for various "-omics" fields like transcriptomics, proteomics, metabolomics, and many more. In microbial studies, NGS has played an important role in performing whole-genome shotgun sequencing (WGS) to analyse microbial genomes. WGS has significantly impacted microbial genomics, making genome sequencing more cost-effective, faster, and simpler due to the accessibility of reference genomes and advancements in technology (Kwong et al. 2015). Wang et al. (2009) highlight RNA sequencing (RNA-seq) as another essential application of NGS, which enables the analysis of all transcripts expressed by an organism's genome through various cellular contexts and developmental stages. RNA-seq offers high sensitivity and accuracy in detecting gene expression levels, splice variants, and noncoding RNAs, and provides valuable understanding of gene regulation and functional genomics. Epigenomics, focusing on DNA methylation and histone modifications, complements genomic studies by revealing heritable gene regulation mechanisms (Kulski 2016).

Quince et al. (2017) describe a common technique used to study microbial communities, which builds on the foundational principles of metagenomics, called shotgun metagenomics. Shotgun metagenomics involves breaking down a large assemblage of microbes and sequencing their DNA to provide a broad overview of the genetic makeup of the microbial community. While this method provides a way to identify sequences linked to individual species and assemble genomes, it struggles to distinguish DNA sequences that are shared among multiple taxa in one

sample, therefore limiting its ability to decipher genomes at the strain level. According to Sheth et al. (2019) and Slizovskiy et al. (2020) metagenomic approaches are also limited in their ability to decipher spatial information of microbes and precise co-localisation of genetic elements, such as plasmids within specific bacteria. Spatial information is hard to obtain because metagenomic approaches require homogenisation of the sample, which leads to loss of information of the spatial structure of microbes. Determining co-localisation of genetic elements is complex and still in early stages. Current approaches, that rely on alignment- or assembly-based methods, frequently generate incomplete or inaccurate information, hindering analysis of the co-localisation of genetic elements.

1.3 Single-Cell Approaches for Microbial Studies

Advancements in varying "-omic" profiling approaches and innovative microbial culturing techniques, such as culturomics and cultivation-based multiplex phenotyping, have greatly enhanced the understanding of diversity within microbial communities (Sommer 2015). Nevertheless, these methods are not without their limitations, and research into cellular heterogeneity and gene expression patterns has been less focused on prokaryotic cells compared to eukaryotic cells (Lloréns-Rico et al. 2022). Differences in mRNA structure, the low abundance of specific microorganisms, and the complexity of cell walls hinder isolation, cell lysis, and molecular workflows, which often results in an underrepresentation of genetic information in prokaryotic single cell genomic studies (Brennan & Rosenthal 2021). However, Lloréns-Rico et al. (2022) state that efforts are underway to tackle this challenge by developing single-cell methods that enable the isolation, cultivation, and analysis of single microorganisms' genomes and transcriptomes in diverse microbial communities. Single-cell sequencing (SCS) techniques provide better understanding of functional cellular heterogeneity and have become essential for studying genomic variations at a single-cell level.

1.3.1 Single-Cell Isolation and Culturing

Single-cell analysis begins with high-throughput isolation of cells while preserving their cellular functions and genetic properties (Chen et al. 2017). Methods for microbial single-cell isolation generally fall into two main categories: isolation using precise manipulation techniques, such as micromanipulation, and random encapsulation through droplet

microfluidics or compartmentalisation in physical microwells (Blainey 2013). Micromanipulation techniques allow precise isolation of individual cells using a microscope but can be labour-intensive and limited in throughput (Picelli 2017). Random cell isolation methods, including flow-assisted cell sorting and microfluidic devices, enable high-throughput isolation of single cells. Flow-assisted cell sorting offers the advantage of sorting cells based on multiple different parameters (Marie et al. 2017), while microfluidic devices provide more integrated single-cell analysis with minimal handling and contamination (de Bourcy et al. 2014). However, a persistent challenge remains in ensuring that each microdroplet contains a single cell (Chen et al. 2017). Techniques that are based on compartmentalising cells to microwells or chambers are alternatives to droplet microfluidics, by providing bigger cultivation chambers for the growth of cells and addressing issues in cell loading and viability (Lloréns-Rico et al. 2022).

1.3.2 Single-Cell Genomic (scDNA-seq) Analysis

Profiling microbial genomes can enhance the comprehension of microbial communities' phylogenetic relationships, ecological roles, and evolutionary processes (Lloréns-Rico et al. 2022). Single-cell DNA genome sequencing is a method where DNA from a single cell is amplified and sequenced, using NGS technologies (Chen et al. 2017). Microbial cells typically hold extremely low amounts of genomic DNA and to overcome this limitation, various whole-genome amplification (WGA) methods have been implemented and refined since the 1990s (Blainey 2013; Kim et al. 2017).

WGA methods include PCR-based methods, isothermal amplification methods, and hybrid approaches (Gawad et al. 2016). PCR-based approaches involve the amplification of gene regions using specific or random primers, but these methods can be low coverage (Chen et al. 2017). Isothermal amplification methods, such as Multiple Displacement Amplification (MDA), utilise polymerases that can displace existing strands or primers, and are particularly effective due to their extensive genome coverage and higher accuracy (Lasken 2012). However, MDA can struggle with uneven amplification of gene regions, leading to overrepresentation of some regions and underrepresentation of others (de Bourcy et al. 2014). Gawad et al. (2016) shed light on hybrid methods as potential solutions for the limitations of PCR-based techniques and the inconsistency of isothermal approaches in single-cell genomic sequencing. Hybrid methods combine the use of isothermal and PCR amplification. Displacement DOP-PCR adds

a common sequence using degenerate primers in the isothermal step, and PCR amplifies this target sequence, while multiple annealing and looping-based amplification cycles (MALBAC) uses random primers and temperature cycling to stimulate loop formation of amplicons, enabling more consistent amplification (Langmore 2002; Zong et al. 2012).

Achieving reliable single-cell genomics outcomes requires the ability to isolate cells without contamination and create an optimal setting for performing concurrent reactions (Hosokawa & Nishikawa 2024). Microfluidic-based single-cell genome analysis has become a key facilitator of the workflow from cell isolation to sequencing. It enables precise encapsulation, lysis, and amplification of single microbial cells in small-volume droplets, increasing the efficiency and reliability of single-cell genome analysis. (Kaminski et al. 2016.) A technique called Microbe-seq, developed by Zheng et al. (2022), has been introduced to better enhance the consistency of the droplet-based workflow for obtaining the entire genomes of single microbes. This workflow enables better throughput compared to traditional methods with microplates. Additionally, Microbe-seq enables the investigation of microbial interactions, such as HGT, within the community, and connections between bacteria and phages, revealing the genomic patterns of interactions among microbes within their populations.

Another method called single-cell genome sequencing at ultra-high-throughput (SiC-seq), presented by Lan et al. (2017), has been developed using droplet microfluidics. In SiC-seq, cells are enclosed in hydrogel microparticles, that allow lysis of cells and genome analysis while keeping each genome isolated. The genomes of lysed cells are fragmented, and unique barcodes are attached to each fragment. The workflow of SiC-seq enables the sequencing of tens of thousands of cells in a couple of hours and generates a database of single-cell genomes. However, while it improves accuracy in biological classification, it faces challenges in coverage and reliability. Nonetheless, SiC-seq offers significant potential for distinguishing biological heterogeneity in various organisms.

1.3.3 Single-Cell Transcriptomic (scRNA-seq) Analysis

While genomic analysis helps recognise and genetically profile microorganisms, further strategies are needed to fully understand the functional heterogeneity in populations of genetically identical bacteria. Recent advancements in single-cell transcriptomics have helped uncover unique gene expression patterns in prokaryotes. These methods overcome challenges

in analysing gene expression in individual bacterial cells, particularly the difficulties of isolating and studying such small entities at a single-cell level. (Lloréns-Rico et al. 2022.)

According to Wang et al. (2023) techniques for RNA sequencing in prokaryotes rely largely on cell and transcript indexing methods. These methods include in situ probe hybridisation, like par-SeqFISH (Dar et al. 2021) and ProBac-seq (McNulty et al. 2023), and split-pool combinatorial indexing, like PETRI-seq (Blattman et al. 2020) and microSPLiT (Kuchina et al. 2021). Par-seqFISH is derived from the seqFISH technique and extends its capabilities by recording gene expression patterns from single bacterial cells, while maintaining their original spatial organization within their environment. This is done by using a series of fluorescently labelled DNA probes in a high-throughput manner. The technique also measures cell size and other cellular characteristics, including chromosome copy number. ProBac-seq employs collections of DNA probes and a microfluidic system to achieve high throughput and sensitivity in uncovering gene expression variation and identifying various cellular conditions. Notably, ProBac-seq enables the identification of toxin expressing subpopulations within pathogenic bacteria, like *C. perfringens*, and provides information of the impact of environmental conditions on single-cell pathogenicity.

Both in situ probe hybridisation methods offer valuable tools for studying bacterial populations in their natural environments and help uncover the spatiotemporal dynamics of gene expression and cellular states. However, while hybridisation methods can detect known genes and RNA molecules using specific probes, they may miss new ones. (Wang et al. 2023.)

1.3.4 Single-Cell Split and Pool Barcoding

Combinatorial indexing methods can be used to help overcome the limitations of traditional approaches that require physically isolating bacteria within small droplets (Lloréns-Rico et al. 2022). Kuijpers et al. (2024) describe an innovative high-throughput approach that uses combinatorial indexing, called SPLiT-seq (depicted in Figure 1), which involves pooling cells, splitting them into a 96-well plate, and adding unique molecular barcodes/identifiers to mRNA/cDNA products through ligation. This process, also known as split and pool barcoding, is repeated until each cell contains a unique combination of barcodes. Within a complex mixture, such as a microbial community, these tagged molecules can be identified and linked back to their original cell using high-throughput sequencing. SPLiT-seq offers a solution to

address the growing demands for scalability and sensitivity in single-cell sequencing applications. It accommodates large projects that involve hundreds of samples and millions of cells, enabling the high-throughput analysis of complex microbiomes.

Two specific techniques, PETRI-seq and Microbial Split-pool Ligation Transcriptomics (Micro-SPLiT), that are built on the existing SPLiT-seq protocol, have successfully used this technology to profile gene expression across large numbers of bacterial cells (Xu et al. 2023). According to Lloréns-Rico et al. (2022), there are minor differences in both approaches, but one noteworthy distinction is that Micro-SPLiT shows a tendency to capture more mRNA from Gram-negative cells compared to positive ones, while PETRI-seq captures mRNA from both. Additional optimisation is also still needed, particularly in applicability and cost-effectiveness. Xu et al. (2023) elaborate that despite efforts to enrich mRNA, a substantial portion of sequenced reads in PETRI-seq and Micro-SPLiT methods consist of non-target ribosomal RNA, leading to increased costs without contributing to the desired data. Therefore, adapting combinatorial barcoding techniques for prokaryotic targets offers a promising direction for enhancing cost-effectiveness in studying transcriptional diversity within microbial communities.

Combinatorial split and pool barcoding can also be applied to genomic targets, as Quinodoz et al. (2022) have done in their protocol, called Split-Pool Recognition of Interactions by Tag Extension (SPRITE). SPRITE identifies interactions in the cell nucleus and spatial arrangements using cross-linking between interacting molecules and split and pool barcoding. During the process, DNA molecules are split to 96 wells, tagged with unique barcode sequences, and pooled. Since the molecules are linked, they stay together during the ligation rounds and end with the same barcode, providing information about which molecules interact with each other, when sequenced.

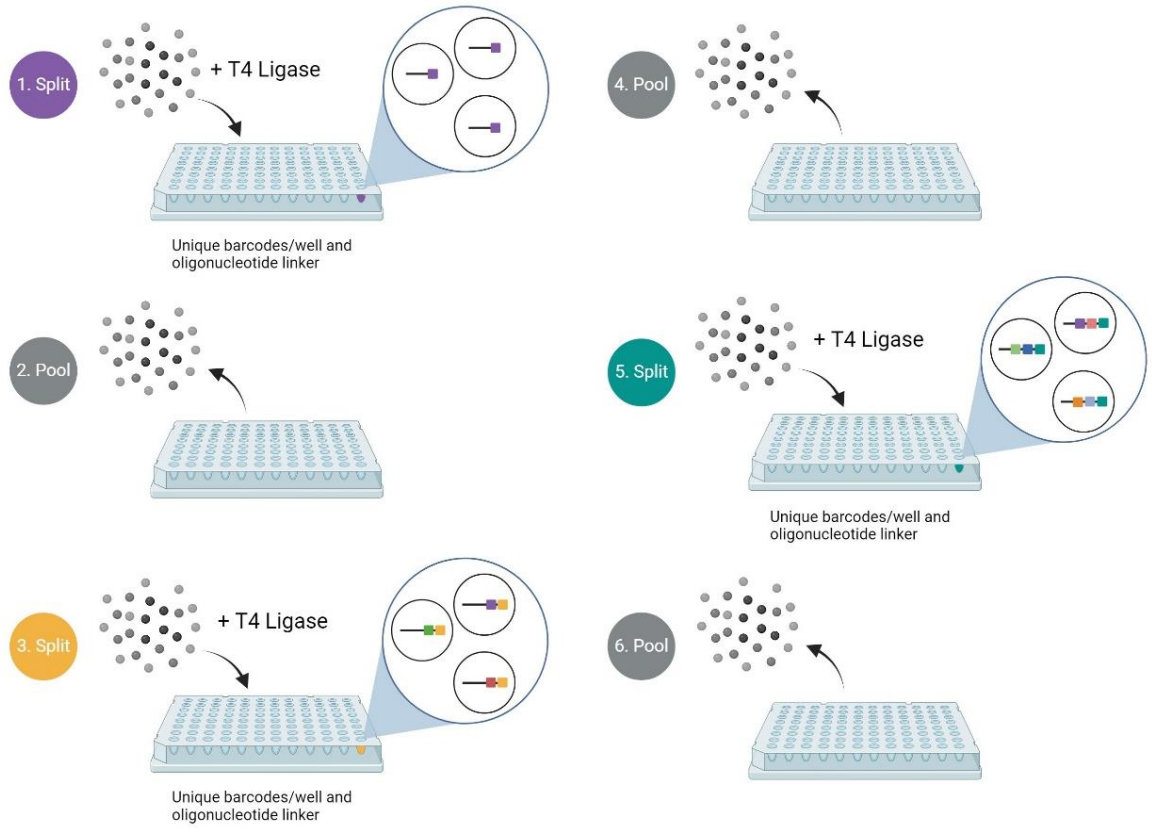


Figure 1. Workflow of the Split-Pool Barcoding protocol. Each round starts with dividing fixed cells (or encapsulated cells) into the wells of a 96-well plate. Unique molecular barcodes are added to the mRNA/cDNA structures with the help of an oligonucleotide linker within each well through a ligation process. The cells are pooled again and redistributed into a new 96-well plate, where a second set of unique barcodes is introduced. This cycle of splitting and pooling is repeated multiple times, with the number of unique barcode combinations increasing exponentially with each round. High-throughput sequencing of the final mixture reveals the unique combinations of barcodes, allowing the identification of individual cells and linking them back to their original biological context. Created with BioRender.com

1.4 Polyacrylamide Beads and Semi-Permeable Capsules for Split and Pool Barcoding

Traditional methods for analysing cell populations with high heterogeneity, such as fluorescence-activated cell sorting (FACS), have limitations in terms of cell throughput, isolating only hundreds or thousands of cells. Moreover, isolating single cells and amplifying their genetic material in microplates needs significant reagent amounts, causing large-scale studies to be expensive. (Zilionis et al. 2017.) Droplet microfluidics technology overcomes these limitations by allowing for the analysis of large numbers of cells with a highly parallel approach that is versatile, precise, has high cell capture efficiency, and is applicable in various single-cell studies (Theberge et al. 2010).

Existing single-cell sequencing approaches, that are based on droplet microfluidics, such as 10X Genomics, inDrop (Klein et al. 2015), and Drop-seq (Macosko et al. 2015), have limitations in terms of fabrication ease, barcode delivery efficiency, and cost-effectiveness. Trapping microbial cells in polyacrylamide beads offers an alternative solution. (Wang et al. 2020.) Tamminen & Virta (2015) describe a method where beads are made by polymerising polyacrylamide around microbial cells within an emulsion suspension. The structure of the beads supports the cell's genetic material, allowing for more efficient and thorough extraction of genetic material, by enabling stronger lysis conditions. This ensures the breakage of cells without compromising the integrity of the genomic material. Additionally, the beads do not need to be broken for chemical protocols, optimising downstream processes, and allowing for smooth integration into other workflows.

Semi-permeable capsules (SPCs), created by Atrandi Biosciences, provide an alternative to polyacrylamide beads for enhancing efficiency and versatility. These structures consist of an aqueous centre which is enveloped by a gel-based membrane, acting as selective barriers to allow small components to pass through while keeping single cells or macromolecules isolated. This capability enables a variety of downstream reactions and high-throughput analysis. Additionally, SPCs are biocompatible, making them suitable for different experimental conditions, including the use of strong chemicals. The advantages of SPCs mirror those of polyacrylamide beads but with additional benefits, such as providing more space to allow for a larger reaction volume and removing the need for immobilisation. This simplifies the experimental setup and reduces the risk of sample loss. (Leonaviciene et al. 2020.)

1.5 Aims and Objectives

This study ultimately aims to offer a way to answer the questions "How do microbial communities assemble?" and "What role do horizontal gene transfer and metabolic cross-feeding play in shaping their dynamics?". While microscopy offers perspective of community structure, understanding the underlying factors that shape these structures requires deeper examination. Split-pool barcoding presents an opportunity for researchers to study microbial interactions with exceptional precision. By elucidating the mechanisms that control the spread of genes within communities, these techniques hold promise not only in advancing scientific understanding but also in practical applications such as tracking the spread of pathogenic traits, for example antibiotic resistance, in clinical settings.

The primary objective of this study is to develop and refine the split-pool barcoding protocol so that it can be used to solve the complex interactions occurring within microbial communities, while also providing information about the practical applicability of the method. The study integrates the SPLiT-seq method, developed by Rosenberg et al. (2018), and a modular barcode approach developed by Delley & Abate (2021), to advance microbial single-cell profiling using polyacrylamide beads and semi-permeable capsules. The goal is that by optimising and testing this method, the study offers a way to gain deeper understanding of the functional diversity, underrepresented genes, and emergent properties that drive community dynamics, in a way that is reliable and cost-effective. Moreover, the study aims to help explain the underlying mechanisms controlling microbial interactions, thereby enhancing the understanding of microbial ecosystems.

2 Material and Methods

As the study focused on testing the split-pool barcoding technique for microbial community analysis, comprehensive optimisation of the pipeline was made through trial and error. Figure 2 illustrates the split-pool barcoding pipeline from start to finish.

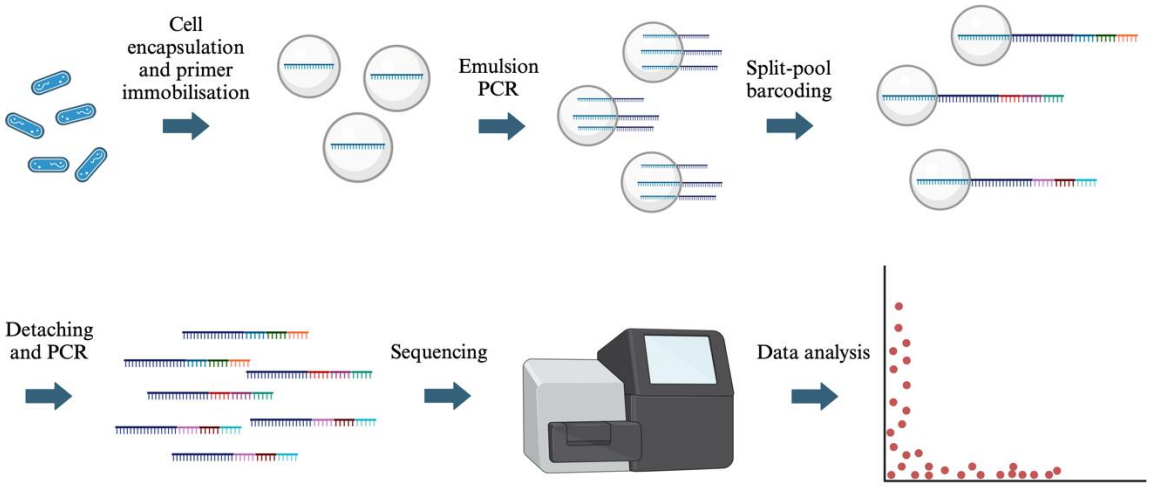


Figure 2. Workflow of the split-pool barcoding protocol optimised in the study. The process started with encapsulating cells and immobilising acrydited primers for attaching DNA to the beads. Subsequent steps included emulsion PCR, split and pool barcoding, detaching the barcoded DNA from the beads, amplification of barcoded product through PCR, sequencing using the Illumina MiSeq™ system, and data analysis. Created with BioRender.com

2.1 Cell Lines

The protocol was developed using three different HAMBI strains; HAMBI105 (*Agrobacterium tumefaciens*), HAMBI403 (*Comamonas testosterone*) and HAMBI262 (*Brevundimonas bullata*), provided by the Microbial Domain Biological Resource Center HAMBI.

The strains were stored in 25% ethanol fixative in -20°C. To prepare the strains for experimentation, 100 µL of each strain was thawed and then centrifuged for 1 minute at 13,000 rcf with a Heraeus Pico 17 centrifuge (Thermo Electron company). Following centrifugation, the supernatant was removed, and the cells were washed with 1x phosphate-buffered saline (1xPBS, Gibco) to remove any residual fixative or debris. After washes, the cells were resuspended in 100 µL of fresh 1xPBS.

2.2 Synthetic Polyacrylamide Beads

2.2.1 Amplification of the 16S rRNA Gene with an Overhang

The synthetic beads, incorporating PCR-amplified 16S rRNA gene region V4 from the target strains, served as positive control beads for refining the experiment protocol. To generate the 16S amplicons for the synthetic polyacrylamide beads, previously extracted DNA from HAMBI105 and HAMBI403 strains was used. The DNA extraction process was performed independently, and specific details regarding the extraction procedure are not provided in this study.

The extracted DNA was thawed at room temperature and amplified using primers designed to target the 16S rRNA gene region V4. The reverse primer contains an overhang, providing a binding site for the S-P linker 1:

16S_519F: CAGCMGCCGCGGTAATWC

16S-rev_S-P_PCR_OH: ACCACGCTCCAATTAAGCGGGGACTACHVGGGTWTCTAAT

A 4x volume PCR reaction mix was prepared, by adding 16 µL of 5x Phusion HF buffer (New England Biolabs, NEB), 1.6 µL of 10mM dNTP Solution Mix (NEB), 0.8 µL of 2,000 units/mL Phusion® Hot Start Flex DNA Polymerase (NEB), 4 µL of each 10 µM primer, and 49.6 µL of sterile H2O. For each reaction, 18 µL of the reaction mix was combined with 2 µL of extracted DNA or sterile H2O for negative control.

The PCR amplification was carried out using the ProFlex PCR System (Applied Biosystems) using the program provided in Table 1.

Table 1. PCR Program Parameters for the amplification of the 16S rRNA gene region V4 with an overhang.

Step	Temperature (°C)	Time	Cycles
Beginning temperature	98	30 s	1
Denaturation Annealing Extension	98 55 72	10 s 30 s 12 s	25
Final extension	72	5 min	1
Storing temperature	4		

Subsequently, the entire PCR sample volumes (20 µL) were loaded onto a 2% agarose E-Gel™ EX (Invitrogen) for gel electrophoresis with a 1:10 dilution of the E-Gel™ 1Kb Plus DNA ladder (0.5 µg/µL, Invitrogen). Gel electrophoresis was performed using an E-Gel™ iBase system (Invitrogen) for 10 minutes, and gel images were captured on a UV-stain-free tray using the SYBR® Gold Gel application on a GelDoc™ Go Gel Imaging System (Bio-Rad) (Appendix 1).

Following gel electrophoresis, the bands corresponding to the expected amplicon size, 285 bp, were excised from the gel and transferred to separate 1.5 mL Eppendorf tubes. The gel bands were then dissolved in 250 µL of Gel Dissolving Buffer (NEB) and incubated at 50°C for 10 minutes using a uniBLOCKTHERM digital drybath (LLG Labware) to dissolve the gel. After gel dissolution, the DNA of the dissolved gel bands was purified using the Monarch® PCR & DNA Cleanup Kit (NEB), according to the manufacturer's protocol (Appendix 2), and eluted in 20 µL of sterile H2O. The concentration and purity of the DNA samples were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific). DNA quantity and quality were deemed sufficient for further processing (Appendix 1).

2.2.2 Incorporation of Acrydited Uracil Primer

To incorporate an acrydited uracil primer into the DNA samples, enabling the attachment of DNA to the beads, a second round of PCR amplification was done. The designed forward primer contained the acrydited uracil modification, while the reverse primer was the same as in the previous PCR:

16S-Acryd-Uracil-For: `/5ACryd/ATGC/ideoxyU/GTGCCAGCMGCCGCGGTAA`

16S-rev_S-P_PCR_OH: `ACCACGCTCCAATTAAGCGGGGACTACHVGGGTWTCTAAT`

A 4x volume PCR reaction mix was made by adding 40 µL of 5x Phusion HF buffer (NEB), 4 µL of 10 mM dNTP Solution Mix (NEB), 10 µL of each 10 µM primer, 2 µL of 2,000 units/mL Phusion® Hot Start Flex DNA Polymerase (NEB) and 126 µL of sterile H2O to a 1.5 mL tube. The reaction mix was divided into three separate 1.5 mL tubes, with 48 µL allocated to each tube. 16 µL of previously amplified and purified template DNA from each HAMBI strain was added to their respective reaction mix tubes and sterile H2O for negative control.

The reaction mix was then divided into eight 50 µL reaction aliquots for each strain and the PCR amplification was carried out using the program provided in Table 1, with an annealing temperature of 60°C.

Following PCR amplification, the samples were pooled into separate 1.5 mL tubes. Gel electrophoresis was performed with 5 µL of PCR product of each HAMBI to visualise the amplicon size (Appendix 3), and bands corresponding to the amplified DNA size, 294 bp, were excised from the gel lanes. The excised bands were dissolved and purified, and their concentration was checked as in section 2.2.1 (Appendix 3).

2.2.3 Synthetic Polyacrylamide Bead Formation

Polyacrylamide bead formation was carried out to create beads containing the amplified DNA of each HAMBI strain. The bead formation process involved preparing two mixtures in a fume hood. An acrylamide-DNA mixture was made by adding 75 µL of HAMBI DNA amplicon, 2 µL of sterile H2O, 3 µL of 10 % ammonium persulfate, and 20 µL of 30 %/0.32 Acrylamide/Bis-acrylamide solution (Sigma-Aldrich) to a 2 mL tube. A RAN-TEMED mixture was made by adding 0.5 µL of >99 % Tetramethyl ethylenediamine (Sigma-Aldrich) and 100 µL of HFE7500 + 20g of 5 weight % 008-FluoroSurfactant RAN oil (RAN Biotechnologies) to a 1.5 mL tube. The acrylamide-DNA mixture was added to the RAN-TEMED solution and emulsified using a 150 µL pipette by pipetting in and out 20-30 times, resulting in a dense liquid emulsion. 100 µL of mineral oil (Sigma-Aldrich) was added on top without mixing, and the emulsion was incubated at 65°C for 16-18 hours.

2.2.4 Polyacrylamide Bead Clean-Up

To extract clean beads from the bead-oil mixture, bead clean-up was performed. This process included breaking the RAN oil emulsion and washing the beads. To break the RAN oil emulsion, the top, and bottom oil phases were removed, the sample was spun for 5 seconds with a uniCFUGE 2 tabletop centrifuge (LLG Labware), and the top and bottom phases were removed. Next, 50 µL of 97% perfluoro-1-octanol (Sigma-Aldrich) and 100 µL of 1xTE buffer were added to the mixture and mixed by tapping. Another 5-second spin was performed, and the bottom phase was discarded. To harvest the beads, 500 µL of 1xTE buffer was added, and

the mixture spun for 5-10 seconds. Most of the top phase (beads) was transferred into a new 2 mL tube and the remaining liquid was discarded. To wash the beads, 1 mL of 1xTE buffer was added and the samples were centrifuged for 1 minute at 13,000 rcf. The supernatant was removed, and the washing step was repeated. After the second wash, the beads were resuspended in 500 µL of 1xTE.

2.2.5 Sodium Hydroxide Treatment and Size Filtering

A NaOH treatment was done to make the DNA within the beads single-stranded. First, a 0.2M NaOH solution (Sigma-Aldrich) was prepared and vortexed. The beads were then combined with the 0.2M NaOH solution in a 1:1 ratio, vortexed, and incubated for 10 minutes at room temperature. The mixture was then vortexed and centrifuged for 1 minute at 13,000 rcf. The supernatant was removed, and the beads were washed twice with 200 µL of 1xTE buffer. Once washed, the beads were resuspended in 1xTE buffer, and size filtered to select only 70 µm sized beads, with a cell strainer (VWR®). The success of the bead formation was checked with a fluorescence microscope (Evos Floid Imaging System, Invitrogen).

2.3 Cell Containing Polyacrylamide Beads

2.3.1 Cell Encapsulation

To create cell containing polyacrylamide beads, cells from HAMBI105, HAMBI403, and HAMBI262 strains were encapsulated. For early testing, beads were made separately with either HAMBI105 or HAMBI403 strains. In later testing, one set of beads contained a mixture of HAMBI105 and HAMBI262, while another set contained only HAMBI403.

The process of cell encapsulating bead formation closely resembled that of synthetic bead formation, with a few key differences. The beads were overloaded with HAMBI cells to ensure that each bead contained at least one cell and the 16S acrydited forward primer was added to the acrylamide-DNA mixture:
16S-Acryd-Uracil-For: `/5ACryd/ATGC/ideoxyU/GTGCCAGCMGCCGCGGTAA`

Two mixtures were made once again in the fume hood, with the acrylamide-DNA mixture containing 10 µL HAMBI cells (suspended in 1xPBS), 62 µL 1xTE buffer, 5 µL 100 µM

acrydited forward primer, 3 µL 10% ammonium persulfate and 20 µL 30% acrylamide/bis-acrylamide suspension (37,5/1 ratio, Sigma-Aldrich). The rest of the protocol was done the same as previously in sections 2.2.2, 2.2.3, and 2.2.4. However, instead of treating the beads with NaOH, they were solely subjected to 70 µm size filtering.

2.3.2 Emulsion PCR

To amplify the 16S rRNA gene region V4 of the bead-encapsulated HAMBI cells, without the risk of cross-contamination between beads, the different HAMBI beads were combined and underwent an emulsion PCR using the same primers as in section 2.2.1.

The PCR reaction mix volume varied based on how much bead material was needed for downstream processes. For one reaction, 20 µL of 5x HF buffer (NEB), 2 µL of 50mM MgCl2 (NEB), 2.5 µL of 10mM dNTPs (NEB), 8 µL of 2,000 units/ml Phusion® Hot Start Flex DNA Polymerase (NEB), 2.5 µL of each 40 µM primer, and 15 µL of sterile H2O was added. A 2 mL tube was prepared for each replicate, with three 4 mm glass beads (Supelco®) added to each tube. The reaction mix was divided into the glass bead containing tubes, with 52.5 µL allocated to each tube followed by the addition of 47.5 µL of previously prepared polyacrylamide beads. 900 µL of a solution containing 4% ABIL EM 90 (Evonik) and 0.05% Triton X-100 (Sigma-Aldrich) in mineral oil was added to each tube and the samples were emulsified by vortexing for 1 minute. Each replicate was divided into sixteen 60 µL aliquots and the following PCR program was run (Table 2). The run was experimented with three different cycle amounts: 32 cycles, 25 cycles, and 20 cycles. Subsequent experiments were carried out using 20 cycles.

Table 2. PCR program parameters for amplifying the 16S rRNA gene region V4 with an overhang in an emulsion. The PCR program was tested with three different cycle amounts: 32, 25, 20.

Step	Temperature (°C)	Time	Cycles
Initial denaturation	94	30 s	1
Denaturation Annealing Extension	94 55 72	10 s 30 s 12 s	32/25/20
Final extension	72	5 min	1
Storing temperature	4		

2.3.3 Emulsion Breaking

After the PCR, the replicates were pooled, and the emulsion was broken to recover the amplified DNA beads. The process of emulsion breaking was carried out in a fume hood and reagents were shaken before use. First, 1 mL of water-saturated diethyl ether was added to each sample and the samples were vortexed and centrifuged for 1 minute at 13,000 rcf to separate the phases. The top phase was discarded, and the step was repeated. Next, 1 mL of water-saturated diethyl acetate was added to each sample. The samples were mixed by tapping and spun for 5-10 seconds. If floating particles remained, the samples were centrifuged as before. The top phase was discarded. Next, 1 mL of diethyl ether was added to each sample. The samples were mixed by tapping gently, so that the bottom phase became transparent, and the top phase was discarded. The diethyl ether wash step was repeated, and the samples were left in the fume hood with the cap open, for at least 3 minutes to allow the remaining ether solvent to evaporate. After the ether had evaporated, the samples were transferred into new 1.5 mL tubes and the previous 2 mL tubes were rinsed with 500 µL 1xTE buffer and transferred to the 1.5 mL tubes. The samples were centrifuged for 1 minute at 13,000 rcf, the supernatant was discarded, and the samples were resuspended in an appropriate amount of 1xTE buffer.

The beads were subjected to a NaOH treatment as detailed in section 2.2.4 and resuspended to a total volume of 530 µL with 1xTE buffer. The success of the bead formation, amplification of the DNA, and the density of beads was checked with a fluorescence microscope (Evos Floid Imaging System, Invitrogen) by mixing 4 µL of the sample with 0.5 µL 1:100 dilution of SYBR® Safe DNA Gel Stain (Invitrogen) and checking the fluorescence of the beads.

2.4 Cell Containing Semi-Permeable Capsules

2.4.1 Cell Encapsulation

To generate cell containing semi-permeable capsules, the cells of the two HAMBI strains were encapsulated using microfluidics. Cell-containing capsules were made, with one set containing HAMBI105 cells and the other containing HAMBI403 cells. This was done to ensure the accuracy of results by verifying against any unintended signal mixing.

The semi-permeable capsules were generated using reagents from the SPC Generation Kit (Atrandi Biosciences). A core mixture was made by mixing 50 µL of 2x Core Reagent, 1 µL of 1 M dithiotreithol (Thermo Scientific), 12.5 µL of photoinitiator, and 36.5 µL of HAMBI cells in 1xPBS (H105 or H403) in a 15 mL falcon tube. A shell mixture was made by mixing 50 µL of 2x Shell Reagent with 50 µL of 1xPBS in another 15 mL falcon tube. The core and shell mixture tubes were attached to a pressurised microfluidics device with an SPC generation chip to create semi-permeable capsules containing HAMBI cells. The generated semi-permeable capsules were intermittently exposed to 365 nm UV light for 30 seconds at a time, approximately every few minutes to polymerise the photoinitiator within the core mixture and form a stable, semi-permeable matrix around the cells.

2.4.2 Emulsion Breaking and Buffer Exchange

Following the generation of semi-permeable capsules, the emulsion was broken according to the 'Emulsion Breaking' protocol detailed in the Flux System (SPC Generation Kit) user guide (pages 9-10) using reagents from the SPC Generation Kit. A Capsule Wash Buffer was prepared by diluting 50x Wash Additive to 1x with 1xPBS. Excess oil from the bottom of the tube was removed and 500 µL of Capsule Wash Buffer and 500 µL of Emulsion Breaker were added to the generated emulsion. The tube was mixed by inverting 5-10 times and spun briefly. The capsules were resuspended to the aqueous phase by pipetting and after 10 seconds, the top phase was transferred to a new 1.5 mL tube.

The buffer was changed by centrifuging the capsules for 1 minute at 300 x g and discarding the supernatant. The capsules were resuspended in 1 mL of 1x Capsule Wash Buffer, and the process was repeated three times. The capsules were stored in Capsule Wash Buffer at 4°C.

2.4.3 Cell Fixation and Lysis

Cell fixation and lysis were done following the Atrandi Single-Microbe DNA Barcoding User Guide for cell fixation and lysis, with modifications to the lysis and wash buffers. First, 100% methanol was chilled on ice. For 100 µL of capsules, 900 µL of cold 100% methanol was added drop by drop while agitating the capsules using a Thermomixer C at 300 rpm. The sample was then incubated at -20°C for one hour.

Subsequently, the sample was washed three times by vortexing, centrifuging for 1 minute at 1000 x g, discarding the supernatant, and aspirating the sample with a custom wash buffer with 1xTE and 0.1% Tween 20 (Sigma). After washing, the sample volume was adjusted to 500 µL with the wash buffer, and a custom lysis buffer was prepared.

500 µL of the lysis buffer (0.38 M potassium hydroxide, 32 mM ethylenediaminetetraacetic acid disodium salt dihydrate (Invitrogen), and 26 mM dithiothreitol (Thermo Scientific) in sterile H2O), was added to the capsule suspension and mixed by vortexing. The sample was rotated for 15 minutes at room temperature using a tube rotator at speed 10 (Fisherbrand™). Following this, the sample was washed five times with a custom neutralisation buffer (1.42 % hydrochloric acid (HCl), and 0.197 M Tris-HCl (pH 8) in sterile H2O).

After neutralisation, the capsules were resuspended in 500 µL of 1xTE buffer, and 10 µL of 20 mg/mL proteinase K (Macherey-Nagel) was added. The sample was incubated at 37°C for 30 minutes. After incubation, the sample was washed five times with the custom wash buffer and resuspended in 100 µL of 1xTE.

2.4.4 Amplification of the 16S Gene Region

Prior to the split-pool barcoding, the 16S rRNA gene region V1-V4 of the HAMBI cells was amplified using primers 27F and Phos-16S-rev_OH, with the latter containing an overhang that provides a binding site for the S_P linker 1 and a 5´-phosphorylation. A longer 16S region was amplified here than with the polyacrylamide beads to prevent the amplicon diffusion out of capsules.

Used primers:

27F: `AGAGTTTGATCCTGGCTCAG`

Phos-16S-rev_OH: `/5Phos/ACCACGCTCCAATTAAGCGGGGACTACHVGGGTWTCTAAT`

A 9x volume PCR reaction mix was prepared to secure enough material for later processes. 90 µL of 5x Phusion HF buffer (NEB) was mixed with 9 µL of 10mM dNTP Solution Mix (NEB), 22.5 µL of each 10µM primer, and 18 µL of 2,000 units/mL Phusion® Hot Start Flex DNA

Polymerase (NEB) in a 1.5 mL tube. Following reaction mix preparation, 18 µL of the reaction mix was combined with 32 µL of lysed capsules in a 0.2 mL PCR tube per reaction.

The PCR amplification was carried out using a specified program, provided in Table 3.

Table 3. PCR program parameters for the amplification of the 16S gene region with an overhang.

Step	Temperature (°C)	Time	Cycles
Beginning temperature	98	30 s	1
Denaturation Annealing Extension	98 55 72	10 s 30 s 12 s	30
Final extension	72	5 min	1
Storing temperature	4		

After amplification, all replicates were pooled and washed twice with 1xTE buffer. The success of the PCR amplification was verified by mixing 4 µL of the sample with 0.5 µL 1:100 dilution of SYBR® Safe DNA Gel Stain (Invitrogen) and checking the fluorescence of the sample with Evos Floid Imaging System (Invitrogen).

2.4.5 Lambda Exonuclease Treatment

For making the DNA inside the capsules single-stranded, the 5' phosphorylated linear double-stranded DNA was degraded from 5' to 3' direction with a 5,000 units/mL Lambda exonuclease (NEB). The pooled sample (400 µL) was divided into four separate 100 µL reactions. Each reaction mix consisted of 50 µL of 16S amplified capsules, 10 µL of 10x Lambda Exonuclease Reaction Buffer, 2 µL of Lambda Exonuclease (5 units) and 38 µL of sterile H2O. The reaction was incubated at 37°C for 30 minutes. To stop the reaction, 1 µL of 0.5M EDTA was added, and a heat inactivation was done at 75°C for 10 minutes. Lastly, the capsules were washed five times with 1xTE buffer.

2.5 Split-Pool Barcoding Protocol

2.5.1 Linker-Barcode Plates

Prior to the split-pool barcoding, three linker-barcode plates were prepared. First, three barcode 96-well plates (Appendix 4) and three linker oligos were diluted to a concentration of 10 µM:

Linker 1: `ACCCTTGCTCAGAACACCACGCTCCAATTA`

Linker 2: `AGTCGTACGCCGATGCGAAACATCGGCCAC`

Linker 3: `CGAATGCTCTGGCCCTCAAGCACGTGGAT`

5 µL of diluted Linker 1 was transferred to each well in a low-profile 96-well plate. Subsequently, 5 µL of barcode from each well of the 10 µM barcode 1 plate was added to the linker 1 plate, adhering strictly to the matching well positions (e.g., A1 barcode transferred to A1 on the linker plate). To hybridise the barcodes to the linkers, the linker-barcode plate was heated to 95°C, followed by cooling to 20°C at a rate of -0.1°C/s, using a C1000 Touch Thermal Cycler (Bio-Rad). The same steps were repeated for linker-barcode plates 2 and 3.

2.5.2 Split-Pool Barcoding

To attach unique combinations of barcodes to the DNA in the beads, as shown in Figure 3, a split-pool barcoding protocol was employed. This protocol involved multiple rounds of ligation, blocking, and washing, with specific optimisations incorporated to enhance the efficiency and accuracy of the process.

Figure 3. The sequence attachment process: starting with a 16S sequence, an overhang is added to which linker 1 attaches. Barcode 1 is then attached directly next to the overhang via linker 1. Subsequently, linker 2 attaches to barcode 1, allowing barcode 2 to attach next to barcode 1. Linker 3 then attaches to barcode 2, enabling barcode 3 to attach next to barcode 2. This results in a unique combination of three barcodes, each connected by specific linkers.

A 106x volume ligation mix was prepared on ice, by adding 265 µL of 10x T4 Ligase buffer (NEB), 530 µL of polyacrylamide beads, 742 µL of sterile H2O and 53 µL of T4 Ligase (NEB) to a 1.5 mL tube. Once prepared, 190 µL of the ligation mix was pipetted into each well of an

8-well strip. From the 8-well strip, 15 µL of the ligation mix was transferred to the wells of the linker-barcode plate 1 using a multichannel pipette. To avoid cross-contamination, the pipette tips were discarded after each transfer.

The linker-barcode-ligation mix plate was incubated for 30 minutes at 37°C with 300 rpm shaking in a ThermoMixer C (Eppendorf). Initially, no heat treatment was performed after this stage, but it was introduced in later iterations and done at 65°C for 20 minutes to ensure the inactivation of any unwanted reactions after ligation.

To prevent hybridisation between nonbarcoded DNA and linkers during the pooling step, a blocking of unbound linkers is required. Three blocking oligos, corresponding to linkers 1, 2 and 3, were diluted to a concentration of 10 µM:

Blocking 1: TAATTGGAGCGTGGTGTTCTGAGCAAGGGT

Blocking 2: GTGGCCGATGTTTCGCATCGGCGTACGACT

Blocking 3: ATCCACGTGCTTGAGGGCCAGAGCATTCG

A blocking mixture was prepared by mixing 66.6 µL of 1st blocking oligonucleotide with 138.8 µL of 10x T4 Ligase buffer and 350 µL of sterile H2O. 66 µL of the blocking mix was pipetted into each well of an 8-well strip, and 5 µL of the blocking mix was added to each well of the linker-barcode-ligation mix plate, using fresh pipette tips for each transfer. The plate was incubated as previously.

After the blocking incubation, the contents of each well were pooled and subjected to a washing procedure: two washes were done with 0.1% SDS followed by three washes with 1xTE buffer. Initially, these washing steps were not part of the protocol, but they were added later to improve the removal of excess reagents and unhybridized oligonucleotides, and to prevent clumping of beads.

The second and third rounds of the split-pool barcoding protocol followed the same steps as the first step, but in the second round the ligation mix was pipetted into linker-barcode plate 2, and the blocking mix was done by mixing 88.8 µL 2nd blocking oligonucleotide with 150 µL 10x T4 Ligase buffer and 316 µL sterile H2O. In the third round, the ligation mix was pipetted into linker-barcode plate 3, and the blocking mix varied from the second blocking mix with the use of a 3rd blocking oligonucleotide.

After the washing steps of the last round, a NaOH treatment was done to remove the excess linker, barcode, and blocking oligos from the barcoded DNA. The protocol followed the same steps as in section 2.2.4. The beads were resuspended in an appropriate volume of 1xTE buffer depending on their amount and their structure was checked with a microscope (Evos Floid Imaging System, Invitrogen).

The semi-permeable capsule split-pool barcoding protocol followed the methods described above, with the exception that all washes between ligation rounds were done with 1xTE buffer + 0.1% Tween 20.

2.5.3 DNA Releasing

To detach the barcoded DNA from the polyacrylamide beads, the barcoded beads were mixed with 5-10 µL 1,000 units/mL USER enzyme (NEB) and incubated at 37°C for 1 hour. After incubation, the sample was vortexed for 1 minute and centrifuged for 1 minute at 13,000 rcf. The supernatant (barcoded DNA) was collected in a separate 1.5 mL tube.

To release the barcoded DNA from the semi-permeable capsules, the capsules were broken with a Release Reagent from the SPC Generation Kit (Atrandi Biosciences). The volume of the capsules was adjusted to 100 µL and the sample was divided in half to separate 1.5 mL tubes. For each tube, 2 µL of Release Reagent was added, and the mixture was incubated for 5 minutes in room temperature until the sample was transparent. Next, 48 µL of sterile H2O was added to get a total mixture of 100 µL per tube. 80 µL of HighPrep^{TM} PCR magnetic beads (MagBio Genomics) was added and thoroughly mixed. The sample was incubated for 5 minutes at room temperature and after incubation, the sample was briefly spun down and placed on a magnetic rack. While keeping the tube on the magnet, the supernatant was removed. 180 µL of 80% ethanol was added without disturbing the pellet, incubated for approximately 30 seconds at room temperature, and removed. The ethanol wash step was repeated once. Following the ethanol washes, the sample was briefly spun down again and placed back on the magnetic rack. Any leftover supernatant was pipetted off, and the beads were dried for 1 minute. Once dried, the tube was removed from the magnetic rack, and the pellet was resuspended in 27 µL of nuclease-free sterile H2O. The sample was incubated for 5 minutes at room temperature to elute the DNA from the beads. After incubation, the beads were pelleted on a magnet until the eluate

was clear and transparent. Lastly, the clear eluate containing the released barcoded DNA was transferred into a clean 1.5 mL tube.

2.5.4 16S-Barcode PCR

To generate sufficient quantities of barcoded DNA for subsequent analysis, a PCR run was done using primers designed to target the 16S rRNA gene region V4 with the barcoded end:
16S_519F: `CAGCMGCCGCGGTAATWC`
BC_R: `TCTCCAAATGGGTCATGATC`

A 2x volume PCR reaction mix was prepared by adding 8 µL of 5x Phusion HF buffer (NEB), 0.8 µL of 10mM dNTP Solution Mix (NEB), 2 µL of each primer, and 0.4 µL of 2,000 units/mL Phusion® Hot Start Flex DNA Polymerase (NEB) in a 1.5 mL tube. Following reaction mix preparation, 6.6 µL of the reaction mix was combined with 13.4 µL of barcoded DNA in a 0.2 mL PCR tube.

The PCR amplification was carried out using the program provided in Table 4. The preamplification was experimented with 25 and 20 cycles. All subsequent experiments were carried out using 20 cycles.

Table 4. PCR Program Parameters for the amplification of barcoded 16S rRNA gene region V4. The PCR program was tested with two different cycle amounts: 25 and 20.

Step	Temperature (°C)	Time	Cycles
Beginning temperature	98	30 s	1
Denaturation Annealing Extension	98 61 72	10 s 30 s 15 s	25/20
Final extension	72	5 min	1
Storing temperature	4		

The amplified barcoded DNA was purified using the Monarch® PCR & DNA Cleanup Kit (NEB) according to the manufacturer's protocol (Appendix 2) and eluted in 20 µL of sterile H2O. The concentration was measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific). DNA quantity and quality were deemed sufficient for further processing.

2.5.5 Frameshift PCR

Frameshift PCR was performed using primers with four forward primers and four reverse primers with 0-3 N-nucleotides between the primer binding site and the overhang for the MiSeq indexing primer (Appendix 5). The binding sites were the same as in the primers used in the 16S-Barcode PCR.

For example:

16S_519F frameshift 2:

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNCAGCAGCCGCGGTAATAC

S-P-BC frameshift 2:

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNCATCTTCTCCAAATGGGTCATGAT

A 3x volume PCR reaction mix was prepared by adding 12 µL of 5x Phusion HF buffer (NEB), 1.2 µL of 10mM dNTP Solution Mix (NEB), 3 µL of primer mix (concentration of each primer in mixture was 2.5 µM), 0.6 µL of 2,000 units/mL Phusion® Hot Start Flex DNA Polymerase (NEB), and 10.2 µL of sterile H2O in a 1.5 mL tube. Following the reaction mix preparation, 10 µL of the reaction mixture was divided into two separate 0.2 mL PCR tubes. 10 µL of the pre-amplified barcoded DNA was added to one reaction mix and 10 µL of sterile H2O to another for negative control.

The PCR amplification was carried out using a specified program, provided in Table 5. The annealing temperature was increased by 0.5°C each round. The preamplification was experimented with 25, 15, and 10 cycles, with subsequent experiments carried out using 10 cycles.

Table 5. PCR Program Parameters for the incorporation of frameshift mutations into the barcoded 16S rRNA gene region V4. The PCR program was tested with three different cycle amounts: 25, 15, and 10.

Step	Temperature (°C)	Time	Cycles
Beginning temperature	98	30 s	1
Denaturation Annealing Extension	98 63 (+0.5°C/round) 72	10 s 30 s 15 s	25/15/10
Final extension	72	5 min	1
Storing temperature	4		

After amplification, gel electrophoresis was performed, gel images were captured, the bands corresponding to the expected amplicon size, approximately 485 bp, were excised from the gel and purified, and the concentration was measured as detailed in section 2.2.1.

2.6 Sequencing

2.6.1 Index PCR

Various indexing oligonucleotides were used throughout the experiment as the number of samples and sequencing runs increased. All used sequencing indices are provided in Appendix 6.

The volume of the reaction mixture varied depending on the amount of sequencing samples. However, each sample had four replicates, so a quadrupled amount of reaction mix was made per sample. For one sample, the quadrupled PCR reaction mix was made by mixing 16 µL of 5x Phusion HF buffer (NEB) with 1.6 µL 10mM dNTP Solution Mix (NEB), 0.8 µL 2,000 units/mL Phusion® Hot Start Flex DNA Polymerase (NEB) and 45.6 µL sterile H2O in a 1.5 mL tube. 64 µL of reaction mix per sample was taken to a new 1.5 mL tube, with 8 µL of the template (barcoded DNA) and 4 µL of 10 µM forward (i5) and reverse (i7) index oligonucleotide added. The mixes were then split into four 20 µL reactions per sample.

The PCR amplification was carried out using a specified program, provided in Table 6.

Table 6. PCR Program Parameters for adding i5 and i7 indices to the samples.

Step	Temperature (°C)	Time	Cycles
Beginning temperature	98	30 s	1
Denaturation Annealing Extension	98 55 72	10 s 20 s 20 s	8
Final extension	72	5 min	1
Storing temperature	4		

After amplification, the PCR replicates were pooled, and 20 µL of the pooled sample was run on a gel. The gel images were captured, with the bands corresponding to the amplified DNA size, 558 bp, excised from the gel and purified as detailed in section 2.2.1.

2.6.2 Library Preparation

Prior to library preparation, the concentration (ng/µL) of indexed samples was measured using a Qubit 2.0 Fluorometer (Thermo Fisher) with a dsDNA High Sensitivity Assay kit (Thermo Fisher). The purified samples were mixed in equimolar concentration to a 4 nM library with sterile H2O. The concentration of the prepared library was checked using the Qubit 2.0 fluorometer (Appendix 7).

The library was denatured and diluted to a final concentration of 10 pM for MiSeq sequencing, following the Illumina MiSeq System Denature and Dilute Libraries Guide, "Protocol A: Standard Normalization Method" (page 5). Briefly, 5 µL of 4 nM library was combined with 5 µL 0.2 N NaOH, vortexed, and centrifuged with uniCFUGE 2 Centrifuge for 1 min at 4000 rpm. After centrifuging, the sample was incubated at room temperature for 5 minutes. 990 µL of 4°C HT1 buffer was added to the denatured library and the library was diluted to 10 pM by combining 300 µL of the library with 300 µL of HT1 buffer.

2.6.3 MiSeq Sequencing

Sequencing was performed on the Illumina MiSeq™ system using the MiSeq® Reagent Nano Kit v2 (500 cycles) according to the steps outlined in the Illumina MiSeq System Guide, specifically in Chapter 3: Sequencing (pages 17-26). The sequencing run was configured for 251 bp paired-end reads. For the analysis, the GenerateFASTQ - 3.1.0 module was used with adapter trimming.

2.7 Bioinformatics and Data Analysis

The raw FASTQ files from the sequencing were analysed using an ad hoc Python script (version 3.10.10), identifying the reads with expected barcode structure, recovering the barcode and unique molecular identifier sequences from these reads and matching sequences specific to

HAMBI105, HAMBI262 and HAMBI403 strains in the reads. Data visualisation was done using ggplot2, version 3.5.1, (Wickham 2016) in the R programming language, version 4.4.1, (R Core Team 2023).

3 Results

3.1 Microscopic Check of Structure and Integrity

3.1.1 Polyacrylamide Beads

The beads were microscopically examined (Figure 4) and confirmed to have formed successfully with structural integrity, and effective encapsulation of DNA/cells. Amplification of the 16S rRNA gene V4 in emulsion PCR could be seen with a fluorescence microscope using SYBR® Safe DNA Gel Stain. After split-pool barcoding, the beads were also microscopically examined and confirmed to have remained intact without clumping together or breaking apart. Figure 4 illustrates the appearance of polyacrylamide beads, after undergoing DNA encapsulation (A), cell encapsulation (B), emulsion PCR (C), and the split-pool barcoding protocol (D).

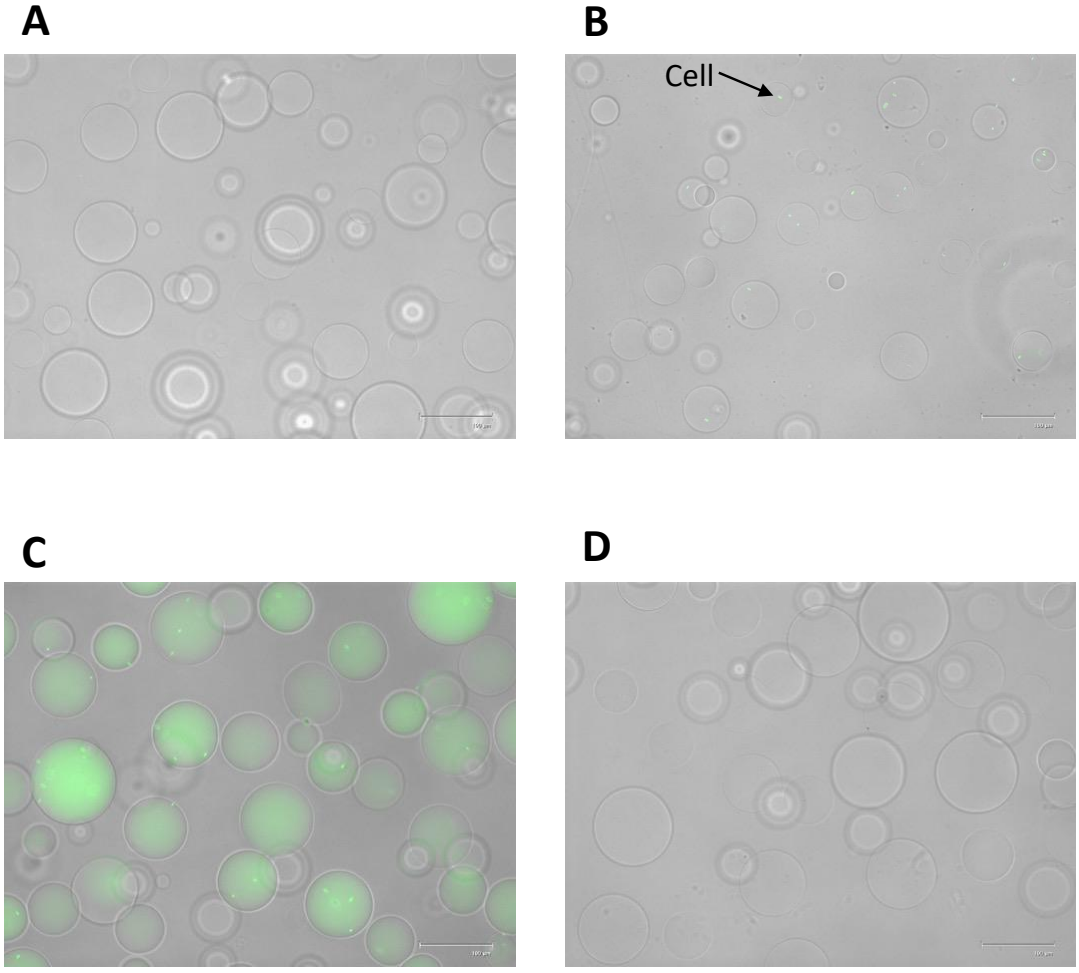


Figure 4. Microscopic examination of polyacrylamide beads at various stages of the experimental protocol. A) Synthetic DNA containing polyacrylamide beads. B) Microbial cell-containing beads, that have been treated with SYBR® Safe DNA Gel Stain (Invitrogen). Cells appear as green, fluorescent dots inside the beads. C) Beads

after the amplification of the 16S rRNA V4 gene region in emulsion, that have been treated with SYBR® Safe DNA Gel Stain (Invitrogen). Amplified DNA bound to the beads appear as green fluorescence inside the beads. D) Beads after split-pool barcoding.

3.1.2 Semi-Permeable Capsules

The formation and structural integrity of the capsules and encapsulation of cells were confirmed microscopically (Figure 5). The presence of encapsulated cells was verified, and the success of the amplification of the 16S rRNA gene region V1-4 was confirmed using a fluorescence microscope with SYBR® Safe DNA Gel Stain. A reduction in fluorescence indicated the success of the lambda exonuclease treatment which the capsules underwent to produce single-stranded DNA. The capsules were also examined and confirmed to remain intact throughout the split-pool barcoding process. Figure 5 illustrates the appearance of semi-permeable capsules, after cell encapsulation and capsule formation (A), 16S PCR (B), lambda exonuclease treatment (C), and the split-pool barcoding protocol (D).

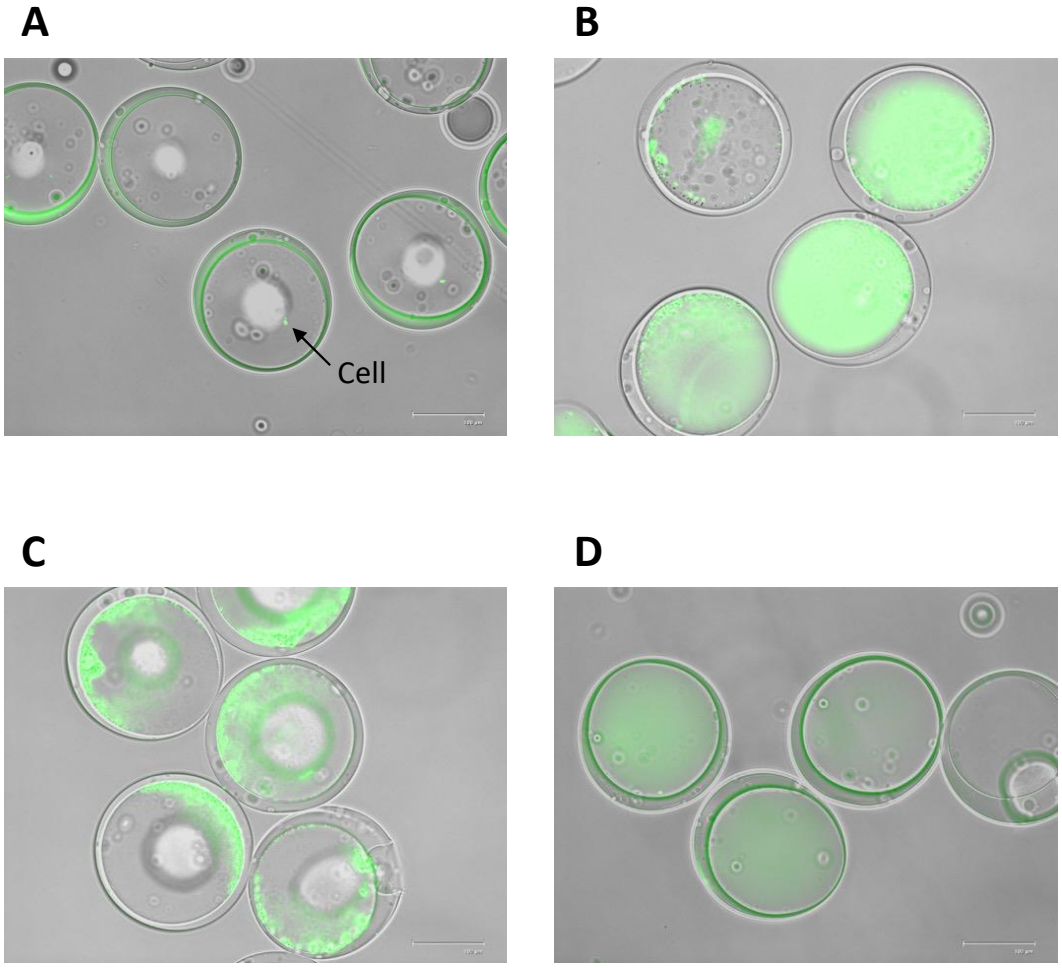


Figure 5. Microscopic examination of semi-permeable capsules at different stages of the experimental protocol. A) Capsules immediately after formation and cell encapsulation. Cells appear as green, fluorescent dots

inside the capsules. B) Capsules after amplification of the 16S rRNA gene region V1-4. Amplified DNA appears as green fluorescence inside the capsules. C) Capsules after lambda exonuclease treatment. Capsules exhibit reduced fluorescence after Lambda treatment. D) Capsules after split-pool barcoding. All samples treated with SYBR® Safe DNA Gel Stain (Invitrogen).

3.2 Frameshift PCR Gel Electrophoresis Results

3.2.1 Polyacrylamide Beads

Throughout the optimisation of the PCR protocols, amplicons from frameshift PCR were consistently analysed using gel electrophoresis to verify amplification and confirm the correct amplicon size, approximately 485 bp (Figure 6). Through gel electrophoresis, the reduction of unwanted by-products in the gel could be assessed and the efficient amplification of the desired product could be confirmed even when the number of PCR cycles decreased.

In the initial experiments with synthetic polyacrylamide beads before optimisation (Figure 6, Image A), the right-sized product appeared smeared with a lot of smaller side products. The smearing was very pronounced in both 10,000 beads and 1,000 beads samples. In the frameshift PCR optimisation experiment with synthetic polyacrylamide beads (Figure 6, Image B), the amplified product was confirmed to be correct in size, and there were much fewer small side products than in the initial experiment. Furthermore, side products were more prevalent in the 10,000 beads samples than in the 1,000 beads samples and 15 cycles produced more side products compared to 10 cycles. The rest of the frameshift PCRs, in subsequent experiments, were conducted with 10 cycles for 1,000 bead samples.

16S-barcode PCR was optimised with cell-containing beads by reducing the number of cycles from 25 to 20. A reduced number of cycles was noted to amplify the correct-sized product, with considerably less side product than in earlier experiments (Figure 6, Image C). Subsequent 16S-barcode PCRs were conducted with 20 cycles. In the emulsion PCR optimisation, the number of emulsion PCR cycles was reduced from 32 to 20. This modification also lessened the amount of side product (Figure 6, Image D), but the amount of amplicon remained sufficient.

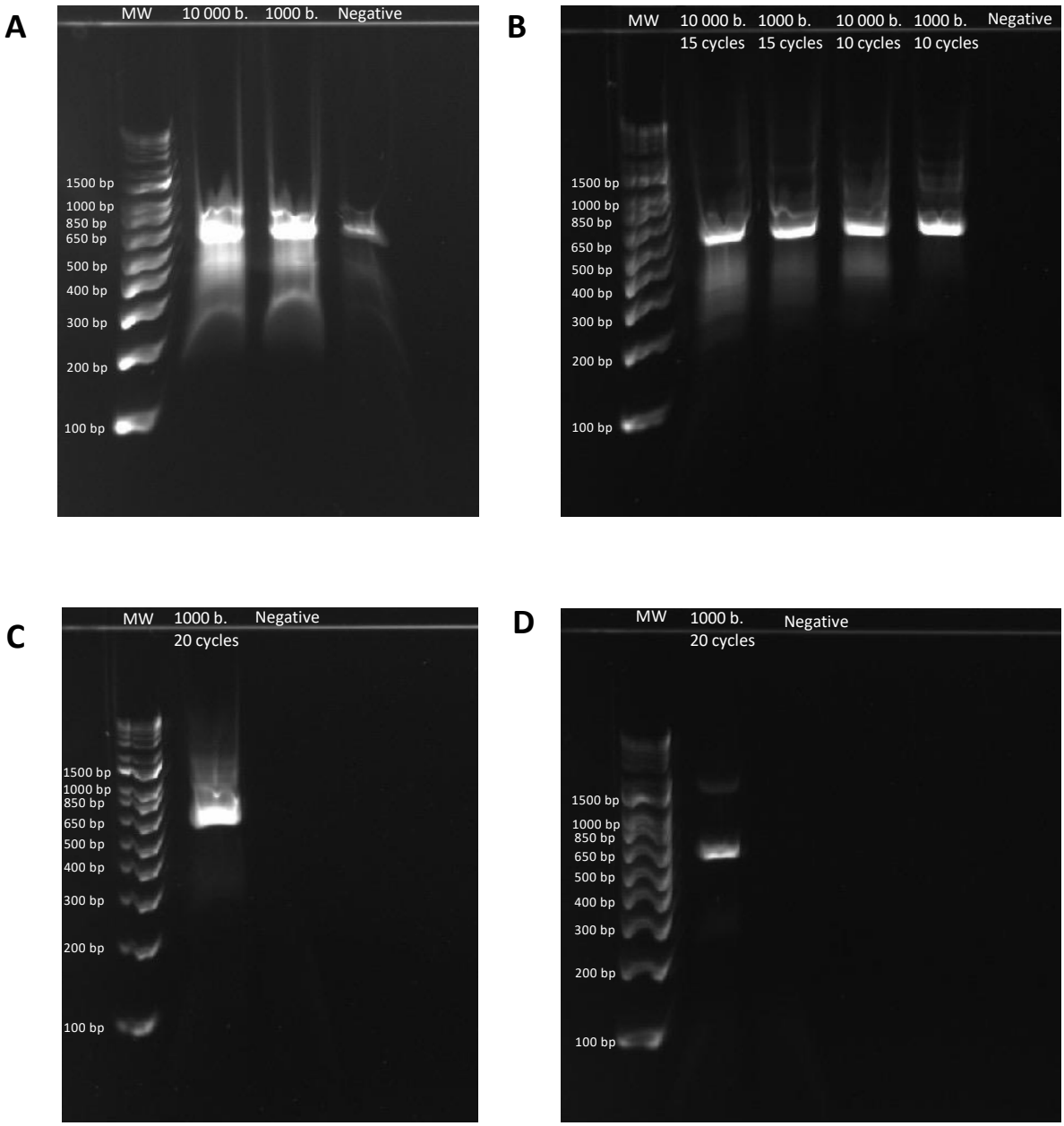


Figure 6. Gel electrophoresis images of PCR optimisations for split-pool barcoded DNA. The expected amplicon length is approximately 485 bp. A) Initial experiment's frameshift PCR gel image with no optimisations made. The DNA of 10,000 and 1,000 synthetic beads was used. Both samples and the negative control appeared smeared with side products. B) Frameshift PCR optimisation gel image. Optimisation experiments were done with 15 and 10 cycles using the DNA of 10,000 and 1,000 synthetic beads. Side products were less prevalent in lower cycles. C) 16S-barcode PCR optimisation gel image. 20 cycles with 1,000 cell-containing beads showed considerably fewer side products. D) Emulsion PCR optimisation gel image. 20 cycles with 1,000 cell-containing beads showed minimal amplification of side products. MW = molecular weight. 1Kb Plus DNA ladder. 2% agarose E-Gel. Gel run 48V, 10 min. Due to the uneven running of the gel, the molecular weight is only an approximate estimate of the DNA size.

3.2.2 Semi-Permeable Capsules

Gel electrophoresis was performed for the frameshift PCR amplicon of semi-permeable capsule DNA to verify the amplification of the product and confirm the correct size (485 bp). Despite using the previously optimised PCR conditions for both 16S-barcode PCR and frameshift PCR, the correct-sized band appeared very smeared and accompanied by a significant amount of larger and smaller side products (Figure 7).

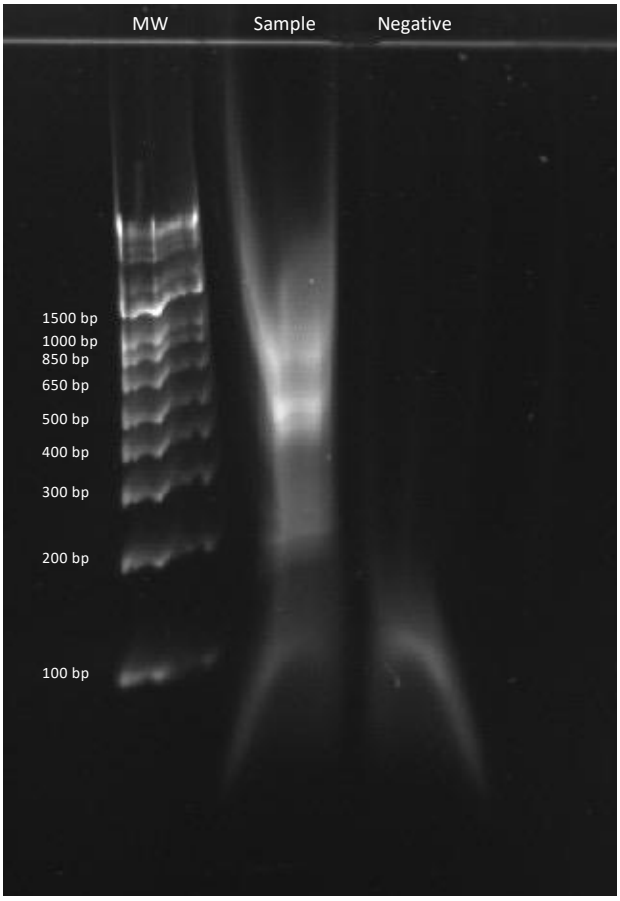


Figure 7. Gel electrophoresis image of frameshift PCR done for split-pool barcoded and 16S-barcode amplified semi-permeable capsule DNA. 20 cycles showed considerable amounts of side-products. MW = molecular weight. 1Kb Plus DNA ladder. 2% agarose E-Gel. Gel run 48V, 10 min.

3.3 Bioinformatic Visualisation of Optimisation

3.3.1 Polyacrylamide Beads

To assess optimisations made to the PCR protocols and to visualise the data from sequencing, scatter plots were used, where each dot represents barcoded DNA from an individual bead. The x- and y-axes correspond to the number of times a barcode sequence was found together with

a specific HAMBI 16S rRNA gene. An optimal result would be a clear 90-degree angle formed by the barcoded DNA fragments, indicating no signal crosstalk, as the beads were designed so that HAMBI105 and HAMBI403 or HAMBI262 and HAMBI403 are not present on the same bead and therefore should not share any barcodes.

The initial experiment, without any optimisations (Figure 8, Plot A), showed a distribution of dots forming an angle smaller than the desired 90 degrees with 1,000 synthetic beads. Similarly, the scatter plot for 10,000 synthetic beads displayed a dense cluster of dots, forming an angle much smaller than the desired 90 degrees. Due to the inability to reach efficient sequencing depth, visualisation of 10,000 beads was left out from subsequent experiments.

Following the initial experiment, optimisations were made to frameshift PCR to reduce signal crosstalk by decreasing PCR cycles from 25 cycles to 15 and 10 cycles. The results were visualised (Figure 8, Plot B) with 1,000 synthetic beads. The sample with 15 PCR cycles, showed a distribution of dots forming a smaller angle, than the desired 90 degrees. However, the sample with 10 PCR cycles formed a closer angle to 90 degrees.

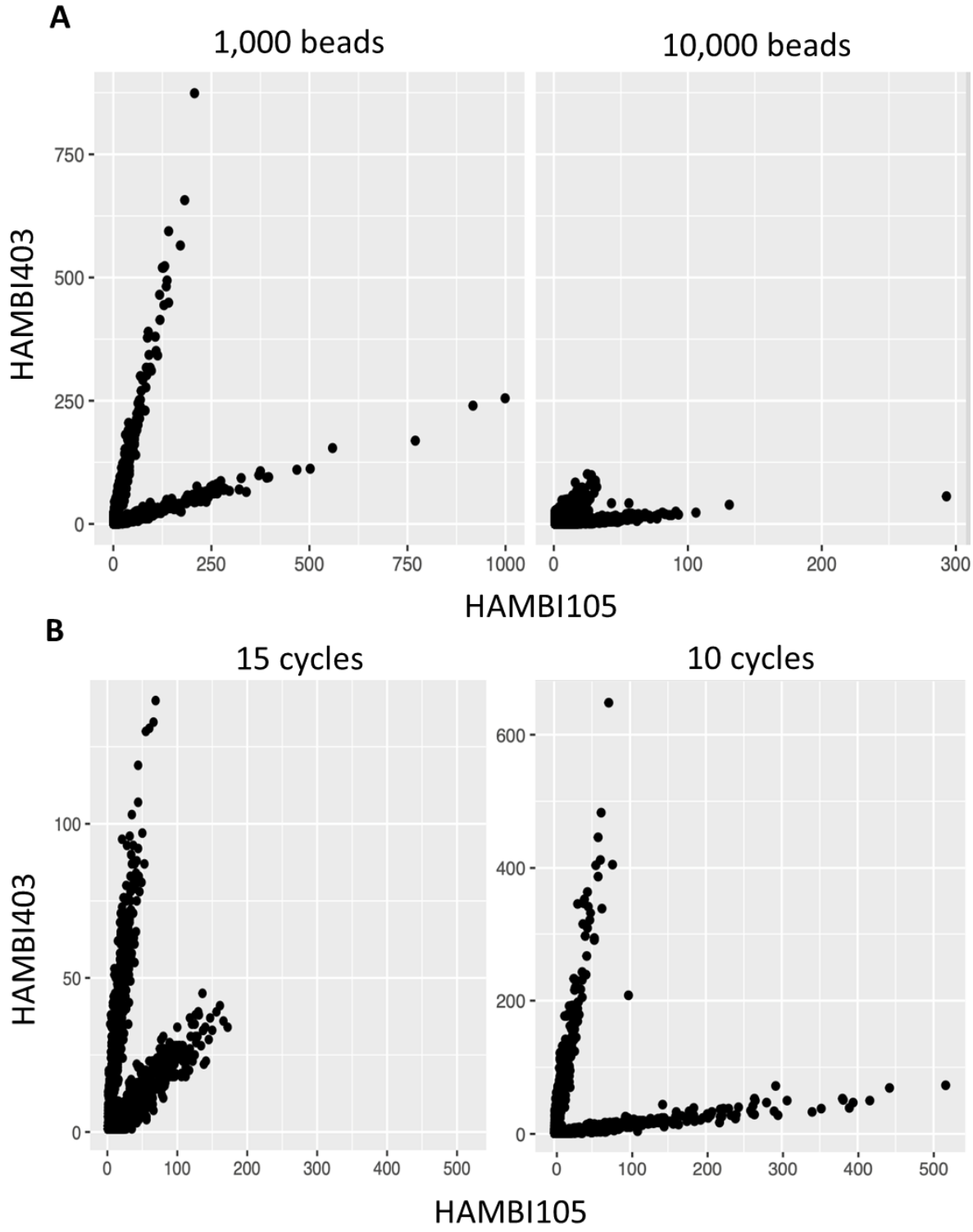


Figure 8. Scatter plots of the initial barcoding protocol with 1,000 and 10,000 synthetic polyacrylamide beads (A) and frameshift PCR optimisations with 15 and 10 cycles, using synthetic polyacrylamide beads (B). Each dot represents barcoded DNA from an individual bead. The x-axis indicates the number of times a barcode sequence was found together with the HAMBI105 16S rRNA gene, and the y-axis indicates the number of times it was found together with the HAMBI403 16S rRNA gene. In the initial protocol, the angles formed by 1,000 and 10,000 beads were smaller than 90 degrees, indicating a need for further optimisations. In the frameshift

PCR optimisation experiments, the sample with 15 PCR cycles did not achieve 90 degrees. However, the sample with 10 PCR cycles formed a closer angle to 90 degrees.

To further reduce signal crosstalk, optimisations were made to 16S-barcode PCR and emulsion PCR. These optimisations were performed on cell-containing beads, with HAMBI262 and HAMBI105 combined in one set of beads and HAMBI403 in another.

In the 16S-barcode PCR optimisation (Figure 9), the number of PCR cycles was reduced from 25 to 20. The distribution of dots formed a clean angle but did not achieve the desired 90 degrees. In the emulsion PCR optimisation, the number of PCR cycles was reduced from 32 to 20. A clean 90-degree angle was obtained, showing no crosstalk. This was achieved by using optimised PCR conditions in all previous PCR steps (16S-barcode PCR, and frameshift PCR), and emulsion PCR.

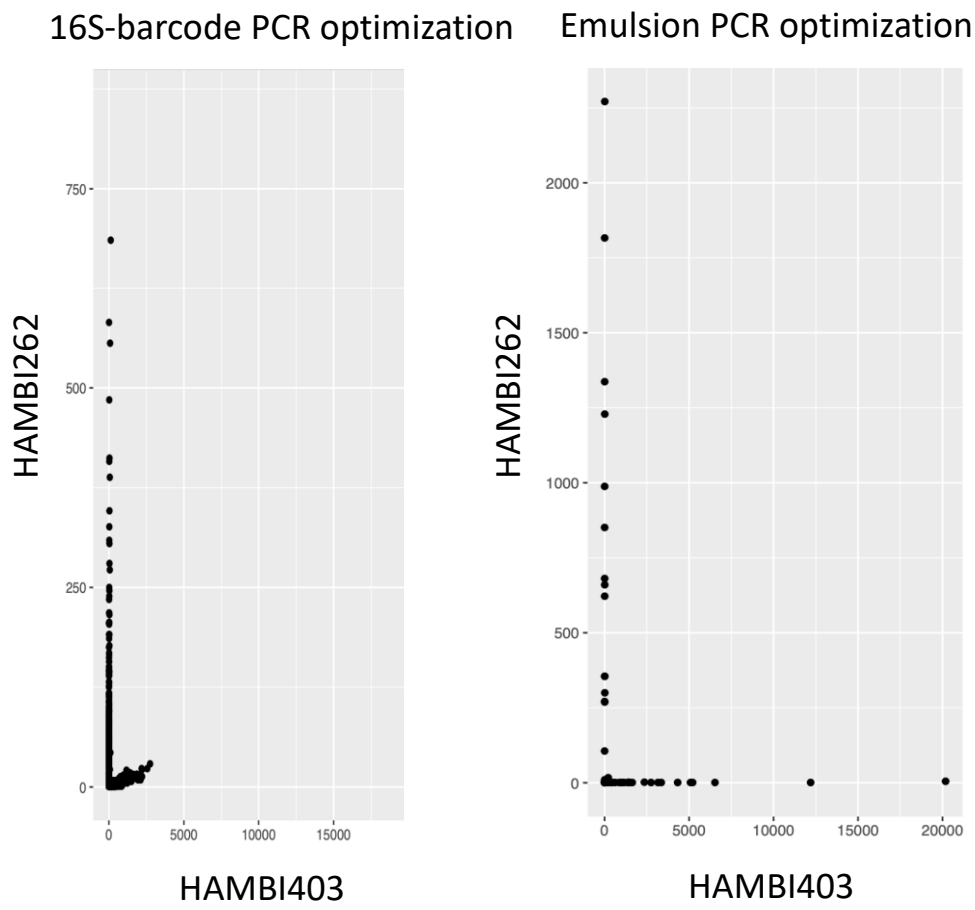


Figure 9. Scatter plots of the optimisations in 16S-barcode PCR and emulsion PCR, using cell-containing beads. Each dot represents barcoded DNA from an individual bead. The x-axis indicates the number of times a barcode sequence was found together with HAMBI403 and, the y-axis together with HAMBI262. In the 16S-barcode PCR optimisation, the angle was clean and closer to the desired 90 degrees. In the emulsion PCR optimisation, a 90-degree angle was achieved.

3.3.2 Semi-Permeable Capsules

The performance of the split-pool barcoding protocol using semi-permeable capsules was also assessed by visualising the data from sequencing through a scatter plot (Figure 10). Despite using optimised PCR conditions, the scattered distribution of dots formed an angle that was much smaller than the desired 90 degrees, and the rest of the dots clustered close to the origin point (0,0) of both axes.

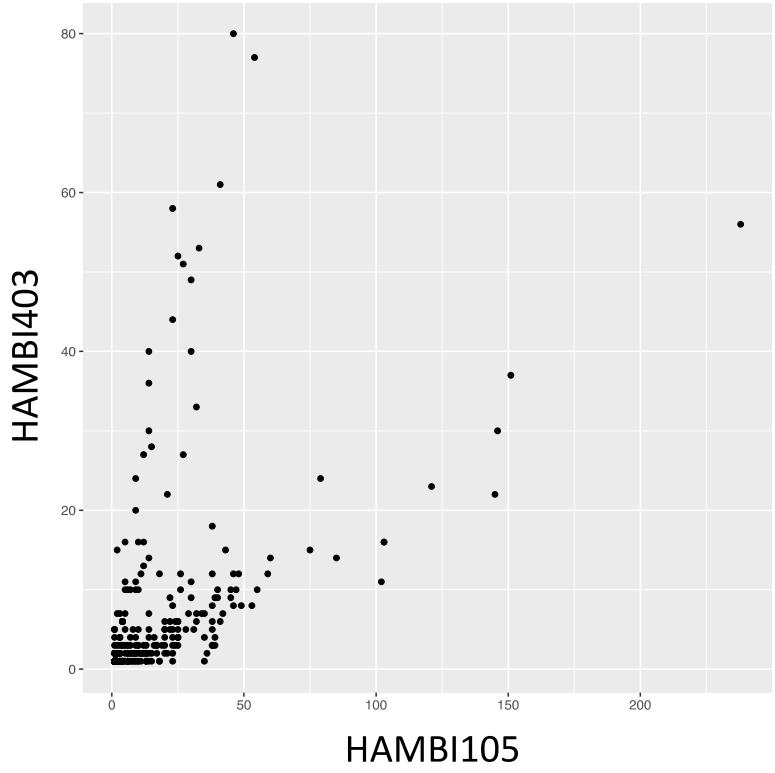


Figure 10. Scatter plot of the barcoded DNA of semi-permeable capsules. Each dot represents barcoded DNA from an individual bead. The x-axis indicates the number of times a barcode sequence was found together with HAMBI105 and, the y-axis together with HAMBI403. The distribution of dots is clustered near the origin, and the lines that were formed did not create the desired 90-degree angle.

4 Discussion

4.1 Optimisation of the Split-Pool Barcoding Protocol

The successful development and optimisation of the split-pool barcoding protocol, done in this study, represents a significant advancement in the field of microbial community analysis. The feasibility of the optimised protocol and the ability to achieve precise and effective microbial single-cell barcoding in a way that is cost-effective and easily repeated was successfully demonstrated. Key findings highlight that the optimised protocol, using polyacrylamide beads, allows for accurate tagging and identification of specific gene targets, e.g., 16S rRNA V4, within a synthetic mock bacterial community, laying a solid foundation for future applications of the technique.

4.2 Key Findings in Protocol Optimisation

Reducing the number of PCR cycles was essential in optimising the protocol, as it resulted in decreased signal cross-contamination between individual cells and improved fidelity of single-cell barcoding, thereby enhancing the reliability of sequencing outcomes. This finding aligns with studies that emphasise the importance of optimising PCR conditions to improve the accuracy of amplification for sequencing. According to Fu et al. (2018), while library preparation and sequencing methods usually come with predefined error percentages, it is a generally acknowledged fact that increased PCR amplification can lead to more artefactual duplicates. Therefore, PCR cycles are frequently modified based on the quantity, source, and quality of the sample. Here, the PCR cycles were systematically reduced, starting with frameshift PCR, then 16S-barcode PCR, and finally emulsion PCR. The reduced amplification of side products could be seen in gel images, which showed the correct product size becoming more prominent with each optimisation step, while side products became less prominent. The optimal number of cycles was determined by ensuring sufficient amplification yield while minimising crosstalk and signal mix-ups. Signal mixing was also checked bioinformatically using scatter plots, where the angles created by barcoded fragments provided perspective of how much each PCR optimisation affected signal mix-ups.

In addition to cycle optimisation, managing unwanted hybridisations between primers, barcode oligos, and linker oligos during ligation was critical for ensuring data integrity. Unwanted

ligations can lead to non-specific hybridisations, which reduce the efficiency and fidelity of downstream processes, such as PCR amplifications, and introduce noise into the sequencing data (Chalaya 2004). To mitigate these issues, heat treatments were implemented and blocking oligos were used to prevent the hybridisation of complementary sequences that could form between primers and other oligos.

Efforts to optimise the protocol with semi-permeable capsules were unsuccessful. In the initial phases, difficulties were encountered with the microfluidic fabrication of the capsules. These challenges were mainly related to maintaining the structural integrity of the capsules during the production process. Leonaviciene et al. (2020) state that achieving consistent capsules with perfect structural properties, in terms of the layering and integrity of the shell around the core, remains a challenge. Despite optimising ligation and PCR conditions, the capsules also faced problems with signal mix-ups and incorrect hybridisations between barcodes and unknown sequences. This could potentially be caused by cross-contamination of barcodes (Clark et al. 2019), mistakes made during the experiment, or problems with the structural integrity of the capsules. Due to time constraints, semi-permeable capsules were ultimately considered unfeasible for this specific application, and their use was abandoned.

4.3 Impact of Study

Current technologies face limitations in observing complex microbial phenomena, such as spatial structures, and HGT, among other interactions between taxa, due to the lack of resolution that is needed to study these processes (Brito 2021; Cordero & Datta 2016).

The location and proximity of microbial cells affect how the community is organised and how it functions (Yanni et al. 2019). Techniques, such as FISH, can provide information about microbes in close proximity, within single-cell level detail, but are not quantitative enough to provide a comprehensive understanding of underlying genetic interactions (Moter & Göbel 2000). Horizontal gene transfer of mobile genetic elements, such as plasmids, is another key process that current methods struggle to accurately capture (Tokuda & Shintani 2024). According to Brito (2021), plasmids have a central role as carriers of antibiotic-resistance genes between microbial cells. Traditional metagenomic approaches or culturing techniques are limited in their ability to distinguish which cells harbour these genes.

Through the development and optimisation of the split-pool barcoding protocol, this study makes significant progress for understanding the physical interactions within microbial communities and provides a step forward in addressing earlier challenges. Unlike traditional methods, which have relied on bulk approaches (Imdahl & Saliba 2020), the optimised protocol enables high-resolution single-cell analysis, providing a tool for investigating genetic elements within individual cells and helps explore underrepresented genes and their roles. Weaver et al. (2014) highlight that increasing the throughput of single-cell genomic methods allows for the analysis of more cells and improves the statistical power of the observations. With higher throughput, a better understanding of rare events, such as transient interactions or the presence of low-abundance gene variants can be gained, which leads to new perspectives in microbial ecology and evolution.

Compared to other existing single-cell sequencing methods, such as FACS and microfluidics-based single-cell sequencing, the use of polyacrylamide beads offers advantages. According to Zilionis et al. (2017), FACS is effective for sorting and analysing individual cells but might be limited by its cell throughput, making the analysis of heterogeneity difficult. Microfluidic approaches, on the other hand, have improved DNA yield (Streets et al. 2014) and lower technical variation (Wu et al. 2014) but they can be challenging to apply to a larger number of cells, as commercial platforms allow for the analysis of only tens to hundreds of cells per sample (Zilionis et al. 2017). The protocol optimised in this study with polyacrylamide beads is more scalable, enabling the analysis of a large number of cells.

Polyacrylamide beads offer a reliable alternative to microfluidics-based approaches, primarily due to their structural advantages and ease of use. While microfluidics systems offer precision, they require specialised equipment and technical expertise to operate (Battat et al. 2022). In contrast, polyacrylamide beads are simpler and less costly to produce and manipulate, which enhances the reliability and reproducibility of the results. Additionally, using polyacrylamide beads for the protocol allows for easy troubleshooting and adaptation to different experimental conditions.

4.4 Challenges and Limitations

Despite the advancements and successful optimisation of the split-pool barcoding protocol, several challenges were encountered. One of the primary challenges faced during the early

stages of the experiments was the amplification of side products during PCR, with little to no amplification of the expected product. The presence of these side products and lack of amplification of the correct sized product complicated the interpretation of results and reduced the efficiency of the overall protocol. Another significant limitation encountered was the ineffectiveness of barcoding in some instances. Before implementing ligase heat inactivation and reducing PCR cycles, results showed that barcodes sometimes hybridised with incorrect regions, leading to signal mixing and incorrect sequencing results.

Optimising the protocol was a laborious and time-intensive process that required a substantial number of resources, as each step had to be assessed and possibly refined. Flaherty & Davis (2015) state that protocols are most commonly optimised by focusing on individual sections of the process until optimal conditions for the whole workflow are found. This is done because it's simpler and provides understanding of the optimisation process at a more in-depth level. This type of optimisation approach used in the study demanded significant time and effort to identify the most effective conditions for individual parts of the protocol. Although it was successfully optimised, early attempts, such as changing reagent concentrations and filtering the beads between ligation steps, often resulted in minimal improvements.

Variability caused by human error in biological research is a known and documented issue, that can largely affect the repeatability and accuracy of experimental outcomes (Pusztai et al. 2013). The manual nature of the workflow of split-pool barcoding added further challenges, as it made the protocol prone to error, making consistent results difficult to achieve. Variability caused by errors was evident in the inconsistent yield of polyacrylamide beads produced, and the number of beads that could still be obtained for downstream processes after undergoing emulsion PCR and the split-pool barcoding protocol. Relying on visual estimation using microscopy and subjectively assessing the number of beads also added uncertainty in determining the exact number of beads used in the following steps.

4.5 Future and Practical Applications

The successful optimisation of the split-pool barcoding protocol using a synthetic mock bacterial community represents a promising step forward in single-cell microbial genomics. While the advancements made in this study provide a foundation for a more detailed exploration of microbial interactions and community dynamics, its true potential lies in its future

applications and continued refinement. The next step would be applying the method to increasingly complex bacterial communities, such as wastewater and soil. Wastewater and soil microbial communities are highly diverse and abundant, containing many unidentified taxa at various physiological states (Hill et al. 2000; Wagner et al. 2002). These complex communities present a variety of challenges, which would test the robustness and adaptability of the protocol.

Another future direction would be to expand the genomic targets of the protocol beyond the 16S rRNA gene to include other genes of interest, for example, those involved in antibiotic resistance. Frieri et al. (2017) define antibiotic resistance as a global health challenge, where bacteria evolve to withstand the effects of drugs that were previously effective against them. This resistance makes bacterial infections harder to treat and can lead to severe health problems, including increased disease and death. Addressing antibiotic resistance is critical because the effectiveness of current antibiotics is diminishing as bacteria develop resistance. Resistance is also increased through overuse and misuse of antibiotics and the current development of antibiotics cannot keep pace with the emergence of resistant strains, making treatment even more challenging. The distribution and movement of antibiotic resistance genes within microbial populations could be examined using the split-pool barcoding protocol, giving insight into how resistance spreads through HGT and clonal expansion.

To increase the scalability of the protocol and decrease human error-caused variability, future efforts could focus on automating some steps of the workflow. These steps would include the ones that are most prone to error, for example, pipetting steps, such as preparing linker-barcode plates, moving the ligation mix to said plates, and sample pooling. According to Holland & Davies (2020), automation has greatly increased the productivity and quality rate in many industries. Automation would not only reduce error rates but would also enable processing of larger sample sizes and increase efficiency, making the protocol more suitable for high-throughput applications.

Incorporating unique molecular identifiers (UMI) into the sequencing workflow would be an effective strategy for minimising signal mix-ups and improving data accuracy. UMIs are short, unique sequences that are added to each DNA molecule before amplification. This unique tagging enables the distinction between true biological duplicates and artefactual duplicates, that arise from PCR amplification errors (Viailly et al. 2021). For the split-pool barcoding protocol, implementing UMI filtering would enhance the reliability of the results, by reducing

the impact of amplification biases, providing a clearer picture of the genetic diversity and abundance within the sample.

Other interesting future directions would be integrating the split-pool barcoding protocol with spatial genomics techniques for providing a multidimensional view of microbial communities by linking genetic information with spatial organisation. This could be particularly valuable for studying biofilms or other complex microbial communities.

4.6 Conclusions

This study presents the successful optimisation of the split-pool barcoding protocol using polyacrylamide beads, offering a novel approach for high-throughput single-cell analysis of microbial communities. By refining key aspects of the protocol, including PCR and ligation conditions, the study provides a potential method that effectively tags and sequences targeted genes within complex microbial populations. These advancements not only address the limitations of existing single-cell sequencing technologies but also provide a more reliable and cost-effective alternative for studying complex microbial interactions and community dynamics.

While the study shows significant progress, it also shows areas for further improvement, such as expansion to additional genomic targets and automation of the workflow. Future work could focus on applying the protocol to more complex communities, testing its scalability, and exploring its practical applications in monitoring antibiotic resistance genes and other clinically relevant markers. Overall, this study provides a solid foundation for advancing single-cell genomic techniques, in both fundamental research and practical applications of microbial science.

Acknowledgements

I would like to sincerely thank my supervisors Manu Tamminen and Niina Smolander for all the hard work they've put into the planning and implementation of the study and their guidance and support throughout my thesis. I also want to thank the rest of the Tamminen group for being such great colleagues to work with. Finally, I would like to thank my partner, family and friends for their constant support during my thesis project.

References

Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nature Reviews Microbiology*, *13*(8), 497–508. https://doi.org/10.1038/nrmicro3491

Anthony J.F. Griffiths. (2024). DNA sequencing. *Encyclopedia Britannica*. https://www.britannica.com/science/DNA-sequencing

Battat, S., Weitz, D. A., & Whitesides, G. M. (2022). An outlook on microfluidics: the promise and the challenge. *Lab on a Chip*, *22*(3), 530–536. https://doi.org/10.1039/D1LC00731A

Blainey, P. C. (2013). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews*, *37*(3), 407–427. https://doi.org/10.1111/1574-6976.12015

Blattman, S. B., Jiang, W., Oikonomou, P., & Tavazoie, S. (2020). Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nature Microbiology*, *5*(10), 1192–1201. https://doi.org/10.1038/s41564-020-0729-6

Brehm-Stecher, B. F., & Johnson, E. A. (2004). Single-Cell Microbiology: Tools, Technologies, and Applications. *Microbiology and Molecular Biology Reviews*, *68*(3), 538–559. https://doi.org/10.1128/MMBR.68.3.538-559.2004

Brennan, M. A., & Rosenthal, A. Z. (2021). Single-Cell RNA Sequencing Elucidates the Structure and Organization of Microbial Communities. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.713128

Brito, I. L. (2021). Examining horizontal gene transfer in microbial communities. *Nature Reviews Microbiology*, *19*(7), 442–453. https://doi.org/10.1038/s41579-021-00534-7

Bustin, S. A., Benes, V., Nolan, T., & Pfaffl, M. W. (2005). Quantitative real-time RT-PCR – a perspective. *Journal of Molecular Endocrinology*, *34*(3), 597–601. https://doi.org/10.1677/jme.1.01755

Chalaya, T. (2004). Improving specificity of DNA hybridization-based methods. *Nucleic Acids Research*, *32*(16), e130–e130. https://doi.org/10.1093/nar/gnh125

Chen, Z., Chen, L., & Zhang, W. (2017). Tools for Genomic and Transcriptomic Analysis of Microbes at Single-Cell Level. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.01831

Clark, M. A., Stankiewicz, S. H., Barronette, V., & Ricke, D. O. (2019). Letter to the Editor – Detecting HTS Barcode Contamination. *Journal of Forensic Sciences*, *64*(3), 961–962. https://doi.org/10.1111/1556-4029.14027

Cordero, O. X., & Datta, M. S. (2016). Microbial interactions and community assembly at microscales. *Current Opinion in Microbiology*, *31*, 227–234. https://doi.org/10.1016/j.mib.2016.03.015

Dar, D., Dar, N., Cai, L., & Newman, D. K. (2021). Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science*, *373*(6556). https://doi.org/10.1126/science.abi4882

de Bourcy, C. F. A., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., & Quake, S. R. (2014). A Quantitative Comparison of Single-Cell Whole Genome Amplification Methods. *PLoS ONE*, *9*(8), e105585. https://doi.org/10.1371/journal.pone.0105585

Delley, C. L., & Abate, A. R. (2021). Modular barcode beads for microfluidic single cell genomics. *Scientific Reports*, *11*(1), 10857. https://doi.org/10.1038/s41598-021-90255-x

DeSantis, T. Z., Brodie, E. L., Moberg, J. P., Zubieta, I. X., Piceno, Y. M., & Andersen, G. L. (2007). High-Density Universal 16S rRNA Microarray Analysis Reveals Broader Diversity than Typical Clone Library When Sampling the Environment. *Microbial Ecology*, *53*(3), 371–383. https://doi.org/10.1007/s00248-006-9134-9

Ducklow, H. (2008). Microbial services: challenges for microbial ecologists in a changing world. *Aquatic Microbial Ecology*, *53*, 13–19. https://doi.org/10.3354/ame01220

Ehrlich, H. (1998). Geomicrobiology: its significance for geology. *Earth-Science Reviews*, *45*(1–2), 45–60. https://doi.org/10.1016/S0012-8252(98)00034-8

Evans, C. R., Kempes, C. P., Price-Whelan, A., & Dietrich, L. E. P. (2020). Metabolic Heterogeneity and Cross-Feeding in Bacterial Multicellular Systems. *Trends in Microbiology*, *28*(9), 732–743. https://doi.org/10.1016/j.tim.2020.03.008

Finishing the euchromatic sequence of the human genome. (2004). *Nature*, *431*(7011), 931–945. https://doi.org/10.1038/nature03001

Flaherty, P., & Davis, R. W. (2015). Robust Optimization of Biological Protocols. *Technometrics*, *57*(2), 234–244. https://doi.org/10.1080/00401706.2014.915890

Frieri, M., Kumar, K., & Boutin, A. (2017). Antibiotic resistance. *Journal of Infection and Public Health*, *10*(4), 369–378. https://doi.org/10.1016/j.jiph.2016.08.007

Fu, Y., Wu, P.-H., Beane, T., Zamore, P. D., & Weng, Z. (2018). Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics*, *19*(1), 531. https://doi.org/10.1186/s12864-018-4933-1

FUKUDA, K., OGAWA, M., TANIGUCHI, H., & SAITO, M. (2016). Molecular Approaches to Studying Microbial Communities: Targeting the 16S Ribosomal RNA Gene. *Journal of UOEH*, *38*(3), 223–232. https://doi.org/10.7888/juoeh.38.223

Garland, J. L., & Mills, A. L. (1991). Classification and Characterization of Heterotrophic Microbial Communities on the Basis of Patterns of Community-Level Sole-Carbon-Source Utilization. *Applied and Environmental Microbiology*, *57*(8), 2351–2359. https://doi.org/10.1128/aem.57.8.2351-2359.1991

Gawad, C., Koh, W., & Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, *17*(3), 175–188. https://doi.org/10.1038/nrg.2015.16

Haack, S. K., Garchow, H., Klug, M. J., & Forney, L. J. (1995). Analysis of factors affecting the accuracy, reproducibility, and interpretation of microbial community carbon source utilization patterns. *Applied and Environmental Microbiology*, *61*(4), 1458–1468. https://doi.org/10.1128/aem.61.4.1458-1468.1995

Hill, G. T., Mitkowski, N. A., Aldrich-Wolfe, L., Emele, L. R., Jurkonie, D. D., Ficke, A., Maldonado-Ramirez, S., Lynch, S. T., & Nelson, E. B. (2000). Methods for assessing the composition and diversity of soil microbial communities. *Applied Soil Ecology*, *15*(1), 25–36. https://doi.org/10.1016/S0929-1393(00)00069-X

Holland, I., & Davies, J. A. (2020). Automation in the Life Science Research Laboratory. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.571777

Hood, L. E., Hunkapiller, M. W., & Smith, L. M. (1987). Automated DNA sequencing and analysis of the human genome. *Genomics*, *1*(3), 201–212. https://doi.org/10.1016/0888-7543(87)90046-2

Hosokawa, M., & Nishikawa, Y. (2024). Tools for microbial single-cell genomics for obtaining uncultured microbial genomes. *Biophysical Reviews*, *16*(1), 69–77. https://doi.org/10.1007/s12551-023-01124-y

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Imdahl, F., & Saliba, A.-E. (2020). Advances and challenges in single-cell RNA-seq of microbial communities. *Current Opinion in Microbiology*, *57*, 102–110. https://doi.org/10.1016/j.mib.2020.10.001

Kamble, A., Sawant, S., & Singh, H. (2020). 16S ribosomal RNA gene-based metagenomics: A review. *Biomedical Research Journal*, *7*(1), 5. https://doi.org/10.4103/BMRJ.BMRJ_4_20

Kaminski, T. S., Scheler, O., & Garstecki, P. (2016). Droplet microfluidics for microbiology: techniques, applications and challenges. *Lab on a Chip*, *16*(12), 2168–2187. https://doi.org/10.1039/C6LC00367B

Kim, S., De Jonghe, J., Kulesa, A. B., Feldman, D., Vatanen, T., Bhattacharyya, R. P., Berdy, B., Gomez, J., Nolan, J., Epstein, S., & Blainey, P. C. (2017). High-throughput automated microfluidic sample preparation for accurate microbial genomics. *Nature Communications*, *8*(1), 13919. https://doi.org/10.1038/ncomms13919

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., & Kirschner, M. W. (2015). Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell*, *161*(5), 1187–1201. https://doi.org/10.1016/j.cell.2015.04.044

Konopka, A. (2009). What is microbial community ecology? *The ISME Journal*, *3*(11), 1223–1230. https://doi.org/10.1038/ismej.2009.88

Konopka, A., Oliver, L., & Jr., R. F. T. (1998). The Use of Carbon Substrate Utilization Patterns in Environmental and Ecological Microbiology. *Microbial Ecology*, *35*(2), 103–115. https://doi.org/10.1007/s002489900065

Kuchina, A., Brettner, L. M., Paleologu, L., Roco, C. M., Rosenberg, A. B., Carignano, A., Kibler, R., Hirano, M., DePaolo, R. W., & Seelig, G. (2021). Microbial single-cell RNA sequencing by split-pool barcoding. *Science*, *371*(6531). https://doi.org/10.1126/science.aba5257

Kuijpers, L., Hornung, B., van den Hout - van Vroonhoven, M. C. G. N., van IJcken, W. F. J., Grosveld, F., & Mulugeta, E. (2024). Split Pool Ligation-based Single-cell Transcriptome sequencing (SPLiT-seq) data processing pipeline comparison. *BMC Genomics*, *25*(1), 361. https://doi.org/10.1186/s12864-024-10285-3

Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and "Omic" Applications. In *Next Generation Sequencing - Advances, Applications and Challenges* (pp. 25–31). InTech. https://doi.org/10.5772/61964

Kwong, J. C., Mccallum, N., Sintchenko, V., & Howden, B. P. (2015). Whole genome sequencing in clinical and public health microbiology. *Pathology*, *47*(3), 199–210. https://doi.org/10.1097/PAT.0000000000000235

Lan, F., Demaree, B., Ahmed, N., & Abate, A. R. (2017). Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nature Biotechnology*, *35*(7), 640–646. https://doi.org/10.1038/nbt.3880

Langmore, J. P. (2002). Rubicon Genomics, Inc. *Pharmacogenomics*, *3*(4), 557–560. https://doi.org/10.1517/14622416.3.4.557

Lasken, R. S. (2012). Genomic sequencing of uncultured microorganisms from single cells. *Nature Reviews Microbiology*, *10*(9), 631–640. https://doi.org/10.1038/nrmicro2857

Leonaviciene, G., Leonavicius, K., Meskys, R., & Mazutis, L. (2020). Multi-step processing of single cells using semi-permeable capsules. *Lab on a Chip*, *20*(21), 4052–4062. https://doi.org/10.1039/D0LC00660B

Lloréns-Rico, V., Simcock, J. A., Huys, G. R. B., & Raes, J. (2022). Single-cell approaches in human microbiome research. *Cell*, *185*(15), 2725–2738. https://doi.org/10.1016/j.cell.2022.06.040

Loreau, M. (2000). Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos*, *91*(1), 3–17. https://doi.org/10.1034/j.1600-0706.2000.910101.x

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, *161*(5), 1202–1214. https://doi.org/10.1016/j.cell.2015.05.002

Marie, D., Le Gall, F., Edern, R., Gourvil, P., & Vaulot, D. (2017). Improvement of phytoplankton culture isolation using single cell sorting by flow cytometry. *Journal of Phycology*, *53*(2), 271–282. https://doi.org/10.1111/jpy.12495

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, *74*(2), 560–564. https://doi.org/10.1073/pnas.74.2.560

McNulty, R., Sritharan, D., Pahng, S. H., Meisch, J. P., Liu, S., Brennan, M. A., Saxer, G., Hormoz, S., & Rosenthal, A. Z. (2023). Probe-based bacterial single-cell RNA sequencing predicts toxin regulation. *Nature Microbiology*, *8*(5), 934–945. https://doi.org/10.1038/s41564-023-01348-4

Moter, A., & Göbel, U. B. (2000). Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *Journal of Microbiological Methods*, *41*(2), 85–112. https://doi.org/10.1016/S0167-7012(00)00152-4

Muyzer, G., de Waal, E. C., & Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, *59*(3), 695–700. https://doi.org/10.1128/aem.59.3.695-700.1993

Picelli, S. (2017). Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biology*, *14*(5), 637–650. https://doi.org/10.1080/15476286.2016.1201618

Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, A. M., Solan, M., van der Gast, C. J., & Young, J. P. W. (2007). The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, *5*(5), 384–392. https://doi.org/10.1038/nrmicro1643

Pusztai, L., Hatzis, C., & Andre, F. (2013). Reproducibility of research and preclinical validation: problems and solutions. *Nature Reviews Clinical Oncology*, *10*(12), 720–724. https://doi.org/10.1038/nrclinonc.2013.171

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*(9), 833–844. https://doi.org/10.1038/nbt.3935

Quinodoz, S. A., Bhat, P., Chovanec, P., Jachowicz, J. W., Ollikainen, N., Detmar, E., Soehalim, E., & Guttman, M. (2022). SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nature Protocols*, *17*(1), 36–75. https://doi.org/10.1038/s41596-021-00633-y

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L. T., Peeler, D. J., Mukherjee, S., Chen, W., Pun, S. H., Sellers, D. L., Tasic, B., & Seelig, G. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, *360*(6385), 176–182. https://doi.org/10.1126/science.aam8999

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: past, present and future. *Nature*, *550*(7676), 345–353. https://doi.org/10.1038/nature24286

Sheth, R. U., Li, M., Jiang, W., Sims, P. A., Leong, K. W., & Wang, H. H. (2019). Spatial metagenomic characterization of microbial biogeography in the gut. *Nature Biotechnology*, *37*(8), 877–883. https://doi.org/10.1038/s41587-019-0183-2

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1). https://doi.org/10.1002/cpmb.59

Slizovskiy, I. B., Mukherjee, K., Dean, C. J., Boucher, C., & Noyes, N. R. (2020). Mobilization of Antibiotic Resistance: Are Current Approaches for Colocalizing Resistomes and Mobilomes Useful? *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01376

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, *321*(6071), 674–679. https://doi.org/10.1038/321674a0

Sommer, M. O. (2015). Advancing gut microbiome research using cultivation. *Current Opinion in Microbiology*, *27*, 127–132. https://doi.org/10.1016/j.mib.2015.08.004

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, *6*(7), 2601–2610. https://doi.org/10.1093/nar/6.7.2601

Streets, A. M., Zhang, X., Cao, C., Pang, Y., Wu, X., Xiong, L., Yang, L., Fu, Y., Zhao, L., Tang, F., & Huang, Y. (2014). Microfluidic single-cell whole-transcriptome sequencing. *Proceedings of the National Academy of Sciences*, *111*(19), 7048–7053. https://doi.org/10.1073/pnas.1402030111

Streit, W. R., & Schmitz, R. A. (2004). Metagenomics – the key to the uncultured microbes. *Current Opinion in Microbiology*, *7*(5), 492–498. https://doi.org/10.1016/j.mib.2004.08.002

Tamminen, M. V., & Virta, M. P. J. (2015). Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.00195

Theberge, A. B., Courtois, F., Schaerli, Y., Fischlechner, M., Abell, C., Hollfelder, F., & Huck, W. T. S. (2010). Microdroplets in Microfluidics: An Evolving Platform for Discoveries in Chemistry and Biology. *Angewandte Chemie International Edition*, *49*(34), 5846–5868. https://doi.org/10.1002/anie.200906653

Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nature Reviews Microbiology*, *3*(9), 711–721. https://doi.org/10.1038/nrmicro1234

Tilman, D., Reich, P. B., & Knops, J. M. H. (2006). Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature*, *441*(7093), 629–632. https://doi.org/10.1038/nature04742

Tokuda, M., & Shintani, M. (2024). Microbial evolution through horizontal gene transfer by mobile genetic elements. *Microbial Biotechnology*, *17*(1). https://doi.org/10.1111/1751-7915.14408

Viailly, P.-J., Sater, V., Viennot, M., Bohers, E., Vergne, N., Berard, C., Dauchel, H., Lecroq, T., Celebi, A., Ruminy, P., Marchand, V., Lanic, M.-D., Dubois, S., Penther, D., Tilly, H., Mareschal, S., & Jardin, F. (2021). Improving high-resolution copy number variation analysis from next generation sequencing using unique molecular identifiers. *BMC Bioinformatics*, *22*(1), 120. https://doi.org/10.1186/s12859-021-04060-4

Wagner, M., Loy, A., Nogueira, R., Purkhold, U., Lee, N., & Daims, H. (2002). Microbial community composition and function in wastewater treatment plants. *Antonie van Leeuwenhoek*, *81*(1/4), 665–680. https://doi.org/10.1023/A:1020586312170

Wang, B., Lin, A. E., Yuan, J., Novak, K. E., Koch, M. D., Wingreen, N. S., Adamson, B., & Gitai, Z. (2023). Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. *Nature Microbiology*, *8*(10), 1846–1862. https://doi.org/10.1038/s41564-023-01462-3

Wang, Y., Cao, T., Ko, J., Shen, Y., Zong, W., Sheng, K., Cao, W., Sun, S., Cai, L., Zhou, Y., Zhang, X., Zong, C., Weissleder, R., & Weitz, D. (2020). Dissolvable Polyacrylamide Beads for High-Throughput Droplet DNA Barcoding. *Advanced Science*, *7*(8). https://doi.org/10.1002/advs.201903463

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Weaver, W. M., Tseng, P., Kunze, A., Masaeli, M., Chung, A. J., Dudani, J. S., Kittur, H., Kulkarni, R. P., & Di Carlo, D. (2014). Advances in high-throughput single-cell microtechnologies. *Current Opinion in Biotechnology*, *25*, 114–123. https://doi.org/10.1016/j.copbio.2013.09.005

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wilmes, P., Simmons, S. L., Denef, V. J., & Banfield, J. F. (2009). The dynamic genetic repertoire of microbial communities. *FEMS Microbiology Reviews*, *33*(1), 109–132. https://doi.org/10.1111/j.1574-6976.2008.00144.x

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, *51*(2), 221–271. https://doi.org/10.1128/MMBR.51.2.221-271.1987

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, *87*(12), 4576–4579. https://doi.org/10.1073/pnas.87.12.4576

Wu, A. R., Neff, N. F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M. E., Mburu, F. M., Mantalas, G. L., Sim, S., Clarke, M. F., & Quake, S. R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nature Methods*, *11*(1), 41–46. https://doi.org/10.1038/nmeth.2694

Xu, Z., Wang, Y., Sheng, K., Rosenthal, R., Liu, N., Hua, X., Zhang, T., Chen, J., Song, M., Lv, Y., Zhang, S., Huang, Y., Wang, Z., Cao, T., Shen, Y., Jiang, Y., Yu, Y., Chen, Y., Guo, G., … Wang, Y. (2023). Droplet-based high-throughput single microbe RNA sequencing by smRandom-seq. *Nature Communications*, *14*(1), 5130. https://doi.org/10.1038/s41467-023-40137-9

Yanni, D., Márquez-Zacarías, P., Yunker, P. J., & Ratcliff, W. C. (2019). Drivers of Spatial Structure in Social Microbial Communities. *Current Biology*, *29*(11), R545–R550. https://doi.org/10.1016/j.cub.2019.03.068

Young, I. M., Crawford, J. W., Nunan, N., Otten, W., & Spiers, A. (2008). *Chapter 4 Microbial Distribution in Soils* (pp. 81–121). https://doi.org/10.1016/S0065-2113(08)00604-4

Zhang, R., Thiyagarajan, V., & Qian, P.-Y. (2008). Evaluation of terminal-restriction fragment length polymorphism analysis in contrasting marine environments. *FEMS Microbiology Ecology*, *65*(1), 169–178. https://doi.org/10.1111/j.1574-6941.2008.00493.x

Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D. M., Evans, E. D., Dai, C. L., Lu, P. J., Alm, E. J., & Weitz, D. A. (2022). High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science*, *376*(6597). https://doi.org/10.1126/science.abm1483

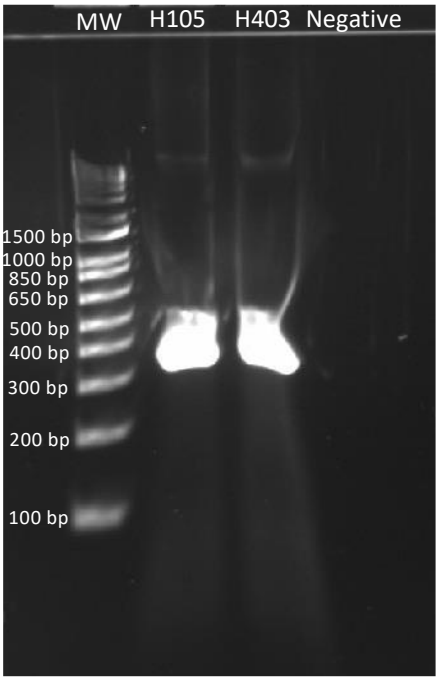
Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A. M., & Mazutis, L. (2017). Single-cell barcoding and sequencing using droplet microfluidics. *Nature Protocols*, *12*(1), 44–73. https://doi.org/10.1038/nprot.2016.154

Zong, C., Lu, S., Chapman, A. R., & Xie, X. S. (2012). Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. *Science*, *338*(6114), 1622–1626. https://doi.org/10.1126/science.1229164

Appendices

Appendix 1. 16S rRNA Gene-Overhang Amplification for Synthetic Beads

The PCR amplification of the 16S rRNA V4 gene region with an overhang for Linker 1, using HAMBI105 and HAMBI403 DNA for synthetic beads, was verified by gel electrophoresis, confirming the correct amplicon size of approximately 285 bp (Appendix Figure 1).



Appendix Figure 1. Gel electrophoresis image showing the PCR-amplified 16S rRNA V4 gene region from HAMBI105 and HAMBI403 DNA, used for creating synthetic polyacrylamide beads. MW = molecular weight. 1Kb Plus DNA ladder. 2% agarose E-Gel. Gel run 48V, 10 min.

The concentrations of the extracted and purified PCR products, measured with NanoDrop 1000, are shown in Appendix Table 1.

Appendix Table 1. The concentrations of gel extracted and purified PCR amplicons, measured with NanoDrop 1000.

	H105	**H403**
Concentration	23.6 ng/µL	21.6 ng/µL
260/280	2.00	2.00
260/230	1.86	1.84

Appendix 2. Protocol for DNA Cleanup and Concentration Using the Monarch® PCR & DNA Cleanup Kit (5 µg)

Protocol for DNA Cleanup and Concentration Using the Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB #T1030)

Important Update: Beginning in May 2021, we will be gradually transitioning the Monarch DNA Cleanup Binding Buffer to a concentrated format which requires the addition of isopropanol by the user. The protocol below has been updated to reflect this change, but please refer to the instructions provided with your products, as your lot may not be affected.

There are two protocols available for this product:

- **DNA Cleanup and Concentration (below):** for the purification of up to 5 µg of DNA (ssDNA > 200 nt and dsDNA > 50 bp) from PCR and other enzymatic reactions.
- **Oligonucleotide Cleanup Protocol:** for the purification of up to 5 µg of DNA fragments ≥ 15 bp (dsDNA) or ≥ 18 nt (ssDNA). Expected recovery is > 70%. When purifying ssDNA of any size, recovery can be increased by using this protocol; however, it is important to note that this protocol shifts the cutoff for smaller fragments to 18 nt (rather than 50 nt for the DNA Cleanup and Concentration Protocol).

Download Quick Protocol Card

General Guidelines:

Input amount of DNA to be purified should not exceed the binding capacity of the column (5 µg). A starting sample volume of 20–100 µl is recommended. For smaller samples, TE can be used to adjust the volume to the recommended volume range. Centrifugation should be carried out at 16,000 x g in a standard laboratory microcentrifuge at room temperature.

Before You Begin:

Add isopropanol to Monarch DNA Cleanup Binding Buffer prior to use*:

- For the 50-prep kit, add 14 ml of isopropanol to the DNA Cleanup Binding Buffer.
- For the 250-prep kit, add 63.5 ml of isopropanol to the DNA Cleanup Binding Buffer.

Add ethanol to Monarch DNA Wash Buffer prior to use (4 volumes of ≥ 95% ethanol per volume of Monarch DNA Wash Buffer)

- For the 50-prep kit, add 20 ml of ethanol to the Monarch DNA Wash Buffer
- For the 250-prep kit, add 100 ml of ethanol to the Monarch DNA Wash Buffer

Always keep all buffer bottles tightly closed when not in use.

Protocol:

All centrifugation steps should be carried out at 16,000 x *g* (~13K RPM in a typical microcentrifuge).

1. Dilute sample with DNA Cleanup Binding Buffer (ensure that isopropanol has been added, as indicated on the bottle label)* according to the table below. Mix well by pipetting up and down or flicking the tube. Do not vortex. A starting sample volume of 20–100 µl is recommended. For smaller samples, TE can be used to adjust the volume. For diluted samples larger than 800 µl, load a portion of the sample, proceed with Step 2, and then repeat as necessary.

   **Beginning in April 2021, the DNA Cleanup Binding Buffer will be changed to a concentrated format which requires the addition of isopropanol by the user. Please refer to the instructions inside of the product that you receive.*

SAMPLE TYPE	RATIO OF BINDING BUFFER: SAMPLE	EXAMPLE
dsDNA > 2 kb (plasmids, gDNA)	2:1	200 µl:100 µl
dsDNA < 2 kb (some amplicons, fragments)	5:1	500 µl:100 µl

ssDNA > 200 nt**	7:1	700 µl:100 µl

** *Please note that recovery of ssDNA < 200 nts can be increased by using the Oligonucleotide Cleanup Protocol, but doing so will shift the cutoff size for DNA binding to 18 nt (versus 50 nt).*

2. Insert column into collection tube and load sample onto column and close the cap. Spin for 1 minute, then discard flow-through.

   To save time, spin for 30 seconds, instead of 1 minute.

   If using a vacuum manifold instead of centrifugation, insert the column into the manifold and switch the vacuum on. Allow the solution to pass through the column, then switch the vacuum source off. Make sure to follow the manifold manufacturer's instructions to set-up the manifold and connect it properly to a vacuum source.

3. Re-insert column into collection tube. Add 200 µl DNA Wash Buffer and spin for 1 minute. Discarding flow-through is optional.

   If using a vacuum manifold, add 200 µl of DNA Wash Buffer and switch the vacuum on. Allow the solution to pass through the column, then switch the vacuum source off.

4. Repeat wash (Step 3).

5. Transfer column to a clean 1.5 ml microfuge tube. Use care to ensure that the tip of the column does not come into contact with the flow-through. If in doubt, re-spin for 1 minute to ensure traces of salt and ethanol are not carried over to next step.

   If using a vacuum manifold: Since vacuum set-ups can vary, a 1 minute centrifugation is recommended prior to elution to ensure that no traces of salt or ethanol are carried over to the next step.

6. Add ≥ 6 µl of DNA Elution Buffer to the center of the matrix. Wait for 1 minute, then spin for 1 minute to elute DNA.

   Note: Typical elution volumes are 6–20 µl. Nuclease-free water (pH 7–8.5) can also be used to elute the DNA. Yield may slightly increase if a larger volume of DNA Elution Buffer is used, but the DNA will be less concentrated. For larger size DNA (≥ 10 kb), heating the elution buffer to 50°C prior to use can improve yield. Care should be used to ensure the elution buffer is delivered onto the matrix and not the wall of the column to maximize elution efficiency.

   To save time, spin for 30 seconds, instead of 1 minute.

Links to this resource

Product Categories: Nucleic Acid Purification Products

Applications: PCR & Reaction Cleanup, Nucleic Acid Purification

Related Products: Monarch® PCR & DNA Cleanup Kit (5 µg), Monarch® DNA Cleanup Columns (5 µg), Monarch® DNA Elution Buffer, | More +

Videos

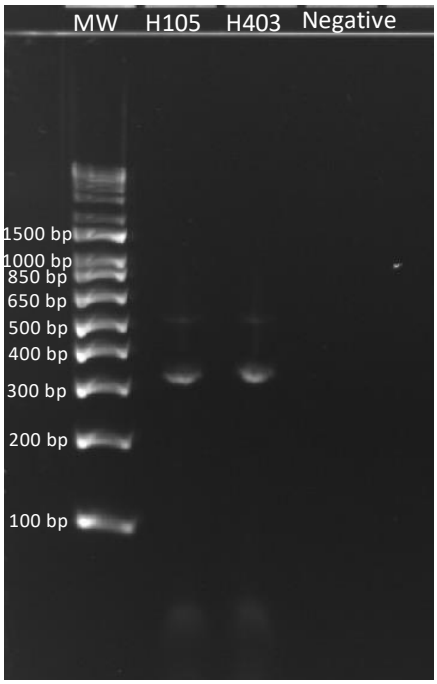
1 of 2

Tips for using the Monarch® PCR & DNA Cleanup Kit

Optimize your DNA isolation from PCR and other enzymatic reactions with our quick tips for using the Monarch PCR & DNA Cleanup Kit.

Appendix 3. Incorporation of an Acrydited Uracil Primer

The incorporation of the acrydited uracil primer to previously amplified HAMBI105 and HAMBI403 DNA was verified using gel electrophoresis. The amplicons were confirmed to be the correct size, approximately 294 bp (Appendix Figure 2).



Appendix Figure 2. Gel electrophoresis image showing HAMBI105 and HAMBI403 DNA amplicons with the incorporated acrydited uracil primer, used for creating synthetic polyacrylamide beads. MW = molecular weight. 1Kb Plus DNA ladder. 2% agarose E-Gel. Gel run 48V, 10 min.

The concentrations of the extracted and purified PCR products, measured with NanoDrop 1000, are shown in Appendix Table 2.

Appendix Table 2. The concentrations of gel extracted and purified PCR amplicons, measured with NanoDrop 1000.

	H105	**H403**
Concentration	16.6 ng/µL	19.0 ng/µL
260/280	1.71	1.76
260/230	-3.69	0.53

Appendix 4. Barcode plates' oligonucleotide sequences

Appendix Table 3. Individual oligonucleotide sequences for each well in Barcode plate 1.

Barcode plate 1		
Well Position	**Name**	**Sequence**
A1	PCR_Barcode1_01	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAACGTGATGTGGCCGATGTTTCG
B1	PCR_Barcode1_02	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAAACATCGGTGGCCGATGTTTCG
C1	PCR_Barcode1_03	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNATGCCTAAGTGGCCGATGTTTCG
D1	PCR_Barcode1_04	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGTGGTCAGTGGCCGATGTTTCG
E1	PCR_Barcode1_05	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACCACTGTGTGGCCGATGTTTCG
F1	PCR_Barcode1_06	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACATTGGCGTGGCCGATGTTTCG
H1	PCR_Barcode1_07	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCAGATCTGGTGGCCGATGTTTCG
G1	PCR_Barcode1_08	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCATCAAGTGTGGCCGATGTTTCG
A2	PCR_Barcode1_09	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCGCTGATCGTGGCCGATGTTTCG

B2	PCR_Barcode1_10	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACAAGCTAGTGGCCGATGTTTCG
C2	PCR_Barcode1_11	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCTGTAGCCGTGGCCGATGTTTCG
D2	PCR_Barcode1_12	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGTACAAGGTGGCCGATGTTTCG
E2	PCR_Barcode1_13	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAACAACCAGTGGCCGATGTTTCG
F2	PCR_Barcode1_14	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAACCGAGAGTGGCCGATGTTTCG
H2	PCR_Barcode1_15	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAACGCTTAGTGGCCGATGTTTCG
G2	PCR_Barcode1_16	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAAGACGGAGTGGCCGATGTTTCG
A3	PCR_Barcode1_17	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAAGGTACAGTGGCCGATGTTTCG
B3	PCR_Barcode1_18	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACACAGAAGTGGCCGATGTTTCG
C3	PCR_Barcode1_19	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACAGCAGAGTGGCCGATGTTTCG
D3	PCR_Barcode1_20	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACCTCCAAGTGGCCGATGTTTCG
E3	PCR_Barcode1_21	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACGCTCGAGTGGCCGATGTTTCG
F3	PCR_Barcode1_22	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACGTATCAGTGGCCGATGTTTCG
G3	PCR_Barcode1_23	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACTATGCAGTGGCCGATGTTTCG
H3	PCR_Barcode1_24	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGAGTCAAGTGGCCGATGTTTCG
A4	PCR_Barcode1_25	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGATCGCAGTGGCCGATGTTTCG
B4	PCR_Barcode1_26	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGCAGGAAGTGGCCGATGTTTCG
C4	PCR_Barcode1_27	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGTCACTAGTGGCCGATGTTTCG
D4	PCR_Barcode1_28	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNATCCTGTAGTGGCCGATGTTTCG
E4	PCR_Barcode1_29	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNATTGAGGAGTGGCCGATGTTTCG
F4	PCR_Barcode1_30	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCAACCACAGTGGCCGATGTTTCG
G4	PCR_Barcode1_31	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGACTAGTAGTGGCCGATGTTTCG
H4	PCR_Barcode1_32	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCAATGGAAGTGGCCGATGTTTCG
A5	PCR_Barcode1_33	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCACTTCGAGTGGCCGATGTTTCG
B5	PCR_Barcode1_34	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCAGCGTTAGTGGCCGATGTTTCG
C5	PCR_Barcode1_35	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCATACCAAGTGGCCGATGTTTCG
D5	PCR_Barcode1_36	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCAGTTCAGTGGCCGATGTTTCG
E5	PCR_Barcode1_37	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCGAAGTAGTGGCCGATGTTTCG
F5	PCR_Barcode1_38	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCGTGAGAGTGGCCGATGTTTCG
G5	PCR_Barcode1_39	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCTCCTGAGTGGCCGATGTTTCG
H5	PCR_Barcode1_40	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCGAACTTAGTGGCCGATGTTTCG
A6	PCR_Barcode1_41	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCGACTGGAGTGGCCGATGTTTCG
B6	PCR_Barcode1_42	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCGCATACAGTGGCCGATGTTTCG
C6	PCR_Barcode1_43	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCTCAATGAGTGGCCGATGTTTCG
D6	PCR_Barcode1_44	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCTGAGCCAGTGGCCGATGTTTCG
E6	PCR_Barcode1_45	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCTGGCATAGTGGCCGATGTTTCG
F6	PCR_Barcode1_46	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGAATCTGAGTGGCCGATGTTTCG
G6	PCR_Barcode1_47	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCAAGACTAGTGGCCGATGTTTCG
H6	PCR_Barcode1_48	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGAGCTGAAGTGGCCGATGTTTCG
A7	PCR_Barcode1_49	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGATAGACAGTGGCCGATGTTTCG
B7	PCR_Barcode1_50	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGCCACATAGTGGCCGATGTTTCG
C7	PCR_Barcode1_51	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGCGAGTAAGTGGCCGATGTTTCG
D7	PCR_Barcode1_52	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGCTAACGAGTGGCCGATGTTTCG
E7	PCR_Barcode1_53	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGCTCGGTAGTGGCCGATGTTTCG
F7	PCR_Barcode1_54	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGGAGAACAGTGGCCGATGTTTCG
G7	PCR_Barcode1_55	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGGTGCGAAGTGGCCGATGTTTCG
H7	PCR_Barcode1_56	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGTACGCAAGTGGCCGATGTTTCG
A8	PCR_Barcode1_57	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGTCGTAGAGTGGCCGATGTTTCG
B8	PCR_Barcode1_58	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGTCTGTCAGTGGCCGATGTTTCG
C8	PCR_Barcode1_59	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGTGTTCTAGTGGCCGATGTTTCG
D8	PCR_Barcode1_60	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTAGGATGAGTGGCCGATGTTTCG
E8	PCR_Barcode1_61	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTATCAGCAGTGGCCGATGTTTCG
F8	PCR_Barcode1_62	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTCCGTCTAGTGGCCGATGTTTCG

G8	PCR_Barcode1_63	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTCTTCACAGTGGCCGATGTTTCG
H8	PCR_Barcode1_64	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTGAAGAGAGTGGCCGATGTTTCG
A9	PCR_Barcode1_65	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTGGAACAAGTGGCCGATGTTTCG
B9	PCR_Barcode1_66	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTGGCTTCAGTGGCCGATGTTTCG
C9	PCR_Barcode1_67	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTGGTGGTAGTGGCCGATGTTTCG
D9	PCR_Barcode1_68	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNTTCACGCAGTGGCCGATGTTTCG
E9	PCR_Barcode1_69	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAACTCACCGTGGCCGATGTTTCG
F9	PCR_Barcode1_70	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAAGAGATCGTGGCCGATGTTTCG
G9	PCR_Barcode1_71	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAAGGACACGTGGCCGATGTTTCG
H9	PCR_Barcode1_72	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAATCCGTCGTGGCCGATGTTTCG
A10	PCR_Barcode1_73	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAATGTTGCGTGGCCGATGTTTCG
B10	PCR_Barcode1_74	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACACGACCGTGGCCGATGTTTCG
C10	PCR_Barcode1_75	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNACAGATTCGTGGCCGATGTTTCG
D10	PCR_Barcode1_76	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGATGTACGTGGCCGATGTTTCG
E10	PCR_Barcode1_77	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGCACCTCGTGGCCGATGTTTCG
F10	PCR_Barcode1_78	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGCCATGCGTGGCCGATGTTTCG
G10	PCR_Barcode1_79	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNAGGCTAACGTGGCCGATGTTTCG
H10	PCR_Barcode1_80	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNATAGCGACGTGGCCGATGTTTCG
A11	PCR_Barcode1_81	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNATCATTCCGTGGCCGATGTTTCG
B11	PCR_Barcode1_82	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNATTGGCTCGTGGCCGATGTTTCG
C11	PCR_Barcode1_83	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCAAGGAGCGTGGCCGATGTTTCG
D11	PCR_Barcode1_84	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCACCTTACGTGGCCGATGTTTCG
E11	PCR_Barcode1_85	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCATCCTCGTGGCCGATGTTTCG
F11	PCR_Barcode1_86	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCGACAACGTGGCCGATGTTTCG
G11	PCR_Barcode1_87	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCTAATCCGTGGCCGATGTTTCG
H11	PCR_Barcode1_88	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCCTCTATCGTGGCCGATGTTTCG
A12	PCR_Barcode1_89	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCGACACACGTGGCCGATGTTTCG
B12	PCR_Barcode1_90	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCGGATTGCGTGGCCGATGTTTCG
C12	PCR_Barcode1_91	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNCTAAGGTCGTGGCCGATGTTTCG
D12	PCR_Barcode1_92	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGAACAGGCGTGGCCGATGTTTCG
E12	PCR_Barcode1_93	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGACAGTGCGTGGCCGATGTTTCG
F12	PCR_Barcode1_94	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGAGTTAGCGTGGCCGATGTTTCG
G12	PCR_Barcode1_95	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGATGAATCGTGGCCGATGTTTCG
H12	PCR_Barcode1_96	/5Phos/GTTCTGAGCAAGGGTNNNNNNNNNNGCCAAGACGTGGCCGATGTTTCG

Appendix Table 4. Individual oligonucleotide sequences for each well in Barcode plate 2.

Barcode plate 2		
Well Position	**Name**	**Sequence**
A1	Barcode2_01	/5Phos/CATCGGCGTACGACTAACGTGATATCCACGTGCTTGAG
B1	Barcode2_02	/5Phos/CATCGGCGTACGACTAAACATCGATCCACGTGCTTGAG
C1	Barcode2_03	/5Phos/CATCGGCGTACGACTATGCCTAAATCCACGTGCTTGAG
D1	Barcode2_04	/5Phos/CATCGGCGTACGACTAGTGGTCAATCCACGTGCTTGAG
E1	Barcode2_05	/5Phos/CATCGGCGTACGACTACCACTGTATCCACGTGCTTGAG
F1	Barcode2_06	/5Phos/CATCGGCGTACGACTACATTGGCATCCACGTGCTTGAG
H1	Barcode2_07	/5Phos/CATCGGCGTACGACTCAGATCTGATCCACGTGCTTGAG
G1	Barcode2_08	/5Phos/CATCGGCGTACGACTCATCAAGTATCCACGTGCTTGAG
A2	Barcode2_09	/5Phos/CATCGGCGTACGACTCGCTGATCATCCACGTGCTTGAG
B2	Barcode2_10	/5Phos/CATCGGCGTACGACTACAAGCTAATCCACGTGCTTGAG
C2	Barcode2_11	/5Phos/CATCGGCGTACGACTCTGTAGCCATCCACGTGCTTGAG
D2	Barcode2_12	/5Phos/CATCGGCGTACGACTAGTACAAGATCCACGTGCTTGAG
E2	Barcode2_13	/5Phos/CATCGGCGTACGACTAACAACCAATCCACGTGCTTGAG

F2	Barcode2_14	/5Phos/CATCGGCGTACGACTAACCGAGAATCCACGTGCTTGAG
H2	Barcode2_15	/5Phos/CATCGGCGTACGACTAACGCTTAATCCACGTGCTTGAG
G2	Barcode2_16	/5Phos/CATCGGCGTACGACTAAGACGGAATCCACGTGCTTGAG
A3	Barcode2_17	/5Phos/CATCGGCGTACGACTAAGGTACAATCCACGTGCTTGAG
B3	Barcode2_18	/5Phos/CATCGGCGTACGACTACACAGAAATCCACGTGCTTGAG
C3	Barcode2_19	/5Phos/CATCGGCGTACGACTACAGCAGAATCCACGTGCTTGAG
D3	Barcode2_20	/5Phos/CATCGGCGTACGACTACCTCCAAATCCACGTGCTTGAG
E3	Barcode2_21	/5Phos/CATCGGCGTACGACTACGCTCGAATCCACGTGCTTGAG
F3	Barcode2_22	/5Phos/CATCGGCGTACGACTACGTATCAATCCACGTGCTTGAG
G3	Barcode2_23	/5Phos/CATCGGCGTACGACTACTATGCAATCCACGTGCTTGAG
H3	Barcode2_24	/5Phos/CATCGGCGTACGACTAGAGTCAAATCCACGTGCTTGAG
A4	Barcode2_25	/5Phos/CATCGGCGTACGACTAGATCGCAATCCACGTGCTTGAG
B4	Barcode2_26	/5Phos/CATCGGCGTACGACTAGCAGGAAATCCACGTGCTTGAG
C4	Barcode2_27	/5Phos/CATCGGCGTACGACTAGTCACTAATCCACGTGCTTGAG
D4	Barcode2_28	/5Phos/CATCGGCGTACGACTATCCTGTAATCCACGTGCTTGAG
E4	Barcode2_29	/5Phos/CATCGGCGTACGACTATTGAGGAATCCACGTGCTTGAG
F4	Barcode2_30	/5Phos/CATCGGCGTACGACTCAACCACAATCCACGTGCTTGAG
G4	Barcode2_31	/5Phos/CATCGGCGTACGACTGACTAGTAATCCACGTGCTTGAG
H4	Barcode2_32	/5Phos/CATCGGCGTACGACTCAATGGAAATCCACGTGCTTGAG
A5	Barcode2_33	/5Phos/CATCGGCGTACGACTCACTTCGAATCCACGTGCTTGAG
B5	Barcode2_34	/5Phos/CATCGGCGTACGACTCAGCGTTAATCCACGTGCTTGAG
C5	Barcode2_35	/5Phos/CATCGGCGTACGACTCATACCAAATCCACGTGCTTGAG
D5	Barcode2_36	/5Phos/CATCGGCGTACGACTCCAGTTCAATCCACGTGCTTGAG
E5	Barcode2_37	/5Phos/CATCGGCGTACGACTCCGAAGTAATCCACGTGCTTGAG
F5	Barcode2_38	/5Phos/CATCGGCGTACGACTCCGTGAGAATCCACGTGCTTGAG
G5	Barcode2_39	/5Phos/CATCGGCGTACGACTCCTCCTGAATCCACGTGCTTGAG
H5	Barcode2_40	/5Phos/CATCGGCGTACGACTCGAACTTAATCCACGTGCTTGAG
A6	Barcode2_41	/5Phos/CATCGGCGTACGACTCGACTGGAATCCACGTGCTTGAG
B6	Barcode2_42	/5Phos/CATCGGCGTACGACTCGCATACAATCCACGTGCTTGAG
C6	Barcode2_43	/5Phos/CATCGGCGTACGACTCTCAATGAATCCACGTGCTTGAG
D6	Barcode2_44	/5Phos/CATCGGCGTACGACTCTGAGCCAATCCACGTGCTTGAG
E6	Barcode2_45	/5Phos/CATCGGCGTACGACTCTGGCATAATCCACGTGCTTGAG
F6	Barcode2_46	/5Phos/CATCGGCGTACGACTGAATCTGAATCCACGTGCTTGAG
G6	Barcode2_47	/5Phos/CATCGGCGTACGACTCAAGACTAATCCACGTGCTTGAG
H6	Barcode2_48	/5Phos/CATCGGCGTACGACTGAGCTGAAATCCACGTGCTTGAG
A7	Barcode2_49	/5Phos/CATCGGCGTACGACTGATAGACAATCCACGTGCTTGAG
B7	Barcode2_50	/5Phos/CATCGGCGTACGACTGCCACATAATCCACGTGCTTGAG
C7	Barcode2_51	/5Phos/CATCGGCGTACGACTGCGAGTAAATCCACGTGCTTGAG
D7	Barcode2_52	/5Phos/CATCGGCGTACGACTGCTAACGAATCCACGTGCTTGAG
E7	Barcode2_53	/5Phos/CATCGGCGTACGACTGCTCGGTAATCCACGTGCTTGAG
F7	Barcode2_54	/5Phos/CATCGGCGTACGACTGGAGAACAATCCACGTGCTTGAG
G7	Barcode2_55	/5Phos/CATCGGCGTACGACTGGTGCGAAATCCACGTGCTTGAG
H7	Barcode2_56	/5Phos/CATCGGCGTACGACTGTACGCAAATCCACGTGCTTGAG
A8	Barcode2_57	/5Phos/CATCGGCGTACGACTGTCGTAGAATCCACGTGCTTGAG
B8	Barcode2_58	/5Phos/CATCGGCGTACGACTGTCTGTCAATCCACGTGCTTGAG
C8	Barcode2_59	/5Phos/CATCGGCGTACGACTGTGTTCTAATCCACGTGCTTGAG
D8	Barcode2_60	/5Phos/CATCGGCGTACGACTTAGGATGAATCCACGTGCTTGAG
E8	Barcode2_61	/5Phos/CATCGGCGTACGACTTATCAGCAATCCACGTGCTTGAG
F8	Barcode2_62	/5Phos/CATCGGCGTACGACTTCCGTCTAATCCACGTGCTTGAG
G8	Barcode2_63	/5Phos/CATCGGCGTACGACTTCTTCACAATCCACGTGCTTGAG
H8	Barcode2_64	/5Phos/CATCGGCGTACGACTTGAAGAGAATCCACGTGCTTGAG
A9	Barcode2_65	/5Phos/CATCGGCGTACGACTTGGAACAAATCCACGTGCTTGAG
B9	Barcode2_66	/5Phos/CATCGGCGTACGACTTGGCTTCAATCCACGTGCTTGAG

C9	Barcode2_67	/5Phos/CATCGGCGTACGACTTGGTGGTAATCCACGTGCTTGAG
D9	Barcode2_68	/5Phos/CATCGGCGTACGACTTTCACGCAATCCACGTGCTTGAG
E9	Barcode2_69	/5Phos/CATCGGCGTACGACTAACTCACCATCCACGTGCTTGAG
F9	Barcode2_70	/5Phos/CATCGGCGTACGACTAAGAGATCATCCACGTGCTTGAG
G9	Barcode2_71	/5Phos/CATCGGCGTACGACTAAGGACACATCCACGTGCTTGAG
H9	Barcode2_72	/5Phos/CATCGGCGTACGACTAATCCGTCATCCACGTGCTTGAG
A10	Barcode2_73	/5Phos/CATCGGCGTACGACTAATGTTGCATCCACGTGCTTGAG
B10	Barcode2_74	/5Phos/CATCGGCGTACGACTACACGACCATCCACGTGCTTGAG
C10	Barcode2_75	/5Phos/CATCGGCGTACGACTACAGATTCATCCACGTGCTTGAG
D10	Barcode2_76	/5Phos/CATCGGCGTACGACTAGATGTACATCCACGTGCTTGAG
E10	Barcode2_77	/5Phos/CATCGGCGTACGACTAGCACCTCATCCACGTGCTTGAG
F10	Barcode2_78	/5Phos/CATCGGCGTACGACTAGCCATGCATCCACGTGCTTGAG
G10	Barcode2_79	/5Phos/CATCGGCGTACGACTAGGCTAACATCCACGTGCTTGAG
H10	Barcode2_80	/5Phos/CATCGGCGTACGACTATAGCGACATCCACGTGCTTGAG
A11	Barcode2_81	/5Phos/CATCGGCGTACGACTATCATTCCATCCACGTGCTTGAG
B11	Barcode2_82	/5Phos/CATCGGCGTACGACTATTGGCTCATCCACGTGCTTGAG
C11	Barcode2_83	/5Phos/CATCGGCGTACGACTCAAGGAGCATCCACGTGCTTGAG
D11	Barcode2_84	/5Phos/CATCGGCGTACGACTCACCTTACATCCACGTGCTTGAG
E11	Barcode2_85	/5Phos/CATCGGCGTACGACTCCATCCTCATCCACGTGCTTGAG
F11	Barcode2_86	/5Phos/CATCGGCGTACGACTCCGACAACATCCACGTGCTTGAG
G11	Barcode2_87	/5Phos/CATCGGCGTACGACTCCTAATCCATCCACGTGCTTGAG
H11	Barcode2_88	/5Phos/CATCGGCGTACGACTCCTCTATCATCCACGTGCTTGAG
A12	Barcode2_89	/5Phos/CATCGGCGTACGACTCGACACACATCCACGTGCTTGAG
B12	Barcode2_90	/5Phos/CATCGGCGTACGACTCGGATTGCATCCACGTGCTTGAG
C12	Barcode2_91	/5Phos/CATCGGCGTACGACTCTAAGGTCATCCACGTGCTTGAG
D12	Barcode2_92	/5Phos/CATCGGCGTACGACTGAACAGGCATCCACGTGCTTGAG
E12	Barcode2_93	/5Phos/CATCGGCGTACGACTGACAGTGCATCCACGTGCTTGAG
F12	Barcode2_94	/5Phos/CATCGGCGTACGACTGAGTTAGCATCCACGTGCTTGAG
G12	Barcode2_95	/5Phos/CATCGGCGTACGACTGATGAATCATCCACGTGCTTGAG
H12	Barcode2_96	/5Phos/CATCGGCGTACGACTGCCAAGACATCCACGTGCTTGAG

Appendix Table 5. Individual oligonucleotide sequences for each well in Barcode plate 3.

Barcode plate 3		
Well Position	**Name**	**Sequence**
A1	Barcode3_01	/5Phos/GGCCAGAGCATTCGAACGTGATGATCATGACCCATTTGGAGAAGATG
B1	Barcode3_02	/5Phos/GGCCAGAGCATTCGAAACATCGGATCATGACCCATTTGGAGAAGATG
C1	Barcode3_03	/5Phos/GGCCAGAGCATTCGATGCCTAAGATCATGACCCATTTGGAGAAGATG
D1	Barcode3_04	/5Phos/GGCCAGAGCATTCGAGTGGTCAGATCATGACCCATTTGGAGAAGATG
E1	Barcode3_05	/5Phos/GGCCAGAGCATTCGACCACTGTGATCATGACCCATTTGGAGAAGATG
F1	Barcode3_06	/5Phos/GGCCAGAGCATTCGACATTGGCGATCATGACCCATTTGGAGAAGATG
H1	Barcode3_07	/5Phos/GGCCAGAGCATTCGCAGATCTGGATCATGACCCATTTGGAGAAGATG
G1	Barcode3_08	/5Phos/GGCCAGAGCATTCGCATCAAGTGATCATGACCCATTTGGAGAAGATG
A2	Barcode3_09	/5Phos/GGCCAGAGCATTCGCGCTGATCGATCATGACCCATTTGGAGAAGATG
B2	Barcode3_10	/5Phos/GGCCAGAGCATTCGACAAGCTAGATCATGACCCATTTGGAGAAGATG
C2	Barcode3_11	/5Phos/GGCCAGAGCATTCGCTGTAGCCGATCATGACCCATTTGGAGAAGATG
D2	Barcode3_12	/5Phos/GGCCAGAGCATTCGAGTACAAGGATCATGACCCATTTGGAGAAGATG
E2	Barcode3_13	/5Phos/GGCCAGAGCATTCGAACAACCAGATCATGACCCATTTGGAGAAGATG
F2	Barcode3_14	/5Phos/GGCCAGAGCATTCGAACCGAGAGATCATGACCCATTTGGAGAAGATG
H2	Barcode3_15	/5Phos/GGCCAGAGCATTCGAACGCTTAGATCATGACCCATTTGGAGAAGATG
G2	Barcode3_16	/5Phos/GGCCAGAGCATTCGAAGACGGAGATCATGACCCATTTGGAGAAGATG
A3	Barcode3_17	/5Phos/GGCCAGAGCATTCGAAGGTACAGATCATGACCCATTTGGAGAAGATG

B3	Barcode3_18	/5Phos/GGCCAGAGCATTCGACACAGAAGATCATGACCCATTTGGAGAAGATG
C3	Barcode3_19	/5Phos/GGCCAGAGCATTCGACAGCAGAGATCATGACCCATTTGGAGAAGATG
D3	Barcode3_20	/5Phos/GGCCAGAGCATTCGACCTCCAAGATCATGACCCATTTGGAGAAGATG
E3	Barcode3_21	/5Phos/GGCCAGAGCATTCGACGCTCGAGATCATGACCCATTTGGAGAAGATG
F3	Barcode3_22	/5Phos/GGCCAGAGCATTCGACGTATCAGATCATGACCCATTTGGAGAAGATG
G3	Barcode3_23	/5Phos/GGCCAGAGCATTCGACTATGCAGATCATGACCCATTTGGAGAAGATG
H3	Barcode3_24	/5Phos/GGCCAGAGCATTCGAGAGTCAAGATCATGACCCATTTGGAGAAGATG
A4	Barcode3_25	/5Phos/GGCCAGAGCATTCGAGATCGCAGATCATGACCCATTTGGAGAAGATG
B4	Barcode3_26	/5Phos/GGCCAGAGCATTCGAGCAGGAAGATCATGACCCATTTGGAGAAGATG
C4	Barcode3_27	/5Phos/GGCCAGAGCATTCGAGTCACTAGATCATGACCCATTTGGAGAAGATG
D4	Barcode3_28	/5Phos/GGCCAGAGCATTCGATCCTGTAGATCATGACCCATTTGGAGAAGATG
E4	Barcode3_29	/5Phos/GGCCAGAGCATTCGATTGAGGAGATCATGACCCATTTGGAGAAGATG
F4	Barcode3_30	/5Phos/GGCCAGAGCATTCGCAACCACAGATCATGACCCATTTGGAGAAGATG
G4	Barcode3_31	/5Phos/GGCCAGAGCATTCGGACTAGTAGATCATGACCCATTTGGAGAAGATG
H4	Barcode3_32	/5Phos/GGCCAGAGCATTCGCAATGGAAGATCATGACCCATTTGGAGAAGATG
A5	Barcode3_33	/5Phos/GGCCAGAGCATTCGCACTTCGAGATCATGACCCATTTGGAGAAGATG
B5	Barcode3_34	/5Phos/GGCCAGAGCATTCGCAGCGTTAGATCATGACCCATTTGGAGAAGATG
C5	Barcode3_35	/5Phos/GGCCAGAGCATTCGCATACCAAGATCATGACCCATTTGGAGAAGATG
D5	Barcode3_36	/5Phos/GGCCAGAGCATTCGCCAGTTCAGATCATGACCCATTTGGAGAAGATG
E5	Barcode3_37	/5Phos/GGCCAGAGCATTCGCCGAAGTAGATCATGACCCATTTGGAGAAGATG
F5	Barcode3_38	/5Phos/GGCCAGAGCATTCGCCGTGAGAGATCATGACCCATTTGGAGAAGATG
G5	Barcode3_39	/5Phos/GGCCAGAGCATTCGCCTCCTGAGATCATGACCCATTTGGAGAAGATG
H5	Barcode3_40	/5Phos/GGCCAGAGCATTCGCGAACTTAGATCATGACCCATTTGGAGAAGATG
A6	Barcode3_41	/5Phos/GGCCAGAGCATTCGCGACTGGAGATCATGACCCATTTGGAGAAGATG
B6	Barcode3_42	/5Phos/GGCCAGAGCATTCGCGCATACAGATCATGACCCATTTGGAGAAGATG
C6	Barcode3_43	/5Phos/GGCCAGAGCATTCGCTCAATGAGATCATGACCCATTTGGAGAAGATG
D6	Barcode3_44	/5Phos/GGCCAGAGCATTCGCTGAGCCAGATCATGACCCATTTGGAGAAGATG
E6	Barcode3_45	/5Phos/GGCCAGAGCATTCGCTGGCATAGATCATGACCCATTTGGAGAAGATG
F6	Barcode3_46	/5Phos/GGCCAGAGCATTCGGAATCTGAGATCATGACCCATTTGGAGAAGATG
G6	Barcode3_47	/5Phos/GGCCAGAGCATTCGCAAGACTAGATCATGACCCATTTGGAGAAGATG
H6	Barcode3_48	/5Phos/GGCCAGAGCATTCGGAGCTGAAGATCATGACCCATTTGGAGAAGATG
A7	Barcode3_49	/5Phos/GGCCAGAGCATTCGGATAGACAGATCATGACCCATTTGGAGAAGATG
B7	Barcode3_50	/5Phos/GGCCAGAGCATTCGGCCACATAGATCATGACCCATTTGGAGAAGATG
C7	Barcode3_51	/5Phos/GGCCAGAGCATTCGGCGAGTAAGATCATGACCCATTTGGAGAAGATG
D7	Barcode3_52	/5Phos/GGCCAGAGCATTCGGCTAACGAGATCATGACCCATTTGGAGAAGATG
E7	Barcode3_53	/5Phos/GGCCAGAGCATTCGGCTCGGTAGATCATGACCCATTTGGAGAAGATG
F7	Barcode3_54	/5Phos/GGCCAGAGCATTCGGGAGAACAGATCATGACCCATTTGGAGAAGATG
G7	Barcode3_55	/5Phos/GGCCAGAGCATTCGGGTGCGAAGATCATGACCCATTTGGAGAAGATG
H7	Barcode3_56	/5Phos/GGCCAGAGCATTCGGTACGCAAGATCATGACCCATTTGGAGAAGATG
A8	Barcode3_57	/5Phos/GGCCAGAGCATTCGGTCGTAGAGATCATGACCCATTTGGAGAAGATG
B8	Barcode3_58	/5Phos/GGCCAGAGCATTCGGTCTGTCAGATCATGACCCATTTGGAGAAGATG
C8	Barcode3_59	/5Phos/GGCCAGAGCATTCGGTGTTCTAGATCATGACCCATTTGGAGAAGATG
D8	Barcode3_60	/5Phos/GGCCAGAGCATTCGTAGGATGAGATCATGACCCATTTGGAGAAGATG
E8	Barcode3_61	/5Phos/GGCCAGAGCATTCGTATCAGCAGATCATGACCCATTTGGAGAAGATG
F8	Barcode3_62	/5Phos/GGCCAGAGCATTCGTCCGTCTAGATCATGACCCATTTGGAGAAGATG
G8	Barcode3_63	/5Phos/GGCCAGAGCATTCGTCTTCACAGATCATGACCCATTTGGAGAAGATG
H8	Barcode3_64	/5Phos/GGCCAGAGCATTCGTGAAGAGAGATCATGACCCATTTGGAGAAGATG
A9	Barcode3_65	/5Phos/GGCCAGAGCATTCGTGGAACAAGATCATGACCCATTTGGAGAAGATG
B9	Barcode3_66	/5Phos/GGCCAGAGCATTCGTGGCTTCAGATCATGACCCATTTGGAGAAGATG
C9	Barcode3_67	/5Phos/GGCCAGAGCATTCGTGGTGGTAGATCATGACCCATTTGGAGAAGATG
D9	Barcode3_68	/5Phos/GGCCAGAGCATTCGTTCACGCAGATCATGACCCATTTGGAGAAGATG
E9	Barcode3_69	/5Phos/GGCCAGAGCATTCGAACTCACCGATCATGACCCATTTGGAGAAGATG
F9	Barcode3_70	/5Phos/GGCCAGAGCATTCGAAGAGATCGATCATGACCCATTTGGAGAAGATG

G9	Barcode3_71	/5Phos/GGCCAGAGCATTCGAAGGACACGATCATGACCCATTTGGAGAAGATG
H9	Barcode3_72	/5Phos/GGCCAGAGCATTCGAATCCGTCGATCATGACCCATTTGGAGAAGATG
A10	Barcode3_73	/5Phos/GGCCAGAGCATTCGAATGTTGCGATCATGACCCATTTGGAGAAGATG
B10	Barcode3_74	/5Phos/GGCCAGAGCATTCGACACGACCGATCATGACCCATTTGGAGAAGATG
C10	Barcode3_75	/5Phos/GGCCAGAGCATTCGACAGATTCGATCATGACCCATTTGGAGAAGATG
D10	Barcode3_76	/5Phos/GGCCAGAGCATTCGAGATGTACGATCATGACCCATTTGGAGAAGATG
E10	Barcode3_77	/5Phos/GGCCAGAGCATTCGAGCACCTCGATCATGACCCATTTGGAGAAGATG
F10	Barcode3_78	/5Phos/GGCCAGAGCATTCGAGCCATGCGATCATGACCCATTTGGAGAAGATG
G10	Barcode3_79	/5Phos/GGCCAGAGCATTCGAGGCTAACGATCATGACCCATTTGGAGAAGATG
H10	Barcode3_80	/5Phos/GGCCAGAGCATTCGATAGCGACGATCATGACCCATTTGGAGAAGATG
A11	Barcode3_81	/5Phos/GGCCAGAGCATTCGATCATTCCGATCATGACCCATTTGGAGAAGATG
B11	Barcode3_82	/5Phos/GGCCAGAGCATTCGATTGGCTCGATCATGACCCATTTGGAGAAGATG
C11	Barcode3_83	/5Phos/GGCCAGAGCATTCGCAAGGAGCGATCATGACCCATTTGGAGAAGATG
D11	Barcode3_84	/5Phos/GGCCAGAGCATTCGCACCTTACGATCATGACCCATTTGGAGAAGATG
E11	Barcode3_85	/5Phos/GGCCAGAGCATTCGCCATCCTCGATCATGACCCATTTGGAGAAGATG
F11	Barcode3_86	/5Phos/GGCCAGAGCATTCGCCGACAACGATCATGACCCATTTGGAGAAGATG
G11	Barcode3_87	/5Phos/GGCCAGAGCATTCGCCTAATCCGATCATGACCCATTTGGAGAAGATG
H11	Barcode3_88	/5Phos/GGCCAGAGCATTCGCCTCTATCGATCATGACCCATTTGGAGAAGATG
A12	Barcode3_89	/5Phos/GGCCAGAGCATTCGCGACACACGATCATGACCCATTTGGAGAAGATG
B12	Barcode3_90	/5Phos/GGCCAGAGCATTCGCGGATTGCGATCATGACCCATTTGGAGAAGATG
C12	Barcode3_91	/5Phos/GGCCAGAGCATTCGCTAAGGTCGATCATGACCCATTTGGAGAAGATG
D12	Barcode3_92	/5Phos/GGCCAGAGCATTCGGAACAGGCGATCATGACCCATTTGGAGAAGATG
E12	Barcode3_93	/5Phos/GGCCAGAGCATTCGGACAGTGCGATCATGACCCATTTGGAGAAGATG
F12	Barcode3_94	/5Phos/GGCCAGAGCATTCGGAGTTAGCGATCATGACCCATTTGGAGAAGATG
G12	Barcode3_95	/5Phos/GGCCAGAGCATTCGGATGAATCGATCATGACCCATTTGGAGAAGATG
H12	Barcode3_96	/5Phos/GGCCAGAGCATTCGGCCAAGACGATCATGACCCATTTGGAGAAGATG

Appendix 5. Frameshift PCR primer oligonucleotide sequences

Appendix Table 6. Frameshift PCR was performed to create frameshift mutations, using four forward and reverse primers with 0-3 N-nucleotides between the binding site and MiSeq indexing primer overhang, identical to those in the 16S-Barcode PCR.

16S_519 forward	**Oligo sequence**
Primer 1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCAGCAGCCGCGGTAATAC
Primer 2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNCAGCAGCCGCGGTAATAC
Primer 3	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNCAGCAGCCGCGGTAATAC
Primer 4	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNCAGCAGCCGCGGTAATAC
S-P-BC reverse	**Oligo sequence**
Primer 1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCATCTTCTCCAAATGGGTCATG AT
Primer 2	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNCATCTTCTCCAAATGGGTCAT GAT
Primer 3	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNCATCTTCTCCAAATGGGTCA TGAT
Primer 4	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNCATCTTCTCCAAATGGGTC ATGAT

Appendix 6. Index PCR Oligonucleotide Sequences

Appendix Table 7. Index PCR was performed using i5 and i7 sequencing indices to label the samples for sequencing runs.

i5 forward	**Oligo sequence**
S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
S508	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
S510	AATGATACGGCGACCACCGAGATCTACACCGTCTAATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
S511	AATGATACGGCGACCACCGAGATCTACACTCTCTCCGTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
i7 reverse	**Oligo sequence**
N701	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
N702	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
N703	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
N704	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
N705	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Appendix 7. Qubit concentrations of samples and sequencing libraries

The concentrations of indexed samples and prepared libraries, measured with Qubit 2.0 Fluorometer (Thermo Fisher) using a dsDNA High Sensitivity Assay kit (Thermo Fisher), are shown in Appendix tables 8-11.

Appendix Table 8. The concentrations of 1,000 and 10,000 bead samples and sequencing library in the initial experiment, measured with Qubit 2.0.

Initial experiment	**Measurement (ng/mL)**
1,000 beads	46.6
10,000 beads	39.0
Library	11.5

Appendix Table 9. The concentrations of frameshift PCR optimisation samples and sequencing library, measured with Qubit 2.0.

Frameshift PCR optimisation	Measurement (ng/mL)
15 cycles 1,000 beads	25.2
15 cycles 10,000 beads	12.1
10 cycles 1,000 beads	32.9
10 cycles 10,000 beads	22.9
Library	12.4

Appendix Table 10. The concentrations of the optimised 16S-barcode PCR sample and sequencing library, measured with Qubit 2.0.

16S-barcode PCR optimisation	Measurement (ng/mL)
20 cycles 1,000 beads	51.2
Library	13.7

Appendix Table 11. The concentrations of the optimised emulsion PCR sample and sequencing library, measured with Qubit 2.0.

Emulsion PCR optimisation	Measurement (ng/mL)
20 cycles emulsion beads	51.0
Library	16.2