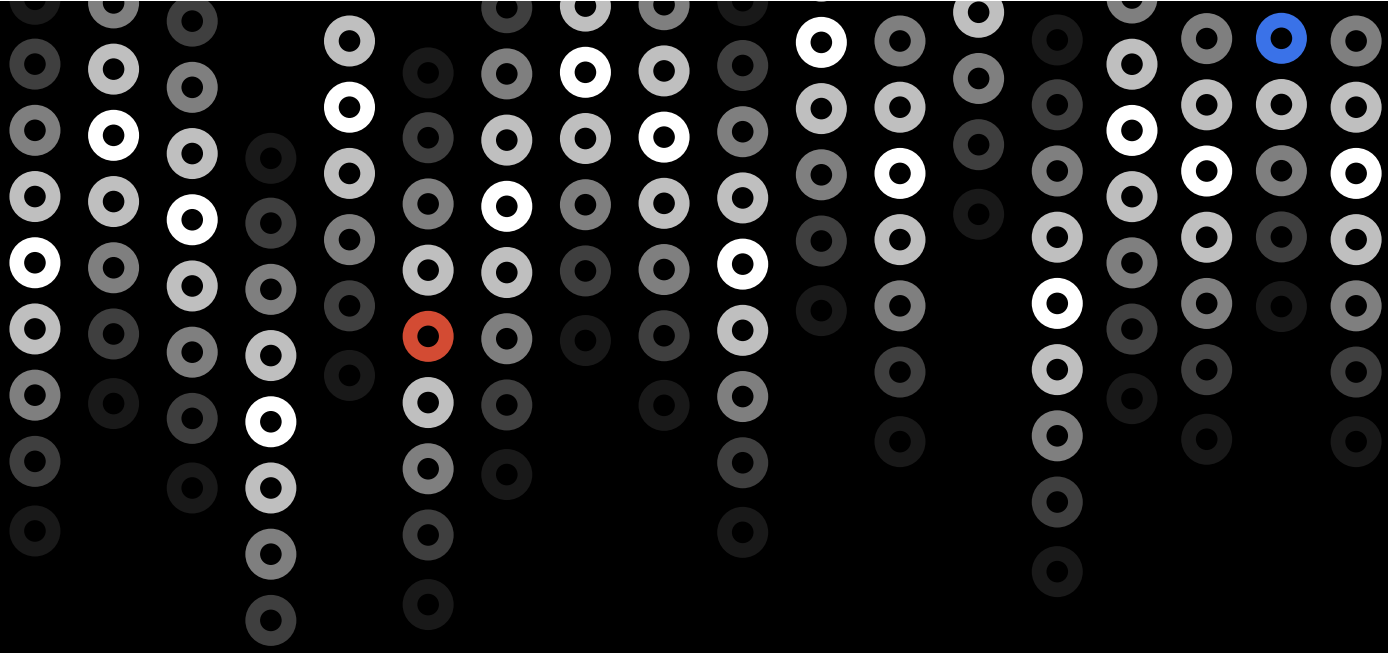

APPLICATION NOTE

DGPM02124281001

INCREASING THE SCALE OF SINGLE-BACTERIAL CELL ANALYSIS BY APPLYING TARGETED SEQUENCING

Atrandi BIOSCIENCES

Increasing the Scale of Single-bacterial Cell Analysis by Applying Targeted Sequencing

SUMMARY

- Our Single Microbe DNA Barcoding Kit allows whole-genome amplification and sequencing of thousands of bacterial cells from standard laboratory strains and heterogeneous environmental samples.
- Here, we introduce another use case of this technology: the targeted sequence enrichment from the barcoded whole genome library. Specifically, we showcase genomic 16S rRNA gene enrichment by employing the myBaits® Expert 16S-Hyb kit.
- Five libraries were enriched and sequenced to demonstrate this application, one of which consisted of standard laboratory strains (and *B. subtilis*) and four of which consisted of environmental samples from three different sources: soil, pond, and saliva.
- By analyzing sequencing data, we confirmed that the 16S genomic targets were enriched 70 to over 500 times compared to the original libraries.
- Importantly, both WGS and targeted sequencing data can be obtained from the same cells.
- The targeted sequencing approach showcases that a deeper understanding of the microbial community and its distribution in the library can be achieved at significantly reduced sequencing costs.

CONTENTS

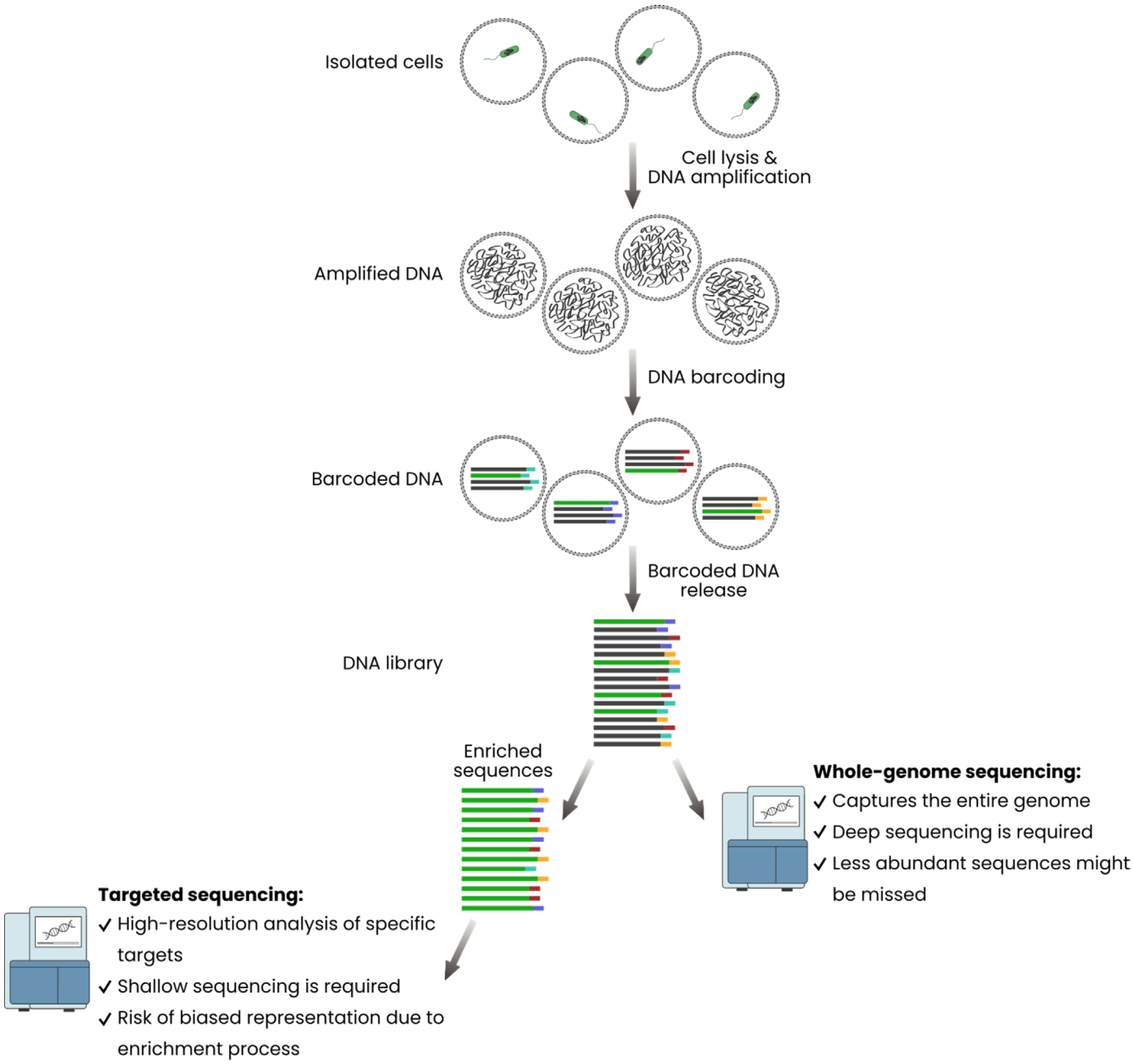

Figure 1. Key steps for targeted sequencing from single-barcoded bacterial cells and comparison to single-cell whole-genome sequencing.

INTRODUCTION

With the constant advancements in next-generation sequencing, the volume of genomic data generated is rapidly increasing. This technology allows for the simultaneous sequencing of multiple samples at high depth. However, many samples are derived from heterogeneous sources, such as mixed cultures of uncultivable organisms or cell cultures that require single-cell resolution to draw accurate conclusions. As a result, there is a growing need for improved methods to trace and differentiate the sources of genomic information accurately.

One solution to this challenge emerged from several technologies that developed sequentially. It began with droplet microfluidics, which enabled scientists to create millions of individual compartments to capture and process cells individually. This advancement paved the way for in-droplet barcoding methods, simplifying the tracing of amplified genomic sequences back to individual organisms [1]. Despite significant progress in the field of single-cell analysis, applying molecular biology reactions within droplets presents challenges, the main one being the difficulty of conducting multi-step reactions due to the technical challenges of modifying droplet content [2] and the inability to remove reagents.

Subsequently, the latest technological advancement came with semi-permeable capsules (SPCs), which opened up new possibilities for single-cell processing [3, 4]. SPCs are water compartments surrounded by a thin, semi-permeable membrane that retains isolated cells and cell-derived nucleic acids while allowing smaller molecules, such as lysis reagents, proteins, and PCR primers, to diffuse between the compartments. Due to this selective permeability, multi-step enzymatic reactions can be conducted by changing the buffer where SPCs are suspended. These unique SPC properties were successfully implemented in our single microbe DNA barcoding technology [5], which allows up to 10,000 bacterial cells to be processed in parallel, including harsh lysis, gDNA purification, whole-genome amplification, split-and-pool barcoding, to obtain single amplified and barcoded genomes. In addition to single-cell whole-genome sequencing, the obtained library can be further processed to enrich specific targets of interest. This is typically done when only a specific set of targets is necessary to gain biological insights about the sample. Furthermore, targeted sequencing can be useful for getting the initial information and making decisions before performing deep whole-genome sequencing.

Many common targeted sequencing methods rely on hybridization, where complementary sequences bind to target sequences of interest, allowing for their retention while unwanted sequences are washed away. This approach commonly uses biotinylated DNA or RNA probes (also known as baits) that hybridize with the target sequences and are then captured using streptavidin-coated beads or substrates. The main advantage of hybridization baits is their flexibility, allowing the creation of probe sets that can target an almost unlimited range of sequences. Baits are widely used for enriching environmental DNA libraries, where contamination with unrelated background DNA is common, particularly in older samples or when target sequences are present in only trace amounts. They are also frequently used in re-sequencing, enabling a more in-depth analysis of target sequences without requiring extensive deep sequencing of the entire library [6, 7, 8]. Other frequently used tools include DNA capture arrays, flow cells, or wells with surface-bound probes, which bind to and retain the target sequences while unbound fragments are removed. [9, 10]. An alternative approach, known as 'capture by circularization,' uses molecular inversion probes (MIPs). MIPs are single-stranded DNA oligonucleotide probes with primer sequences at both ends. Following gap filling and a ligation reaction, they form circularized DNA. Exonucleases are then used to degrade linear DNA, leaving only the circular DNA containing the sequences of interest, which can subsequently be re-amplified to enrich the library [9]. Another common approach for sample preparation for targeted sequencing is multiplex PCR, which amplifies multiple target regions in a single PCR reaction. This approach uses multiple pairs of primers designed to amplify specific regions of interest, allowing for efficient and cost-effective enrichment. Multiplex PCR is

particularly advantageous when working with a limited number of targets or when rapid enrichment of known sequences is required [11].

In this application note, we demonstrate the enrichment of targets of interest from whole-genome libraries using a hybridization-based approach. Specifically, we aimed to enrich 16S rRNA gene sequences from libraries prepared with standard laboratory strains, *E. coli* and *B. subtilis*, as well as environmental samples. The results provided in this application note show that our Single Microbe DNA Barcoding Kit is fully compatible with standard enrichment techniques, thereby expanding its applications.

STUDY OVERVIEW

This study demonstrates the feasibility of targeted sequence enrichment from the amplified and barcoded genomes using the myBaits biotinylated DNA probe kit. For this demonstration, we chose 16S rRNA gene (hereafter referred to as 16S) sequences as the enrichment target. Overall, we used a total of five libraries generated using our Single Microbe DNA Barcoding Kit (CKP-BARK1). Specifically, one library was prepared with standard bacterial strains, *E. coli* MG1655 and *B. subtilis* 23857, and four libraries were prepared from the environmental samples, including soil, pond, and saliva (refer to Table 4 for more details). All five libraries were sequenced before and after conducting 16S target enrichment experiments.

Before conducting 16S target enrichment, each library was prepared using the standard procedure described in the Single Microbe DNA Barcoding Kit's User Guide (refer to Figure 2). The first step was to obtain SPCs with amplified gDNA. For this, single bacterial cells were isolated in SPCs, lysed, and then SPCs with purified gDNA were used for the whole-genome amplification reaction. The second step was library preparation, including debranching, split-and-pool barcoding, final fragmentation, sequencing adapter ligation, and barcoded DNA enrichment. Note that libraries can be sequenced after this step to obtain genome-wide resolution. Finally, the library was incubated with biotinylated probes to enrich sequences of interest, which were then captured on streptavidin beads, followed by the final target sequence enrichment.

The application note is divided into four main sections for better readability. The first section details how the whole-genome library should be prepared prior to the enrichment step. The second section describes the complete enrichment procedure from start to finish. The third section outlines the analyses performed on the sequenced libraries, while the final section presents the anticipated results.

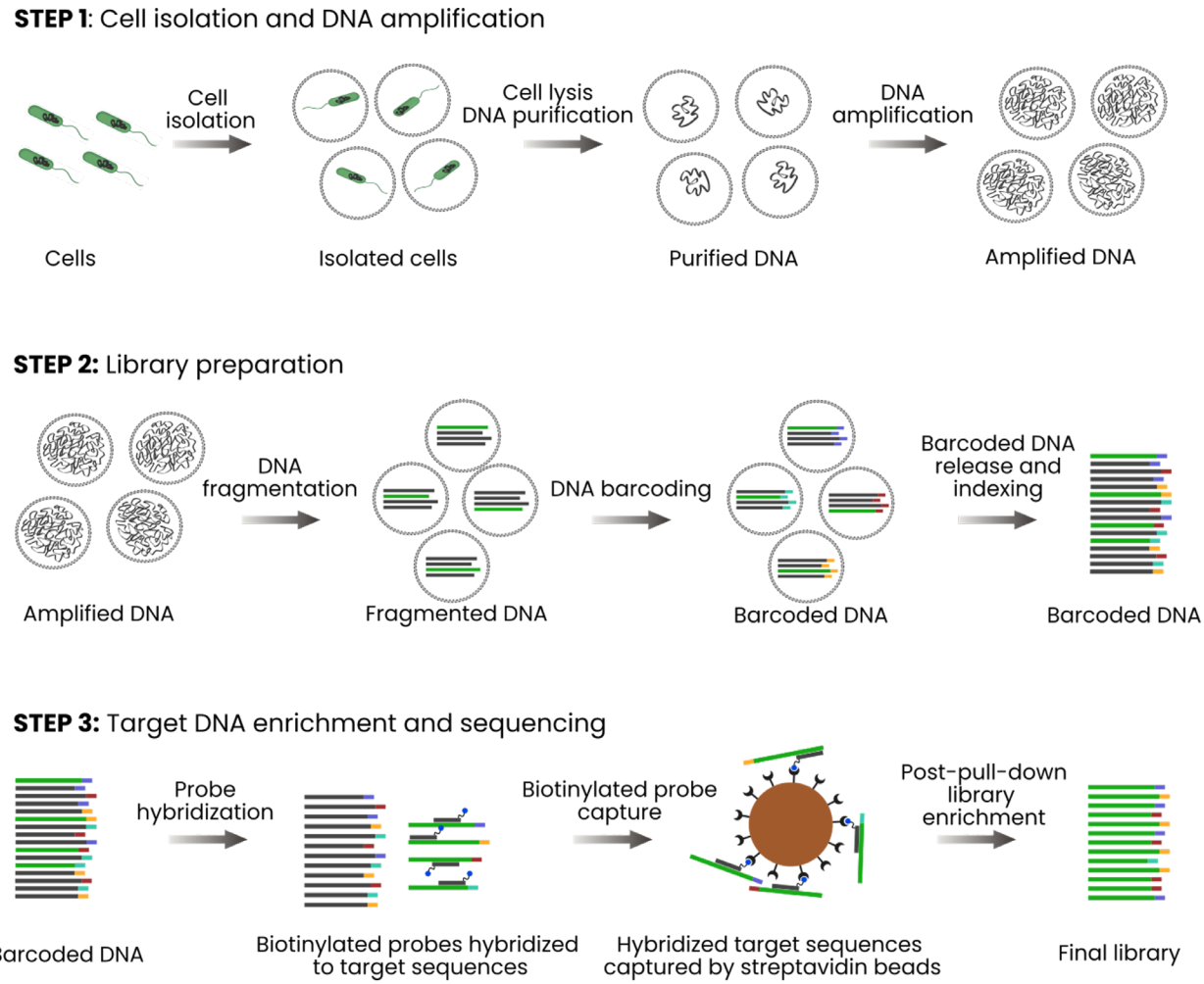

Figure 2. A detailed workflow for obtaining an enriched single-cell barcoded library. Note that the experimental steps described in Steps 1 and 2 are from the Single Microbe DNA Barcoding Kit.

EQUIPMENT AND OTHER ITEMS

Table 1. *Laboratory equipment required to complete this protocol.*

Category	Reagent/consumable	Manufacturer	P/N
DNA barcoding	Single Microbe DNA Barcoding Kit	Atrandi Biosciences	CKP-BARK1
Target enrichment	Nuclease-free water	Invitrogen	10977-035
	myBaits Expert 16S-Hyb kit	Arbor Biosciences	308616.v5
PCR amplification	NEBNext® Ultra II Q5® Master Mix	NEB	M0492S
	Primers (see Table 3)	-	-
DNA purification	Ethanol (96%)	-	-
	AMPure XP Reagent	Beckman	A63881

Table 2. *Laboratory equipment used in the demonstrated experiments.*

Item	Manufacturer
Agilent High Sensitivity DNA Kit	Agilent Technologies
Agilent Bioanalyzer	Agilent Technologies
Qubit™ 1X dsDNA High Sensitivity (HS)	Thermo Fisher Scientific
Qubit 4 Fluorometer	Thermo Fisher Scientific
Mastercycler™ X50s or similar thermocycler	Eppendorf or MLS
Concentrator plus/Vacufuge® plus	Eppendorf
Centrifuge with swinging bucket rotor	MLS*
Vortex mixer	MLS
Non-stick 1.5 mL and 0.2 mL tubes	MLS
Pipettors: 10–µL, 20–µL, 200–µL, 1000–µL	MLS
Agilent Bioanalyzer	Agilent Technologies
Agilent High Sensitivity DNA kit	Agilent Technologies
Qubit™ Fluorometer with dsDNA HS Assay kit	Thermo Fisher Scientific

Table 3. *Primer sequences used in this application note. Prepare 5 µM (each) primer mixtures for each amplicon before starting the experiment.*

Indexing primer pair	Primer	Sequence (5‘-3‘)
Ix3	p5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
	p7_ix 3	CAAGCAGAAGACGGCATACGAGATGGTTCAGTGACTGGAGTTCAGACGTGT
Ix4	p5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
	p7_ix 4	CAAGCAGAAGACGGCATACGAGATTTGGACGTGACTGGAGTTCAGACGTGT
Ix5	p5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
	p7_ix 5	CAAGCAGAAGACGGCATACGAGATACCACTGTGACTGGAGTTCAGACGTGT
Ix6	p5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
	p7_ix 6	CAAGCAGAAGACGGCATACGAGATCAGTGGGTGACTGGAGTTCAGACGTGT
Ix8	p5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC
	p7_ix 8	CAAGCAGAAGACGGCATACGAGATTGACAAGTGACTGGAGTTCAGACGTGT

PART 1. LIBRARY PREPARATION FOR TARGET ENRICHMENT

Table 4. *SAG libraries used in these experiments. Note that all libraries had different indexing primer pairs, which allowed us to sequence all libraries in a single NextSeq run.*

Category	Reagent/consumable	Expected number of cells	Additional information
EC-BS	*E. coli* MG1655 and *B. subtilis* 23857 were separately encapsulated in SPCs. After SPC pooling (to reach a 1:1 species ratio), SPCs were processed using the Single Microbe DNA Barcoding Kit.	6600	Library index: Ix8
ENV-1	Bacterial samples collected from pond, soil, and saliva were separately encapsulated in SPCs, followed by processing using the Single Microbe DNA Barcoding Kit. Different source samples were multiplexed in a single barcoding run.	2000	Library index: Ix3
ENV-2		2000	Library index: Ix4
ENV-3		2000	Library index: Ix5
ENV-4		2000	Library index: Ix6

SAG library amplification

1) Prepare the reaction as follows:

Reagent	Volume, µL
SAG library	X (40 ng of DNA library)
Nuclease-free Water	80–X
Indexing primer Mix (5 µM each)	20
NEBNext® Ultra II Q5® Master Mix	100
Total:	**200**

2) Optional: Split 200 µL PCR mixes into four tubes (50 µL per tube).
3) Apply the following PCR cycling program:

Cycle step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	1
Denaturation	98 °C	20 s	10
Annealing	54 °C	30 s	
Extension	72 °C	20 s	
Final extension	72 °C	1 min	1
Hold	4 °C	-	-

4) After PCR amplification, follow with 0.6–0.8X double-sided AMPure XP purification as described below.

0.6–0.8X double-sided AMPure XP purification

1) Resuspend the AMPure XP magnetic beads. Add 60 µL of resuspended AMPure XP beads to the 100 µL sample after PCR and mix by pipetting.
2)Incubate at room temperature (20 to 25°C) for 5 minutes.
3) Briefly spin the sample down and place the tube on a magnetic rack. Keep the tube on a magnet and transfer the supernatant into a clean PCR tube. Keep the supernatant!
4) Add 20 µL of fresh resuspended AMPure XP beads to the supernatant from the previous step and mix by pipetting.
5) Incubate for 5 minutes at room temperature (20 to 25°C).
6) Briefly spin the sample down and place the tube on a magnetic rack. Keep the tube on a magnet and remove the supernatant.
7) Wash the beads by adding 180 µL 80% ethanol without disturbing the pellet and incubate for ~30 s at room temperature (20 to 25°C).
8) Remove and discard ethanol. Repeat 3–4 steps.
9) Briefly spin down and place the tube back on a magnetic rack. Remove any residual supernatant.
10) Allow the beads to dry for 1 minute.
11) Remove the tube from the magnetic rack and resuspend the pellet in 32 µL nuclease-free water. Incubate for 5 minutes at room temperature (20 to 25°C).
12) Pellet the beads on a magnetic rack. Wait until the eluate is clear and colorless, and transfer it to a 1.5–mL DNA LoBind tube.
13) Measure the DNA concentration by Qubit using a dsDNA HS kit.

PART 2. 16S FRAGMENT ENRICHMENT USING THE myBaits Expert 16S-Hyb kit

In this application note, we utilized the myBaits Expert 16S-Hyb kit to target bacterial 16S encoding DNA sequences. The kit consists of 37,745 baits, which were derived from the 16S reference sequence database Greengenes (v. 13.5). All featured 16S sequences have at least one bait with no less than 78% identity over all of its regions and, on average, 3000 probes covering each nucleotide position [12].

Library preparation

1) Each library enrichment is prepared separately. Capture is carried out with 2 µg of DNA library dissolved in 7 µL of nuclease-free water. If the library is not at the working volume of 7 µL, follow steps 2 and 3.
2) Concentrate or dry the DNA library in a vacuum concentrator (Speed Vac) under V-AQ mode. Do not apply heat.
Notes: For a library of 100 µL in a PCR tube, Eppendorf “Concentrator plus” takes about 2.5 hours to dry. If hybridization is not performed on the same day, store the samples at –20°C.
3) Before hybridization, dissolve or dilute concentrated library up to 7 µL using nuclease-free water.

16S sequence hybridization and enrichment

1) Heat the Hyb S and Hyb N reagents to 60°C and vortex them until all precipitate dissolves before preparing the Hyb Mix — "HYB".

2) For each enrichment reaction, prepare the following mixes; pipetting error is built in:

Hybridization Mix	
Component	µL / Reaction
Hyb N	9.25
Hyb D	3.5
Hyb S	0.5
Hyb R	1.25
Baits	5.5
Total:	**20 µL**

Blockers Mix	
Component	µL / Reaction
Block X	0.5
Block C	2.5
Block O	2.5
Total:	**5.5 µL**

3) After pre-warming the Hybridization mix for 10 minutes at 60°C, take an aliquot of 18.5 µL of Hybridization Mix—now "HYBs"—for each reaction.

4) For each reaction, take an aliquot of 5 µL of Blockers Mix and then add 7 µL of each library – now "LIBs".

5) Incubate the LIBs in the thermal cycler for 5 minutes at 95°C and then drop to the hybridization temperature (65°C). Be sure to use a heated lid (105°C).

6) Put the HYBs in the thermal cycler and warm to the hybridization temperature for 5 minutes.

7) Transfer 18 µL of each HYB to each LIB, mix by pipetting, and incubate for 16–24 hours (the current target is 20–24 hours).

8) 1.5 hours before step 6, prepare Wash Buffer X. Vortex thoroughly and warm to the hybridization temperature for at least 45 minutes.

Blockers Mix	
Component	Volume for one reaction
Hyb S	10 µL
Nuclease-Free Water	990 µL
Wash Buffer	250 µL
Total:	**1.25 mL**

9) Prepare 30 µL of beads per reaction by washing three times in 200 µL Binding Buffer.
10) Resuspend the washed bead aliquots in 70 µL Binding Buffer and warm the suspensions to the hybridization temperature (65°C) for at least 2 minutes.
11) Combine the warmed beads with the hybridization reactions and incubate for 5 minutes at the hybridization temperature, agitating at 2.5 minutes to keep the beads suspended.
12) Pellet the beads and remove the supernatant. If using microcentrifuge tubes for cleanup, wash the beads three times with 375 µL warmed Wash Buffer X, incubating for 5 minutes at the hybridization temperature (65°C).
16) Resuspend the beads in 30 µL Buffer E and then use the full supernatant volume with beads for PCR amplification reactions.

Enriched target amplification

1) Prepare the reaction as follows:

Reagent	Volume, µL
Purified DNA	X (~30 µL of eluted library with streptavidin beads)
Nuclease Free Water	40–X
Indexing primer Mix (5 µM each)	10
NEBNext® Ultra II Q5® Master Mix	50
Total:	**100**

2) Optional: Split 100 µL PCR mixes into two tubes (50 µL per tube).
3) Apply the following program:

Cycle step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	1
Denaturation	98 °C	20 s	10
Annealing	54 °C	30 s	
Extension	72 °C	20 s	
Final extension	72 °C	1 min	1
Hold	4 °C	-	-

4) After PCR amplification, follow with 0.6–0.8X double-sided AMPure XP purification.

0.6–0.8X double-sided AMPure XP purification

1) Resuspend the AMPure XP magnetic beads.
2) Add 60 µL of resuspended AMPure XP beads to the 100 µL sample after PCR and mix by pipetting.
3) Incubate at room temperature (20 to 25°C) for 5 minutes.
4) Briefly spin the sample down and place the tube on a magnetic rack. Keep the tube on a magnet and transfer the supernatant into a clean PCR tube. Keep the supernatant!
5) Add 20 µL of fresh resuspended AMPure XP beads to the supernatant from the previous step and mix by pipetting.
6) Incubate for 5 minutes at room temperature (20 to 25°C).
7) Briefly spin the sample down and place the tube on a magnetic rack. Keep the tube on a magnet and remove the supernatant.
8) Wash the beads by adding 180 µL 80% ethanol without disturbing the pellet and incubate for ~30 s at room temperature (20 to 25°C).
9) Remove and discard ethanol.
10) Repeat the previous step.
11) Briefly spin down and place the tube back on a magnetic rack. Remove any residual supernatant.
12) Allow the beads to dry for 1 minute.
13) Remove the tube from the magnetic rack and resuspend the pellet in 32 µL nuclease-free water. Incubate for 5 minutes at room temperature (20 to 25°C).
14) Pellet the beads on a magnetic rack. Wait until the eluate is clear and colorless and transfer it to a 1.5–mL DNA LoBind tube.
15) Measure the DNA concentration by Qubit using a dsDNA HS kit.
Note: If library DNA is undetectable after 10 PCR cycles, amplify the whole library with an additional 10 PCR cycles. In the experiments described here, 30 µL of post-capture bead suspension required 20 amplification cycles consisting of two 10–cycle runs.
16) Run the Agilent Bioanalyzer for library QC, as described in the "Agilent High Sensitivity DNA Kit Quick Start Guide" protocol (refer to Figure S1 for expected library profiles).

PART 3. DATA ANALYSIS

A preliminary analysis of demultiplexed enriched library sequencing data was performed to determine the fold enrichment of 16S ratio change, the breadth of 16S sequence coverage by sequenced reads, and the species associated with each barcode (Figure 3). Prior to the analysis, Illumina outputs were processed using a barcode demultiplexing pipeline specifically designed for data generated with the Atrandi Biosciences' Single-Microbe DNA Barcoding Kit.

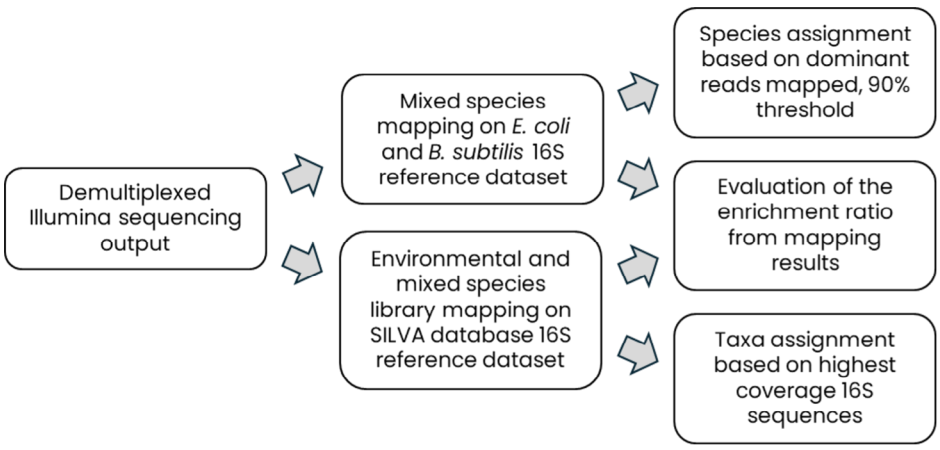

Figure 3. Schematic depiction of data analysis.

Data analysis consisted of 4 steps:

1) Mapping enriched and pre-enriched sequenced library data on SILVA or *E. coli* and *B. subtilis* 16S (EcBs 16S) reference datasets using minimap2.

2) Extracting the percentage of reads which could be mapped and could not be mapped from the mapping data and determining 16S percentage before and after enrichment.

3) Extracting mapping data with coverage information per individual 16S and reads mapped per each 16S sequence.

4) Finally, depending on the sequenced library type, one of two different approaches was applied:

- For the ENV libraries, the closest single shared taxonomic level was assigned to each barcode based on identified organisms with the highest 16S coverage. The assigned taxa were then visualized in taxonomic abundance charts.
- For the EC-BS library, species were assigned to each barcode based on the dominant proportion of mapped reads. A 90% read threshold (out of all mapped reads) was applied to determine whether a barcode belonged to *E. coli* or *B. subtilis*. These results were then displayed in a species mixing chart.

PART 4. ANTICIPATED RESULTS

Evaluation of enrichment based on the read mapping

Table 5. *Enrichment results of all five libraries.*

	Mixed EC-BS – minimap2		ENV – minimap2 – SILVA			
	SILVA	EcBs 16S	ENV1	ENV2	ENV3	ENV4
Reads	136 049 580		26 981 472	56 823 406	29 026 604	52 373 252
Reads mapped	59 306 171	125 764 323	16 654 331	36 328 932	20 529 907	41 210 504
Reads unmapped	76 743 409	10 285 257	10 327 141	20 494 474	8 496 697	11 162 748
% of 16S reads	**43.6**	**92.4**	**61.7**	**63.9**	**70.7**	**78.7**
% of 16S in pre-enriched library	**0.241**	**0.174**	**0.319**	**0.302**	**1.011**	**0.264**
16S reads in pre-enriched library	608 580	440 926	971	801	1386	981
Reads in pre-enriched library	253 024 392		304 250	265 016	137 354	371 308
Enrichment ratio	180.9	531.3	193.5	211.7	70.0	298.1

The results in **Table 5** show that the myBaits kit can enrich the 16S sequences by 70– to over 500–fold, depending on the library, with a maximum of around 90% of all reads being 16S. Fold enrichment is calculated by dividing the percentage of 16S reads in the post-enrichment dataset by the percentage of 16S reads in the pre-enrichment dataset. Note, that the enrichment ratio value is particularly sensitive to the denominator, i.e. the estimate of the fraction of 16S reads in the pre-enrichment library, which tends to vary across libraries and sample types. Another observed source of variability is the choice of applied analysis methods. For example, in Table 5, the EC-BS library was mapped to two different reference datasets. One was from the SILVA database, which includes rRNA gene sequences from the Bacteria, Archaea, and Eukarya domains, while the other consisted of only *E. coli* and *B. subtilis* 16S sequences. Both cases delivered significantly different conclusions about 16S enrichment efficiency. While only 43.6% of reads were mapped to the SILVA database, 92.4% were mapped to the *E. coli* and *B. subtilis* 16S reference. This corresponded to a 181–fold enrichment using SILVA and a 531–fold enrichment using the *E. coli* and *B. subtilis* 16S reference. This inconsistency possibly emerged from minimap2 default parameters not being fine-tuned for mapping a large number of low-diversity sequences on such varied reference library as SILVA. For this reason, EC-BS library mapping result analysis was carried out using only the EcBs 16S reference dataset.

Another tested 16S identification approach to calculate fold enrichment involves searching for high-conserved regions of the 16S sequence **(Table S1)**, which in enriched libraries analysis produced results similar to minimap2, with 16S content ranging from 55% to 89%, and a 3% to 8% difference between the two methods. This conserved region search allows fast preliminary evaluation of enrichment. However, it underestimates the fraction of 16S reads in pre-enrichment libraries, likely due to shallow sequencing, resulting in a 2– to 10–fold lower estimate of 16S-mapping reads compared to minimap2 outputs **(Table S2)**.

An important conclusion derived from the calculated enrichment ratio is that targeted enrichment serves as a valuable tool in improving the sequencing depth of regions of interest. In a hypothetical scenario, the 16S coverage depth that would require 20 billion unpaired reads in a whole-genome library (NovaSeq 6000) can be achieved by sequencing a 70–fold enriched library with only 285.7 million unpaired reads (71.4% of the capacity of a NextSeq 550). In contrast, a 500–fold enriched library would deliver the same coverage depth with just 40 million unpaired reads (10% of a NextSeq 550).

Mixed species EC-BS library mapping results

The EC-BS library was mapped to the *E. coli* (Accession: NC_000913.3) and *B. subtilis* (Accession: NC_000964.3) 16S reference dataset (EcBs 16S), which consisted of extracted 16S sequences from both species. The mapping results were summarized by counting the number of reads mapped to the 16S sequences corresponding to each species for every barcode. Species assignment was performed using the combined data of *E. coli* and *B. subtilis* mapped reads. A barcode was assigned to E. coli if more than 90% of the counted 16S reads corresponded to that organism. Likewise, a barcode was assigned to B. subtilis if more than 90% of the counted 16S reads corresponded to that organism. Cases under the 90% threshold were labeled as 'Mixed.' The final data was visualized in species mixing graphs **(Figure 4A).**

At a sequencing depth of 104–105 **(Figure 4B)**, 91% of all barcodes had at least one 16S sequence with full 100% coverage **(Figure 4C)**.

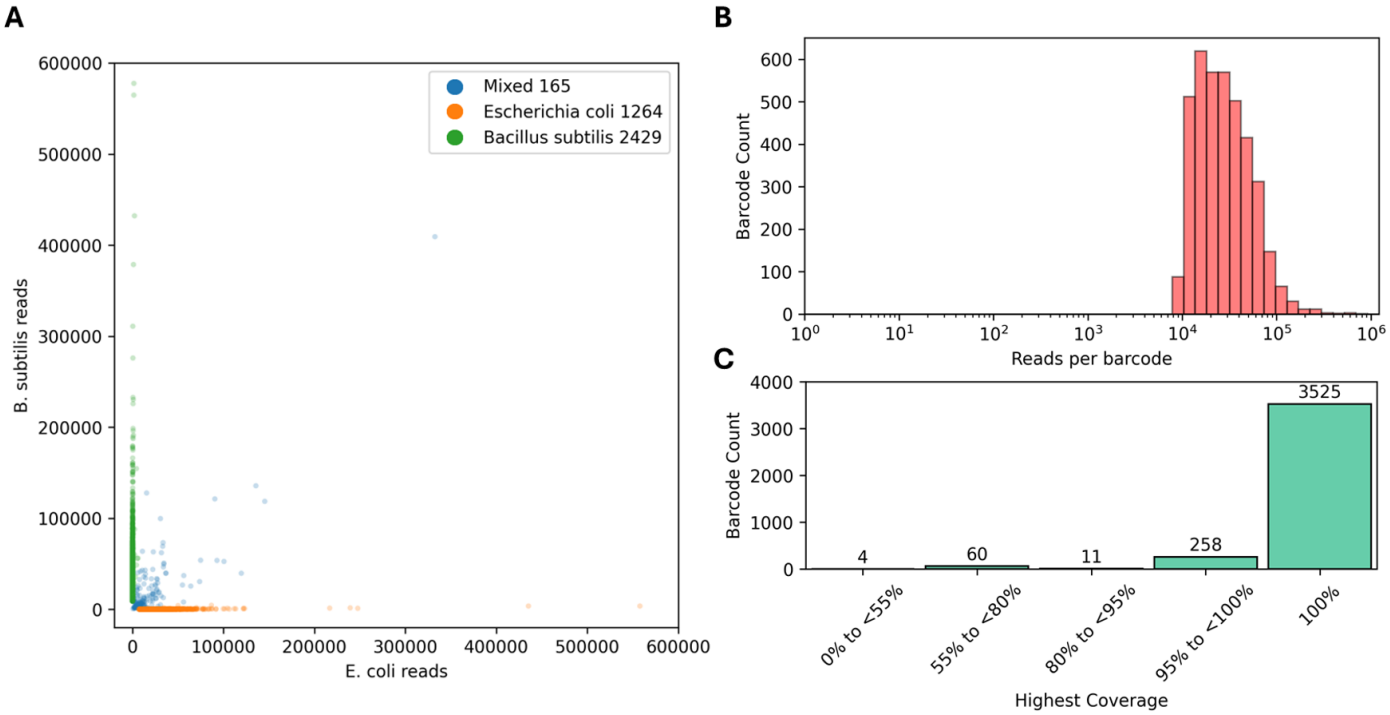

Figure 4. Species mixing chart (A), reads per barcode (B), and highest coverage distribution per barcode (C) charts.

Using assigned species and coverage per barcode data, a coverage probability heatmap was calculated (**Figure 5**). This heatmap shows the percentage of barcodes (n(*E. coli*) = 1264, n(*B. subtilis*) = 2429) covering at least once specific nucleotides for each 16S sequence. The main observation is that most 16S sequences are fully covered across all barcodes, with blue and green gaps indicating individual nucleotide differences between the reference and the sequenced organism. Although some regions display lower coverage probability, these do not suggest any significant bias in target selection during hybridization with the myBaits enrichment kit.

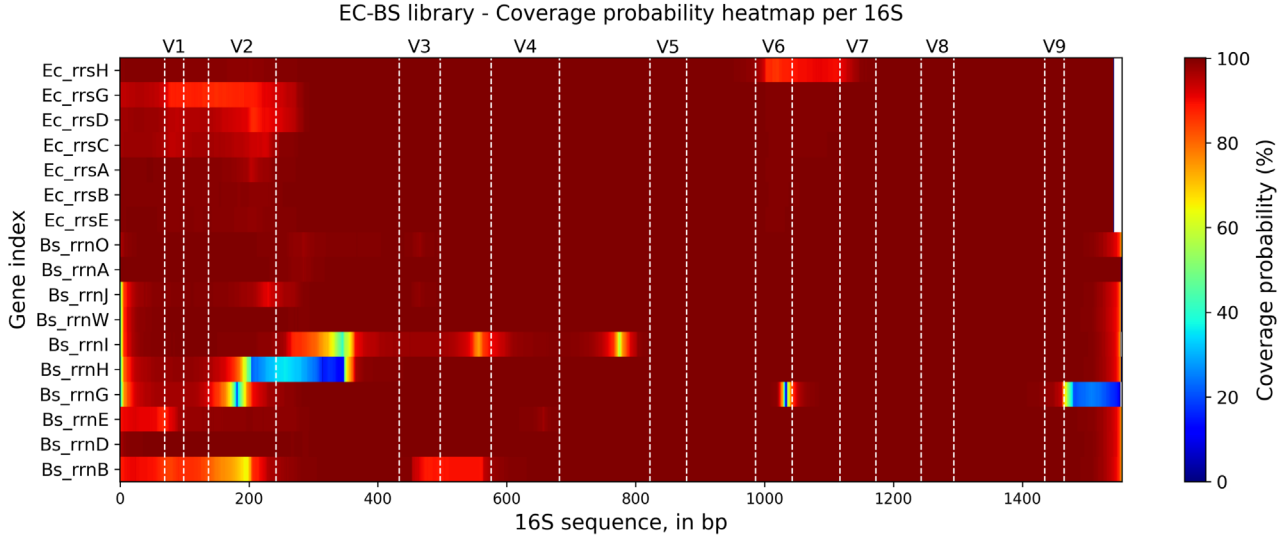

Figure 5. Coverage distribution probability heatmap. The color indicates the percentage of barcodes with a given nucleotide covered at least once. The gene indexes on the Y-axis represent the organism and 16S gene, Ec being *E. coli*, and Bs-*B. subtilis*. V1 - V9 – represents approximate hypervariable region positions on the heatmap.

As we explored additional visualizations of coverage per barcode, we calculated the count of 16S sequences for each barcode using coverage breadth thresholds of >99% and exactly 100% (**Figure S2A** and **B**). These results, consistent with those in **Figure 5**, reveal that *B. subtilis* exhibits less uniformity in 16S coverage compared to *E. coli*. Another visualization, applying the same thresholds, presented the counts for each 16S gene index (**Figure S2C**), which also indicated that the lowest coverage occurred in the *B. subtilis rrnI*, *rrnH*, and *rrnG* 16S genes.

To summarize, based on mixed known-species library analysis and coverage distribution heatmap, single-cell whole-genome libraries are fully compatible with probe-based enrichment.

Environmental library mapping results

Environmental sample libraries ENV1, ENV2, ENV3 and ENV4 were processed by mapping their reads to 16S reference sequences from SILVA (SILVA_138.2_SSURef_NR99) database. The mapping results were filtered per barcode to retain only cases with mapped 16S sequences, while also removing barcodes where the taxonomic assignment was likely to be incorrect or ambiguous. The following rules were applied:

- If the 16S sequence with the highest coverage achieves 100%, then only those 16S sequences with 100% coverage are retained.
- If the highest coverage is X% (where X is less than 100%), all 16S reference sequences with coverage between (X - 1%) and X are retained.
- If a barcode's highest 16S coverage is less than 8%, or if the number of mapped unpaired reads is below 20, the barcoded genome is discarded due to insufficient data.

The application of minimal coverage and read count rules, varying by library, removed 3% to 8% of reads and 35% to 45% of all barcodes in each ENV library (**Figures 6** and **7**).

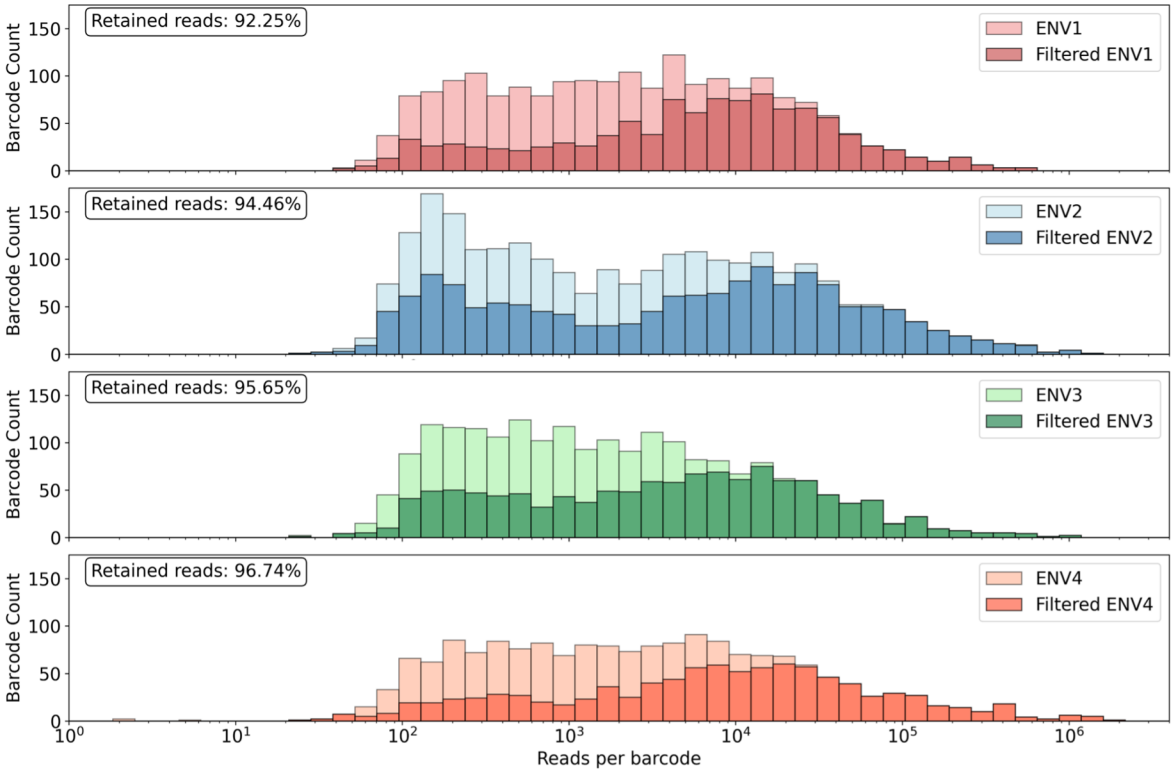

Figure 6. Read per barcode distribution charts, before filtering and after filtering.

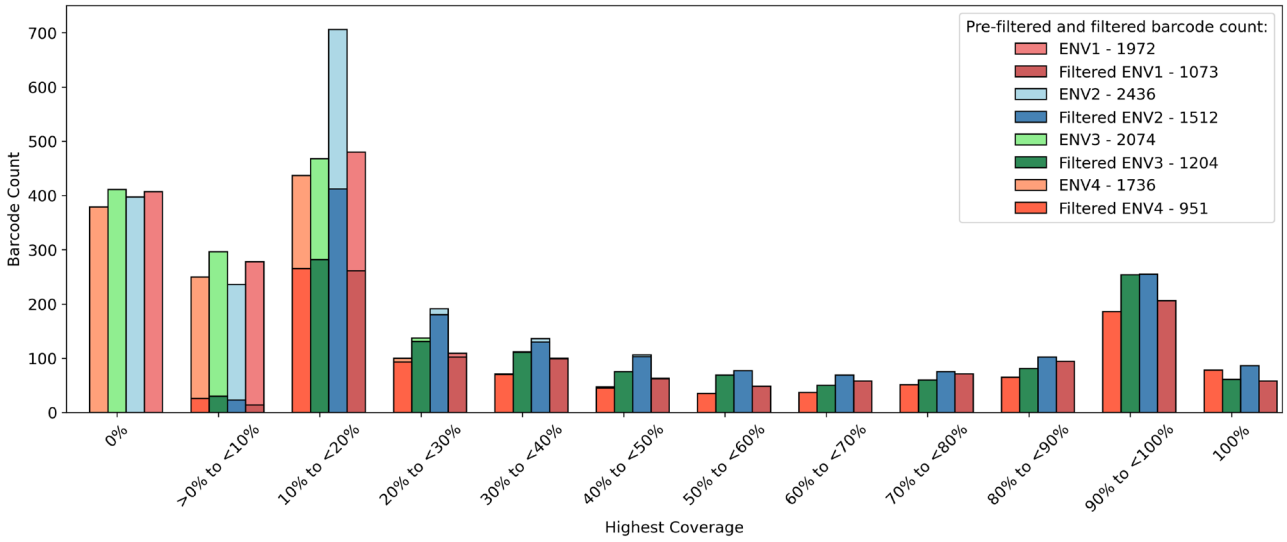

Figure 7. Highest 16S coverage per SAG distribution before filtering, and the bottom figure after filtering.

Extracted and filtered barcodes from all ENV libraries were combined, after which taxa for each barcode were assigned based on the following parameters:

- Taxa are assigned at the lowest possible taxonomic level where only a single taxonomic name exists.
- Species are assigned only if there is a unique species-level taxon, the 16S coverage is greater than 99%, and the assignable species is not labeled as "unclassified", "uncultivated", or any similar undefined designation.
- If the lowest assignable taxonomic level is *incertae sedis*, a higher taxonomic level is considered.
- If conflicting taxa are present at the phylum level (i.e. two different phyla can be assigned), the case is categorized as "Mixed Species".

Following taxa assignment, all ENV libraries were demultiplexed to trace back the sample origin - soil, pond or saliva. The final output consists of taxonomic abundance charts (**Figure 8, 9** and **10**), which display the depth and diversity of known bacterial species in the collected samples. Moreover, these charts indicate the "unknown", "unsequenced" and "unclassified" species, which appear as white spaces.

Based on the observed results, saliva samples contained the well-characterized and extensively studied taxa, with the highest abundance of individual barcodes assigned at the genus or species levels. The dominant phyla in saliva samples included Pseudomonadota, Bacillota, Bacteroidota, Patescibacteria, and Actinomycetota (**Figure 8**), aligning with the typical phylum composition of saliva [13]. An exception was Patescibacteria, also known as 'Candidate Phyla Radiation' (CPR) - a large infrakingdom of uncultivable bacteria, primarily known from metagenomic and sequencing data.

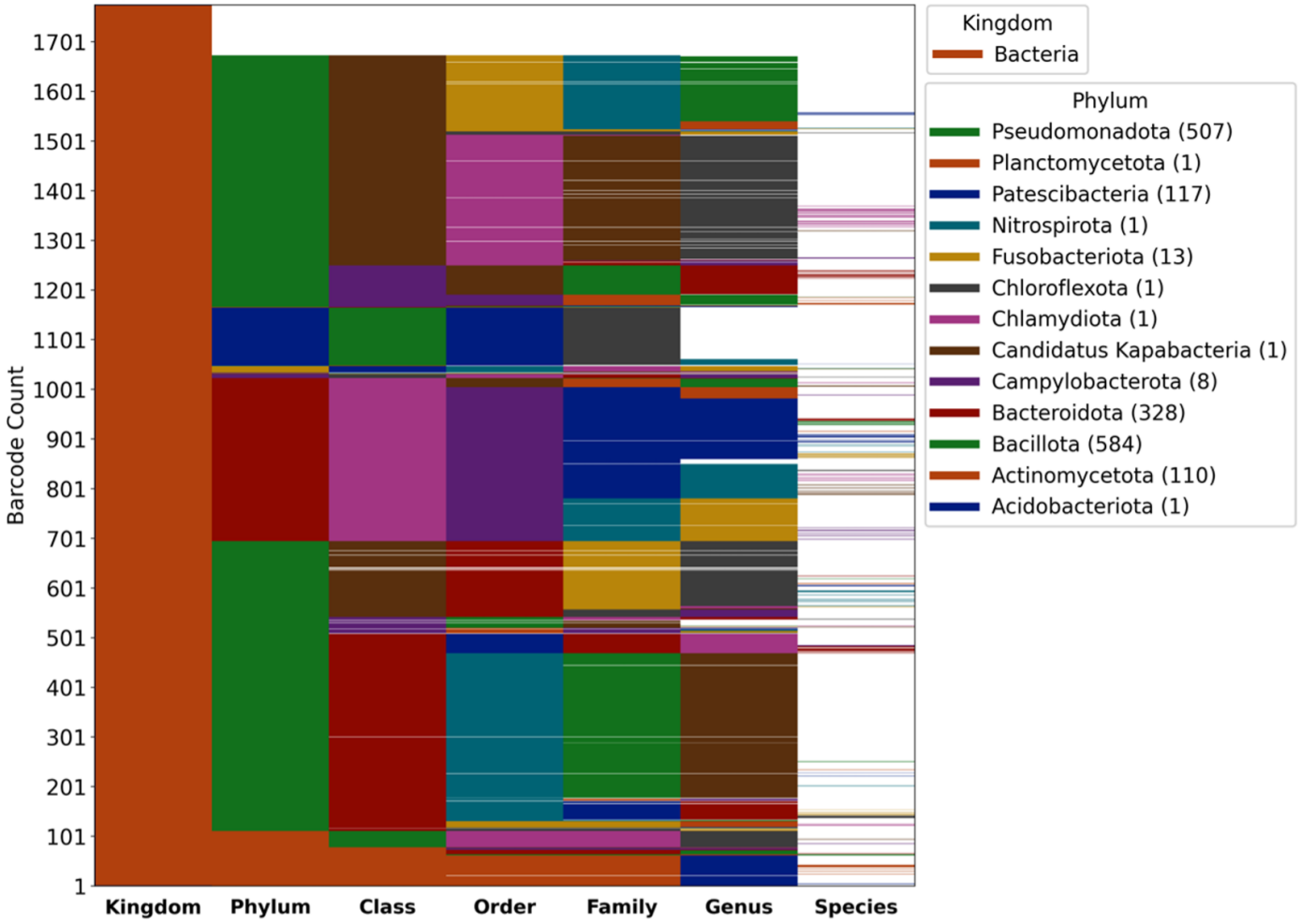

Figure 8. Abundance chart of identified taxons in the saliva sample. Barcodes with no assigned phylum signify mixed species cases.

Pond samples, in contrast, had fewer well-defined species with most barcodes assigned at the genus or family level. The dominant phyla included Pseudomonadota, Planctomycetota, Bacteroidota, Candidatus Kapabacteria, Cyanobacteriota and Verrucomicrobiota (Figure 9), all of which are relatively common members of freshwater communities [14].

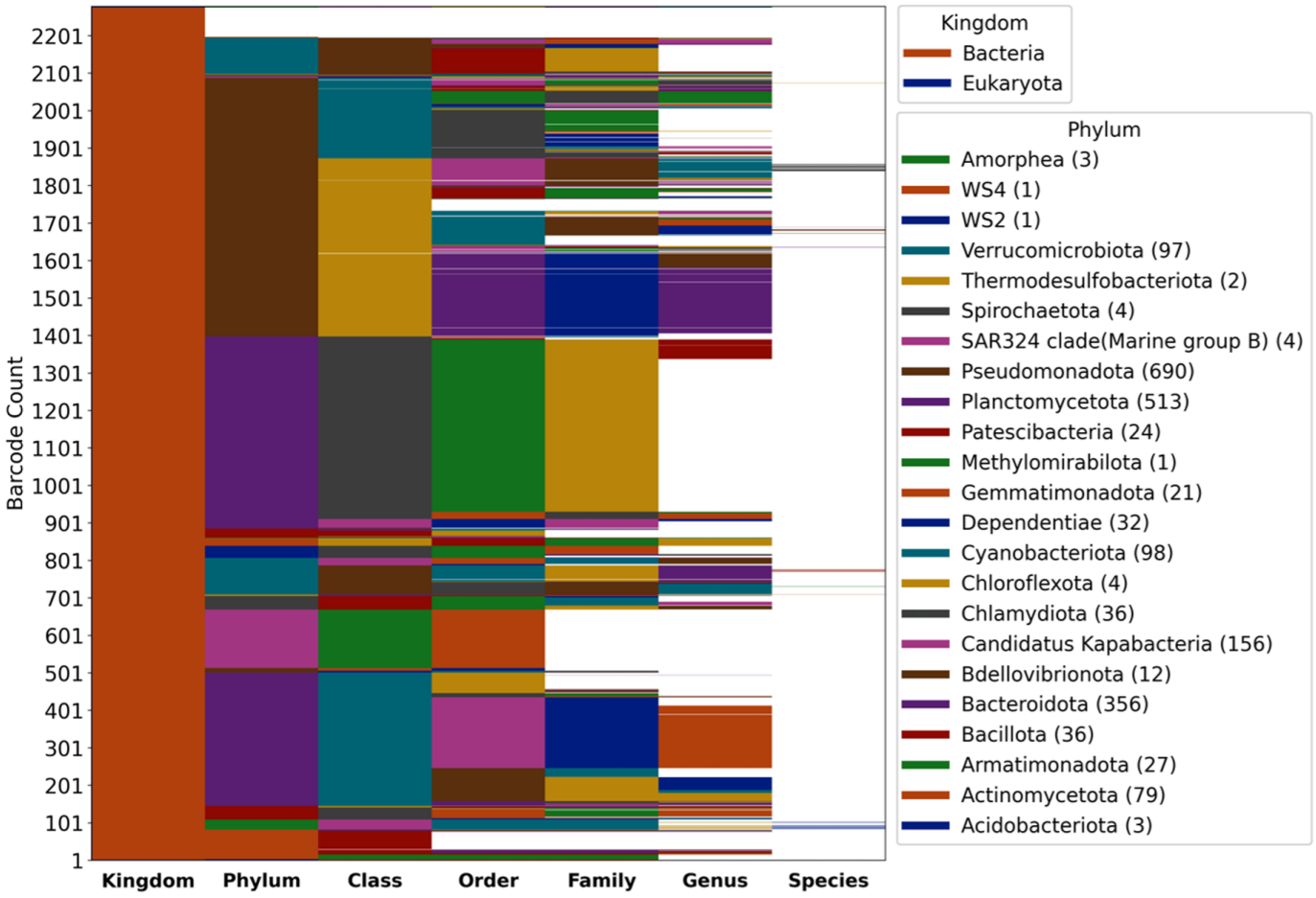

Figure 9. Abundance chart of identified taxons in the pond water sample. Barcodes with no assigned phylum signify mixed species cases.
Note: Observable Eukaryota consists of three fungal members of Amorphea phylum.

Among the three sample sources, soil samples exhibited the highest variability in abundance, despite having the lowest barcode count and only a few barcodes assigned at the species level. These samples had lower taxonomic depth, with most taxa identified at the genus, family and order level. The dominant phylum was Pseudomonadota, followed by significantly less predominant phyla such as Actinomycetota, Acidobacteriota, Planctomycetota, and Bacillota (Figure 10), all of which are common in garden soil environments [15].

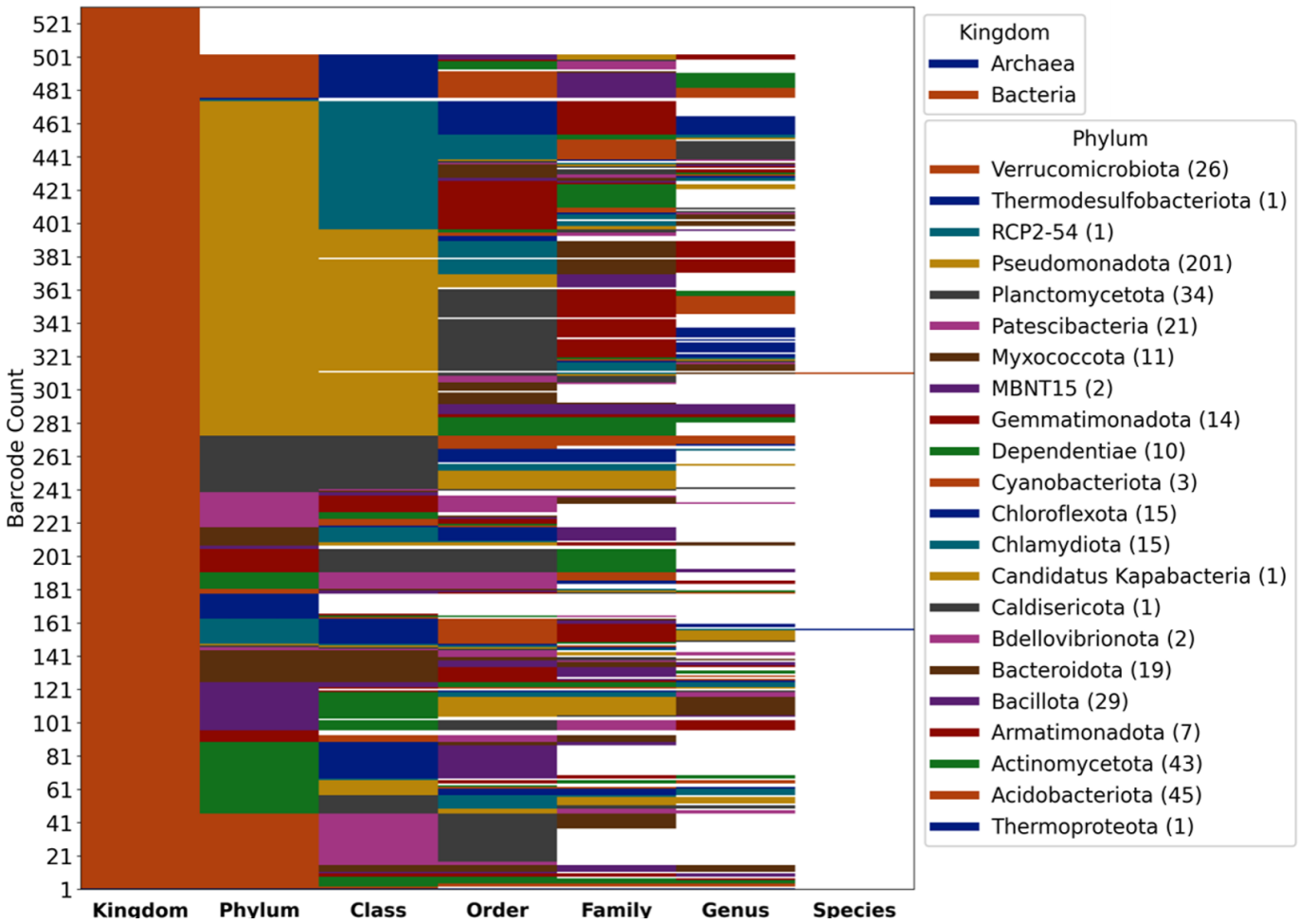

Figure 10. Abundance chart of identified taxons in soil sample. Barcodes with no assigned phylum signify mixed species cases.

Overall, the sequencing analysis of environmental samples aligns with the expected microbial composition at the phylum level, indicating accurate taxonomic assignment. Furthermore, the detection of uncultivable organisms, such as Patescibacteria, demonstrates the capability of the SPC technology and the Single Microbe DNA Barcoding kit in facilitating the study of organisms that are difficult or impossible to cultivate at both the scale of metagenomics and the resolution of individual genomes.

CONCLUSIONS

1. Experimental results show that our Single Microbe DNA Barcoding kit is fully compatible with biotinylated probe-based targeted enrichment techniques.
2. The observed enrichment of a mixed *E. coli* and *B. subtilis* genome library shows an increase in 16S reads from 0.174% to 92%, which can be expressed as over 500–fold enrichment.
3. Enrichment of environmental sample libraries showed variable but, in all cases, positive results, with 16S reads increasing from 0.26–1% to 61–78%, which can be expressed as 70– to 300–fold enrichment.
4. Targeted enrichment enables deeper sequencing on specific targets while narrowing down the scale and allowing higher resolution on more elusive sequences of interest.

REFERENCES

1. Matuła, K., Rivello, F., Huck, W. (2019). Single-Cell Analysis Using Droplet Microfluidics. Advanced Biosystems, 2020, 4, 1900188. https://doi.org/10.1002/adbi.201900188
2. Kim SC., Clark IC., Shahi P., Abate AR. (2018) Single-Cell RT-PCR in Microfluidic Droplets with Integrated Chemical Lysis. Analytical Chemistry, 2018 Jan 16;90(2):1273–1279. https://doi.org/10.1021/acs.analchem.7b04050
3. Leonaviciene, G., Leonavicius, K., Meskys, R., Mazutis, L. (2020). Multi-step processing of single cells using semi-permeable capsules. Lab on a Chip, 2020 (21). https://doi.org/10.1039/d0lc00660b
4. Leonaviciene, G., & Mazutis, L. (2023). RNA cytometry of single-cells using semi-permeable microcapsules. Nucleic Acids Research, 51(1), e2. https://doi.org/10.1093/nar/gkac918
5. Atrandi Biosciences 2024, Single-Cell Metagenomics, accessed 7 October 2024, URL: https://atrandi.com/application/single-cell-metagenomics/technology
6. Murchie, TJ., Karpinski, E., Eaton, K., Duggan, AT., Baleka, S., Zazula, G., MacPhee, RDE., Froese, D., Poinar, HN. (2022). Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost sediments. Current Biology 2022 32(4), 851. https://doi.org/10.1016/j.cub.2021.12.023
7. Beaudry, MS., Wang, J., Kieran, T., Thomas, J., Bayona-Vasquez, NJ., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, JS., Rhodes, OE., Glenn, TC. (2020). Improved microbial community characterization of 16S rRNA via metagenome hybridization capture enrichment. Frontiers in Microbiology, 2021, 12, https://doi.org/10.3389/fmicb.2021.644662
8. Beaudry, MS., Thomas, JC., Baptista, RP., Sullivan, AH., Norfolk, W., Devault, A., Enk, J., Kieran, TJ., Rhodes OE., Perry-Dow, A., Rose, LJ., Bayona-Vásquez, NJ., Oladeinde, A., Lipp, EK., Sanchez, S., Glenn, TC. (2021). Escaping the fate of Sisyphus: assessing resistome hybridization baits for antimicrobial resistance gene capture. Environmental Microbiology, 2021, 23(12), 7523. https://doi.org/10.1111/1462–2920.15767
9. Kozarewa, I., Armisen, J., Gardner, AF., Slatko BE., Hendrickson, CL. (2015). Overview of Target Enrichment Strategies. Current Protocols Molecular Biology, 2015 Oct 1;112:7.21.1–7.21.23. https://doi.org/10.1002/0471142727.mb0721s112
10. Gasc, C., Peyretaillade, E., Peyret P. (2016). Sequence capture by hybridization to explore modern and ancient genomic diversity in model and nonmodel organisms. Nucleic Acids Research, 2016, 44(10). https://doi.org/10.1093/nar/gkw309
11. Singh, RR. (2022). Target Enrichment Approaches for Next-Generation Sequencing Applications in Oncology. Diagnostics 2022, 12(7), 1539. https://doi.org/10.3390/diagnostics12071539
12. Beaudry, MS., Wang, J., Kieran, TJ., Thomas, J., Bayona-Vásquez, NJ., Gao, B., Devault, A., Brunelle, B., Lu, K., Wang, JS., Rhodes OE., Glenn, TC. (2021). Improved Microbial Community Characterization of 16S rRNA via Metagenome Hybridization Capture Enrichment. Frontiers in Microbiology, (2021), 12. https://doi.org/10.3389/fmicb.2021.644662
13. Bang, E., Oh, S., Ju, U., Chang, HE., Hong, JS., Baek, HJ., Kim, KS., Lee, HJ., Park, KU. (2023). Factors influencing oral microbiome analysis: from saliva sampling methods to next-generation sequencing platforms. Scientific Reports, 2023, 13, 10086. https://doi.org/10.1038/s41598–023–37246–2
14. Qin, Y., Hou, J., Deng, M., Liu, Q., Wu, C., Ji, Y., He, X. (2016). Bacterial abundance and diversity in pond water supplied with different feeds. Scientific Reports, 2016, 6, 35232. https://doi.org/10.1038/srep35232
15. Mhete, M., Eze, PN., Rahube, TO., Akinyemi, FO. (2020). Soil properties influence bacterial abundance and diversity under different land-use regimes in semi-arid environments. Scientific African, 2020, 7. https://doi.org/10.1016/j.sciaf.2019.e00246

SUPPLEMENTARY MATERIAL

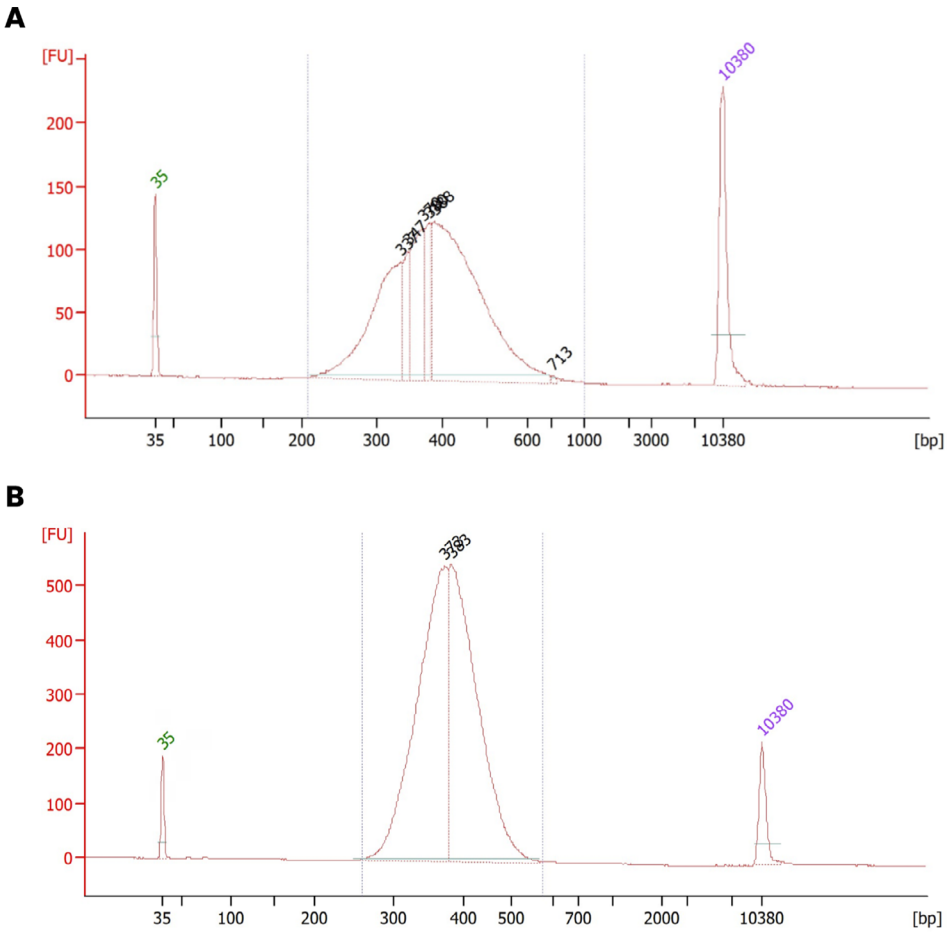

Figure S1. Results of Agilent Bioanalyzer run with pre-enriched (A) and post-enrichment (B) libraries. Here FU - Fluorescent Units, bp – length of DNA fragment.

Aside from using minimap2, we initially employed a conserved 16S sequence search to evaluate enrichment quality. This approach involved designing a set of 14 short, conserved 16S sequences distributed across the full length of the 16S gene (**Table S1**). Using these sequences, we performed an in-string search, requiring at least a 15–nucleotide perfect match, without gaps, to the paired read being analyzed. While the final results closely resembled those from minimap2 in enriched libraries, they varied by a factor of 2 to 10 when examining pre-enriched libraries (**Table S2**).

Table S1. Custom sequences targeting conserved 16S bacterial DNA regions. Y - C or T, M - A or C, S - G or C, N – any base, R - A or G, H - A or C or T, V - A or C or G, K - G or T, W - A or T.

Name of 16S conserved region	Sequence (5'-3')
16S_8_27	AGAGTTTGATYMTGGCTCAG
16S_100_119	ASYGGCGNACGGGTGAGTAA
16S_337_357	ACTCCTACGGRNGGCNGCAGT
16S_504_521	GGCTAACTHCGTGNCVGC
16S_682_704	GTGTAGMGGTGAAATKCGTAGAT
16S_784_807	GGATTAGAWACCCNNGTAGTCCAC
16S_917_938	AATWGRCGGGGRCCCGCACAAG
16S_966_984	CAACGCGARGAACCTTACC
16S_1046_1079	GGTGNTGCATGGYYGYCGTCAGCTCGTGYYGTGA
16S_1082_1114	TGTTGGGTTAAGTCCCRYAACGAGCGCAACCCT
16S_1176_1192	GGAAGGYGGGGAYGACG
16S_1219_1236	GGGCKACACACGYGCTAC
16S_1386_1406	GCCTTGYACWCWCCGCCCGTC
16S_1486_1509	GGGTGAAGTCRTAACAAGGTANCC

Table S2. Library 16S composition evaluation based on conserved 16S sequence search.

Library name	**EC-BS-1**	**ENV1**	**ENV2**	**ENV3**	**ENV4**
Number of paired reads	74 505 789	13 445 701	28 344 205	14 456 643	26 148 244
16S paired reads	66 360 975	7 456 309	16 615 692	9 083 398	18 878 661
% of paired reads being 16S	**89.07**	**55.45**	**58.62**	**62.83**	**72.20**
Number of paired reads in pre-pulldown library	1 000 000	152 125	132 508	68 677	185 654
16S paired reads in pre-pulldown library	612	116	85	71	128
% of 16S in pre-pulldown library	**0.0612**	**0.0762**	**0.0641**	**0.1034**	**0.0689**
Enrichment ratio	1 455.36	727.25	913.86	607.76	1 047.18

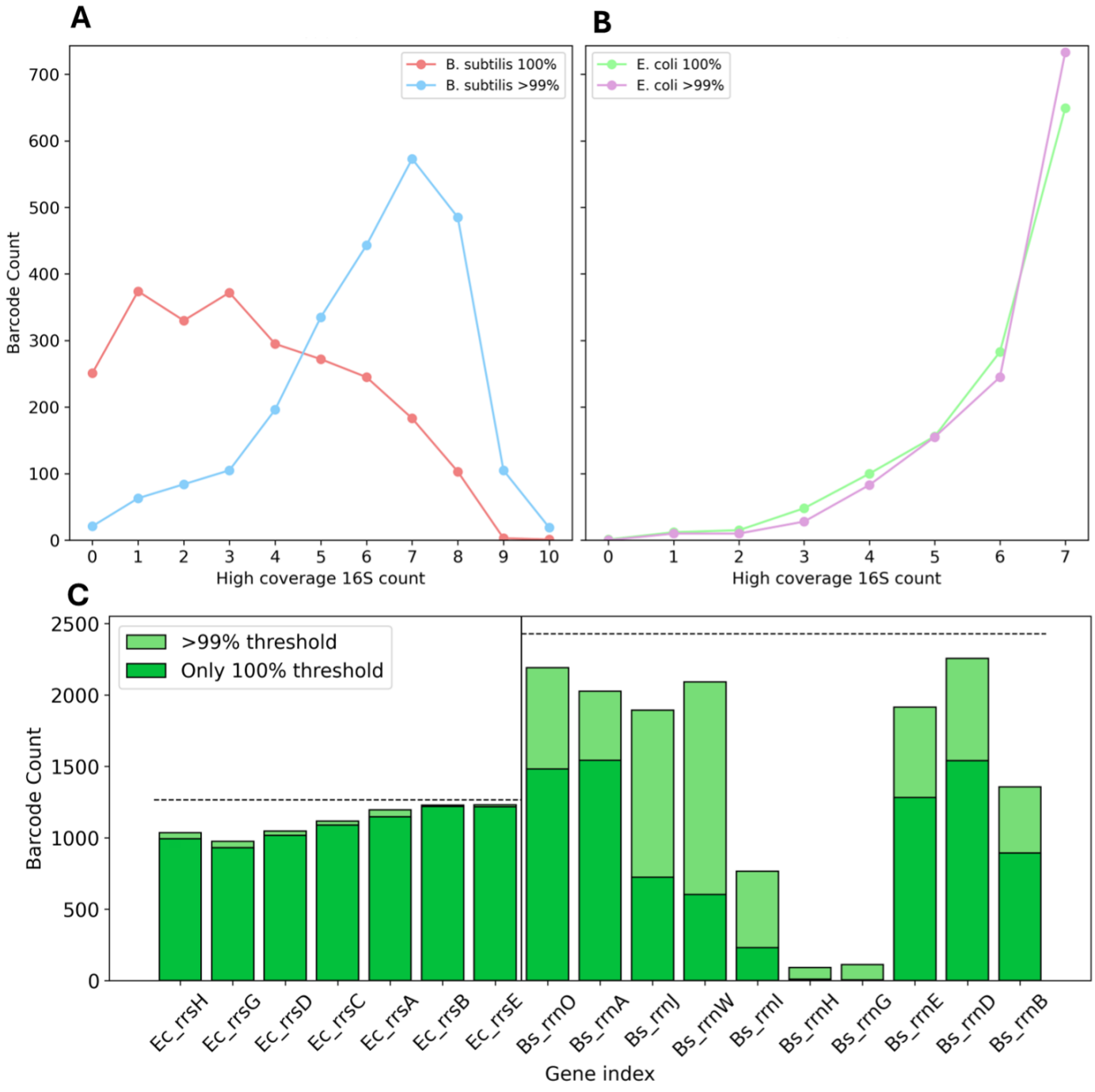

Figure S2. Alternative approaches to visualize 16S coverage quality. A and B – distribution of barcodes by number of 16S sequences detected. The maximum number of 16S sequences (x-axis) is 10 and 7 for *B. subtilis* and *E. coli*, respectively. Data are shown for two coverage breadth thresholds for a sequence to be considered detected. C – individual threshold fitting 16S sequence count from the whole dataset; the horizontal dashed lines mark the total number of barcodes used in the analysis for each species (n(*E. coli*) = 1264, n(*B. subtilis*) = 2429).

Atrandi
BIOSCIENCES

APPLICATION NOTE

ENABLING
THE CENTURY OF BIOLOGY

ATRANDI.COM