Linking plasmid-based beta-lactamases to their bacterial hosts using single-cell fusion PCR

Peter J. Diebold[1], Felicia N. New[1], Michael Hovan[2], Michael J. Satlin[3], Ilana L. Brito[1]*

[1] Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY

[2] Robert Wood Johnson Medical School, New Brunswick, NJ

[3] Weill Cornell Medicine, Cornell University, New York, NY

* Please send all correspondences to: ibrito@cornell.edu

Abstract

The horizonal transfer of plasmid-encoded genes allows bacteria to adapt to constantly shifting environmental pressures, bestowing functional advantages to their bacterial hosts such as antibiotic resistance, metal resistance, virulence factors, and polysaccharide utilization. However, common molecular methods such as short- and long-read sequencing of microbiomes cannot associate extrachromosomal plasmids with the genome of the host bacterium. Alternative methods to link plasmids to host bacteria are either laborious, expensive or prone to contamination. Here we present the One-step Isolation and Lysis PCR (OIL-PCR) method, which molecularly links target ARGs with the bacterial 16S rRNA gene via fusion PCR performed within an emulsion. After validating this method, we apply it to identify the bacterial hosts of three clinically relevant beta-lactamases in a neutropenic patient population who are particularly vulnerable multidrug-resistant infections. We detect novel associations of two low-abundance genera, *Romboutsia* and *Agathobacter*, with a multi-drug resistant plasmid harbored by *Klebsiella pneumoniae*. We put forth a robust, accessible, and high-throughput platform for sensitively surveying the bacterial hosts of mobile genes in complex microbial communities.

Introduction

The emergence of multidrug-resistant (MDR) pathogens is a grave public health threat that occurs when pathogenic bacteria acquire antibiotic resistant genes (ARGs) through horizontal gene transfer (HGT) with bacteria in their proximal environment. The gut microbiome harbors a diverse repertoire of ARGs and these genes have been proposed to serve as a reservoir for HGT with MDR pathogens[1]. ARGs are often carried on mobilizable plasmids that impose technical challenges to surveying the set of bacteria affiliated with these genes. Standard molecular tools such as PCR and next-generation sequencing often fail to associate mobile ARGs with their bacterial hosts because they cannot capture the cellular context of extrachromosomal genes in the case of plasmids. Novel untargeted sequencing methods, such as bacterial Hi-C[2] and methylation profiling[3], provide broad reconstruction of plasmid-host relationships in metagenomes, as a trade-off for sensitivity. Alternatively, single-cell whole genome sequencing offers an ideal solution to this problem, but may be lower throughput, more expensive and require specialized equipment[4,5]. Targeted methods, such as bacterial cell culture under antibiotic selection, require that the ARG is expressed, functional, and selective in all hosts. Applying this broadly to capture the full diversity of the gut microbiome is complicated by the need for wide-ranging media and growth conditions[6,7].

Single-cell qPCR is a targeted method to identify the hosts of specific genes, however each use specialized microfluidic devices, are limited in bacterial taxa they can capture, and most do not allow direct sequencing of the PCR products[8–10]. Alternatively, epicPCR[11], uses fusion PCR and two emulsion steps to associate a taxonomic marker with a functional gene. Sequencing the fused PCR products provides accurate and sensitive associations between 16S sequence taxonomy and a given target gene. However, this method can be challenging to execute, difficult to scale up for multiple samples, and utilizes toxic and difficult-to-acquire reagents.

Here, we put forth One-step Isolation and Lysis PCR (OIL-PCR), a method that detects host-ARG associations from complex microbial communities through cellular emulsion and fusion PCR. Our streamlined method, based on the innovation of epicPCR, simplifies the procedure by combining the two

emulsion steps of cell lysis and fusion PCR into a single emulsion PCR reaction that can be performed in a 96-well format using robotic automation. Furthermore, OIL-PCR can be multiplexed to target at least three genes in the same reaction, uses non-toxic commercially available reagents, and can be performed without relying on microfluidics or specialized equipment. Validation experiments on three environmental bacterial communities reveal that OIL-PCR is highly accurate and specific. We demonstrate the utility of this approach in examining the novel association of three extended spectrum beta-lactamase (ESBL) genes with two commensal organisms in the gut. Our results highlight the utility of this method in defining mobile ARG distribution within microbial communities as complex as the human gut microbiome.

Results

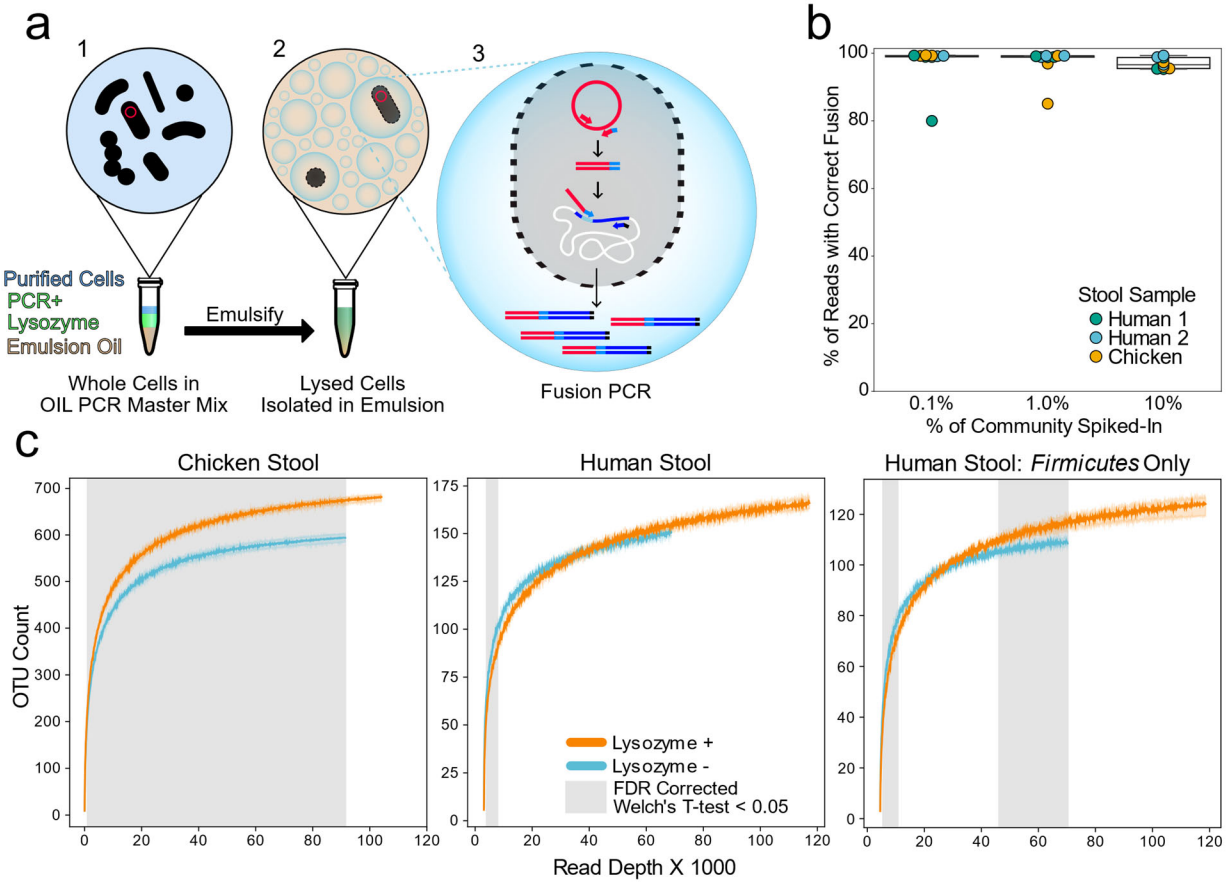
Development of a One-step Isolation and Lysis PCR method

OIL-PCR applies established fusion PCR methods to fuse any gene of interest to the 16S rRNA gene using three primers: two primers hybridize to the target gene, and a universal 16S reverse primer hybridizes to the V4 region. Amplification of the target gene appends a universal 16S forward primer sequence to the end of the target amplicon via a tailed reverse primer. The target gene amplicon then acts as a primer for amplification and hybridizes to the 16S rRNA gene as a forward primer, producing a fused gene product containing both the target gene and the 16S V4 sequence (Fig. 1a, Supplementary Fig. 1).

For fusion PCR to accurately link target genes with host marker genes, cells must be isolated to prevent the formation of non-specific fusion products. Oil emulsions and microwells have long been used to isolate eukaryotic cells, however, it is difficult to lyse bacteria in this format, especially gram-positive bacteria due to their thick cell walls. Existing single-cell isolation methods for bacteria either do not address this problem[9,10], rely on specialized microfluidics[12] or use time-consuming methods to encapsulate bacteria within hydrogel beads before performing multi-step chemical and enzymatic lysis

71 procedures[11,13]. To address this problem, OIL-PCR combines bacterial isolation, lysis, and fusion PCR
72 into a single streamlined reaction.

73



74 **Figure 1. OIL-PCR can specifically link plasmid-encoded genes with their hosts**

75 a) Depiction of the OIL-PCR method. (1) Nycodenz-purified cells are mixed with PCR master mix,
76 lysozyme, and emulsion oil and shaken to create an emulsion. (2) Cells are lysed within the
77 emulsion. (3) Fusion PCR is performed in droplets containing cells harboring the targeted gene.
78 Fused amplicons between the gene of interest and the 16S rRNA gene are the product.
79 b) A boxplot showing the percent of Illumina reads containing correct fusion products, namely the
80 fusion of plasmid-borne *cmR* and the 16S rRNA gene of *E. coli* MG1655. OIL-PCR was

performed on two individuals' and one chicken's gut microbiome sample, spiked with varying concentrations of *E. coli.*

c) Rarefaction analysis of chicken (left) or human gut microbiome sample (middle) with (orange) and without (blue) lysozyme treatment. At right is the rarefaction analysis performed on Firmicutes only in the human stool sample. Grayed regions in the plot represent areas where the curves are significantly different ($p<0.05$) from one another, according to an FDR-corrected Welch's t-test.

We developed a protocol that allows for the incorporation of Ready-Lyse (RL) Lysozyme into the fusion PCR master mix. Whole bacterial cells are added directly to the master mix while on ice to inhibit lytic activity during sample preparation. Vigorous shaking of the mixture then encapsulates the individual cells in an emulsion. Warming the emulsion to 37 °C activates the enzyme, lysing the cells. Next, a standard PCR thermocycler carries out the fusion PCR reactions in the single-cell emulsions. Fused PCR products are purified from the emulsion and amplified further with a nested primer to filter out off-target PCR products and add Illumina adapters. Lastly, custom indexing primers are used to index the fused products before Illumina sequencing. Our experiments confirmed the compatibility of the RL Lysozyme with the fusion PCR reaction, but required the addition of bovine serum albumin, a globular protein known to reduce protein aggregation[14] (Supplementary Fig. 2a). We found that RL retained full activity in the standard NEB Phusion HF buffer (Supplementary Fig. 2c).

Next, we optimized the fusion PCR master mix to maintain a stable emulsion and amplify efficiently in picoliter droplets. PCR emulsions were prepared with fluorinated oil as used in modern emulsion-based methods, such as Drop-Seq[15]. We combined the fusion PCR master mix with bacterial cells and emulsion oil in either a 1.5 ml tube or a 0.5 ml deep-well plate before emulsifying the mixture using a tabletop bead

105 homogenizer. Unlike microfluidic-enabled emulsions, our protocol leverages equipment commonly found
106 in most molecular biology laboratories. We stabilized the emulsion by using detergent-free buffers and
107 improved the efficiency of the PCR amplification within the emulsion by adding additional polymerase,
108 BSA, dithiothreitol (DTT), and ammonium sulfate. We found that the addition of extra $MgCl_2$ mitigated
109 the inhibitory effects of extremely high concentrations of cell debris within droplets after lysis
110 (Supplementary Fig. 1b).

111 **OIL-PCR accurately associates plasmid genes with the host in a binary community**

112 In any emulsion-based method, it is essential to optimize the concentration of input cells to prevent the
113 encapsulation of two or more cells in the same droplet. When using a monodisperse emulsion such as
114 those generated using microfluidics, the ideal concentration of input cells is chosen using a Poisson
115 distribution[8–10]. However, these calculations are not reliable in the case of a polydisperse emulsion, as
116 employed here. We therefore developed a probe-based TaqMan qPCR assay to experimentally verify the
117 optimal concentration of input cells that prevented non-specific gene fusions (Supplementary Fig. 3a).
118 OIL-PCR was performed on a binary community consisting of *E. coli* carrying the chloramphenicol
119 resistance gene *cmR* on a plasmid and WT *V. cholerae*. The two strains were mixed 1:1 and we performed
120 OIL-PCR with a fusion primer set specific to *cmR* and universally targeting the 16S rRNA gene[11]
121 (Supplementary Table T2 and T3). A gradient of cell input concentrations was used and the final PCR
122 products were recovered and purified. We then performed probe-based qPCR on the purified product
123 using a nested primer for *cmR*, two blocking primers to inactivate any unfused amplicons, and two
124 distinct fluorescent TaqMan probes (Thermo-Fisher 4316034) to specifically target the V4 region of
125 either *E. coli* or *V. cholerae* (Supplementary Table T2). The fluorescent signal from each probe measured
126 the relative ratio of specific to non-specific gene fusions present in the final amplicon pool
127 (Supplementary Fig. 3a). When the input concentration of cells was at or lower than 400 cells/µl, or 40 k
128 cells per reaction, non-specific gene fusion detection was reduced to undetectable levels (Supplementary
129 Fig. 3b). As well as confirming that bacterial cells were isolated within the emulsion, we further

confirmed that droplets did not coalesce by performing the TaqMan assay on OIL-PCR products from *E. coli* and *V. cholerae* cells combined after they were individually emulsified (Supplementary Fig. 3b). Our results confirmed that the emulsion is highly stable and coalescence was undetected.

Application of OIL-PCR to environmental microbial communities allows robust and sensitive association of extrachromosomal elements with their host.

Using OIL-PCR on environmental microbial communities requires clean bacterial cell preparations free of environmental contaminants which may inhibit PCR. To address this concern, cells were purified using Nycodenz density gradient centrifugation[16,17], a simple method that can isolate clean bacterial fractions with minimal handling time to reduce contamination. Additionally, concerned that cell-free DNA can stick to the membranes and cell walls of bacteria[18] thus introducing noisy associations in the data, we treated cells with heat-liable double strand-specific DNase (dsDNase). This enzyme only digests unprotected double stranded genomic DNA present in the samples without degrading single strand primers. By controlling the enzyme concentration, temperature, and speed at which cells were processed, we were able to digest extra-cellular DNA without impacting PCR efficiency of cellular contents. Using our Taqman assay, we demonstrated that including dsDNase treatment has the potential to increase the total cell input per reaction tenfold (Supplementary Fig. 3d).

To test the accuracy of our method on environmental samples, we spiked *Escherichia coli MG1655*[19] containing plasmid pBAD33[20] harboring *cmR* into two human and one chicken stool samples that lacked the gene according to PCR screening. Our results show that when *E. coli* was incorporated at 0.1%, or about 20 cells total, 97.8% of the reads (or 99.2%, excluding a single outlier) demonstrated the correct association when the test strain of *E. coli* was incorporated at 0.1%, or about 20 cells total, highlighting the sensitivity of OIL-PCR to detect the associations of genes in low abundant species across different sample types. The accuracy of OIL-PCR decreases slightly when the targeted sequence increases to 10% of the community composition, although associations were still 97% correct on average.

Lysozyme improves capture of difficult-to-lyse gram positive bacteria

To achieve our goal of robust lysis and amplification to screen all bacteria within a complex community, we measured the effect lysozyme had on bacterial detection. We performed standard 16S sequencing on human and chicken stool communities using OIL-PCR, testing three variables: the effect of lysozyme, dsDNase, and heat inactivation of dsDNase on total bacterial recovery (Supplementary Fig. 4). All eight combinations of the three variables were tested in duplicate for two stool samples using robotic automation. For our analysis, we chose to focus on the total number of operational taxonomic units (OTUs) captured in our data rather than relative abundance metrics, as this better reflects our goal of detecting species, rather than recapitulating the starting community structure.

First, we assayed how each of the three variables (RL, dsDNase, and heat inactivation) affected OTU recovery. Based on rarefaction curves, we found dsDNase and Heat inactivation had no significant effect on OTU recovery in human and chicken stool, while RL lysozyme significantly increase OTU recovery in chicken stool based on Welch's T-test with Benjamini-Hochberg FDR correction (Fig 1, Supplementary Fig. 4). RL was the only variable that significantly changed OTU recovery and therefore it was the only variable included in our final protocol.

Next, we looked to see whether any taxonomic groups were being enriched or depleted with lysozyme. Technical replicate OTU tables were combined to allow for deeper sampling depth. Results show that no phylum was significantly depleted, and importantly, at higher sequencing depths, *Firmicutes* were enriched in both the human and chicken samples. Additionally, *Bacteroides* species were enriched in the chicken sample. These results demonstrate the benefit of RL Lysozyme for capturing difficult to lyse gram-positive bacteria in the Firmicutes phylum, which account for much of the commensal diversity within the gut microbiome

We noticed that the total number of OTUs recovered from OIL-PCR was significantly lower than 16S sequencing of the input community at the same sampling depth (Supplementary Fig. 5). We hypothesized

the reason for this dramatic reduction in OTUs was due to subsampling bias introduced through low cell input and variable amplification efficiency in OIL-PCR. To test our hypothesis, we combined OTU tables from two, four and eight technical replicates and found a consistent up-shift for each rarefaction curve as we combined more tables. This up-shift was not observed when combining the input Nycodenz sequencing, indicating that the reduced OTU counts were due in part to sub-sampling bias and not an inherent failure to capture bacterial taxa (Supplemental Fig. 5). We therefore recommend OIL-PCR to be performed in replicates such that a sufficient number of cells are being sampled.

Increased throughput through automation and multiplexing

To further improve the efficiency and throughput of OIL-PCR, we sought to transition the method from 1.5 ml centrifuge tubes to a 96-well plate format using the Eppendorf epMotion liquid handling robot. The liquid handling robot can perform certain parts of the PCR preparation as well as DNA recovery and purification. The automated workflow allowed us to process up to 48 samples simultaneously with fewer manual steps overall.

We next tested whether OIL-PCR could simultaneously target multiple genes though multiplexing. We repeated the previously described TaqMan assay using a strain of *V. cholerae* containing the ampicillin resistance gene *ampR* and *E. coli* with *cmR;* both on a plasmid (Supplementary Fig. 3c). Our results demonstrate that OIL-PCR can be multiplexed while still accurately maintaining the correct associations of target genes with their host bacteria.

Bacterial hosts are identified for several clinically important ß-lactamase genes

We analyzed metagenomic sequencing of stool samples that were collected from a cohort of patients who were neutropenic because of chemotherapy administered for a hematopoietic cell transplant. Two patients, B335 and B314, were chosen for OIL-PCR based on the presence of three class-A beta-lactamase genes, bla_{TEM}, bla_{SHV}, and $bla_{CTX\text{-}M}$ in the metagenomes. We tested a three-sample time course from patient B335: before antibiotic treatment, after four days of trimethoprim-sulfamethoxazole and one

day of levofloxin, and lastly after an additional two days of levofloxin (Fig. 2a). Patient B335 carried all three genes across three time points with *bla*$_{TEM}$ and *bla*$_{CTX-M}$ on an 80 kb *Klebsiella* plasmid and *bla*$_{SHV}$ on a contig that was annotated as *Klebsiella* (Fig. 2b). We tested one sample from patient B314 from before antibiotic treatment which carried multiple *bla*$_{SHV}$ genes(Supplementary Fig. 6). We hypothesized that OIL-PCR could be used to sensitively and accurately detect additional hosts of these genes.

We designed three degenerate fusion primer sets to broadly target most variants of *bla*$_{TEM}$, *bla*$_{SHV}$, and *bla*$_{CTX-M}$ (Supplementary Table T2, T3), and performed multiplexed OIL-PCR with robotic automation. Samples were processed in quadruplicates along with negative and positive controls. We set a threshold for defining positive gene-taxa associations, as having 0.5% of total reads across the four replicates.

Our OIL-PCR results largely confirm findings in the metagenomic assemblies from Kent et al.[2]. In B314, we found *bla*$_{SHV}$ associated with *Klebsiella* as suggested by metagenomic assemblies. However, we also detected two other class-A beta-lactamase genes, *bla*$_{LEN}$ and *bla*$_{OXY}$, which were present in the metagenomes but we did not expect to amplify with our primers. *bla*$_{LEN}$ amplified with the primers designed for *bla*$_{SHV}$ and *bla*$_{OXY}$ amplified with primers for *bla*$_{CTX-M}$. Curiously, *bla*$_{OXY}$ is an exceptionally poor match for our *bla*$_{CTX-M}$ primers, having a mismatch one base away from the 5' end of the fusion primer. We hypothesize that the low annealing temperature and modified buffer used in the emulsion PCR is highly permissive to priming mismatches. We see permissive annealing as an advantage for the method because it allows for amplification of unknown variants of target genes while amplification due to off-target priming is filtered out during the nested PCR step (Supplementary Fig. 1), leaving only the true amplicons in the final sequencing. This permissive annealing behavior of OIL-PCR can be leveraged in the future to design broad-range primers for diverse gene groups such as metallo-beta-lactamases[24].

Results from patient B335's time course also matched the metagenomic sequencing from Kent et al., associating *bla*$_{TEM}$, *bla*$_{SHV}$, and *bla*$_{CTX-M}$ with *Klebsiella* in all three time points (Fig. 2c,d). We also found that all three genes strongly associated with the commensal genus *Romboutsia* in time points T2 and T3 and to a lesser extent with *Agathobacter* in time point T1 (Fig. 2c,d). A strain of *Escherichia* with a

distinct variant of *bla*$_{TEM}$ was detected at time point T2, but did not pass the detection threshold across all replicates in timepoint T3. We repeated OIL-PCR on all three samples from B335, this time in triplicate without multiplexing to further confirm these results. The singleplex experiment perfectly mirrored the multiplex results, excluding one replicate of T2/CTX-M which failed to sequence, indicating that these genes are linked with organisms other than *Klebsiella*. As further confirmation of this result, we targeted two Tn3-like transposon genes situated in close proximity to *bla*$_{TEM}$ and *bla*$_{CTX\text{-}M}$ on the 80 kb *Klebsiella* plasmid. We hypothesized that these genes should also be associated with the same genera as the ARGs. Remarkably we observed the identical pattern with *Klebsiella, Romboutsia,* and *Agathobacter* as with the three beta-lactamases, but not *Escherichia,* which carried a distinct variant of *bla*$_{TEM}$ (Fig 2c,d).

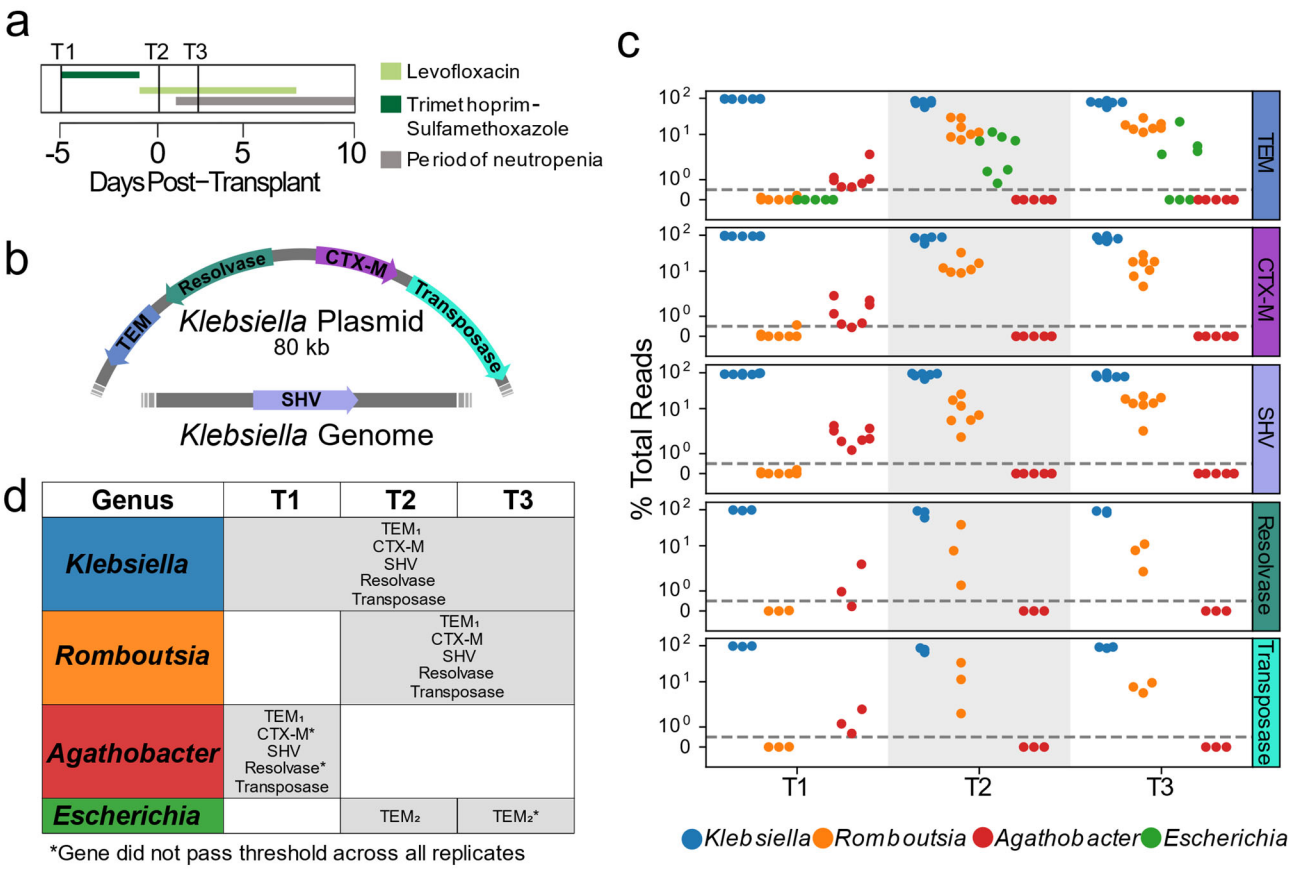


Figure 2. Extended spectrum beta-lactamase genes are associated with both pathogenic and commensal species

a) Summary of treatment and sample time points for patient B335

b) Depiction of an 80kb plasmid carried by *K. pneumoniae* harboring the $bla_{CTX\text{-}M}$, bla_{TEM}, Tn3 transposase and resolvase genes. The bla_{SHV} gene is presumed to be carried within the *K. pneumoniae* genome. Placement of these genes was inferred from metagenomic assemblies of patient B335's gut microbiome sample.

c) OIL-PCR results for each of the genes depicted in (a) patient B335 at 3 time points. For all gene-taxa associations, the percent of total OIL-PCR reads for that gene-time point is plotted. All species passing our detection threshold of 0.5% (dotted line) at any of the three time points is included in this plot.

d) A table summarizing the results in (b). All gene-taxa associations for each time point passing our detection thresholds are listed. Two SNP variants of TEM were detected and denoted with subscript numbering. Gene-taxa associations which did not consistently pass our detection threshold across all replicates are noted (*).

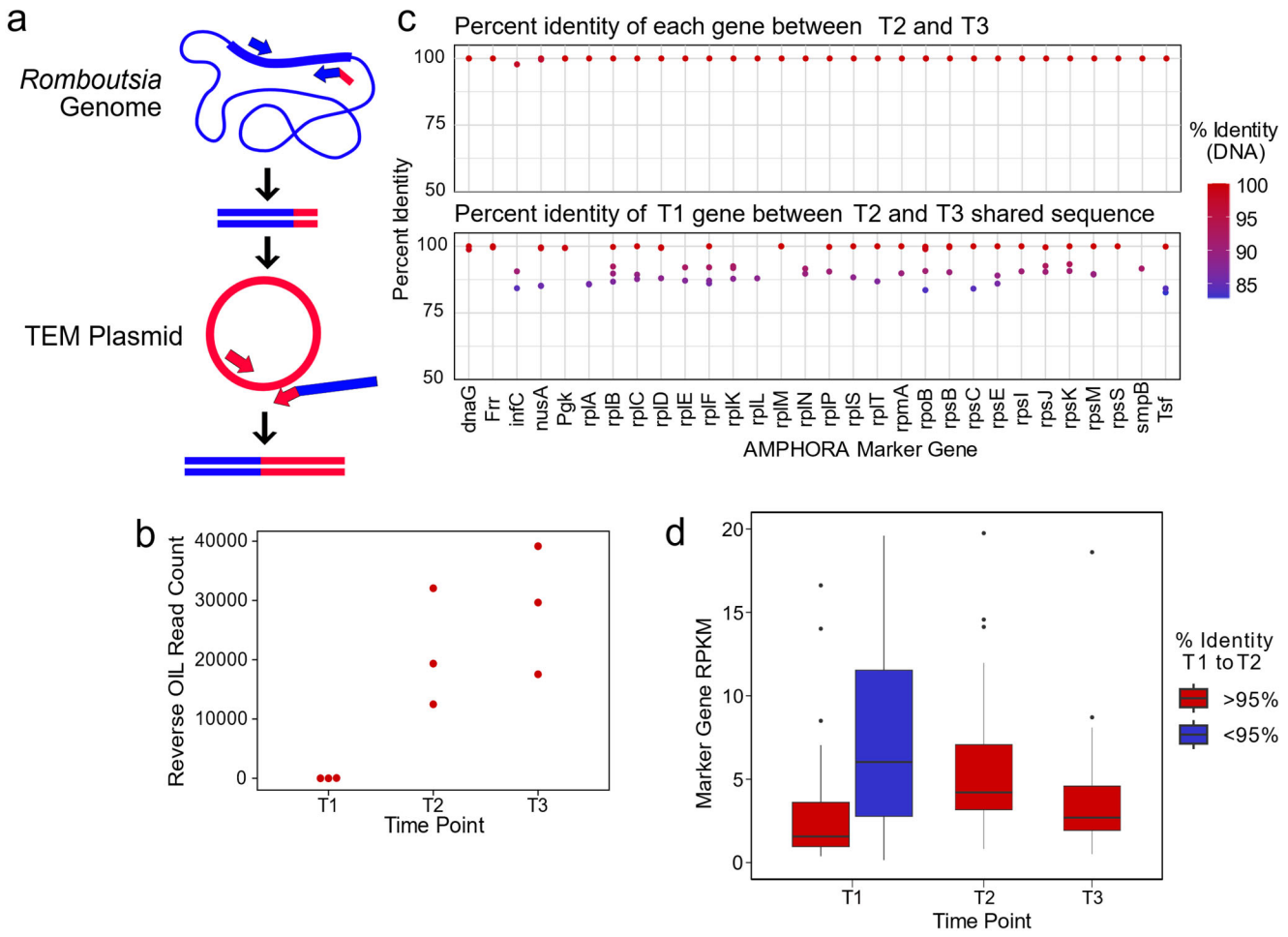
OIL-PCR provides further evidence of the association of beta-lactamases with the commensal *Romboutsia*

We next investigated whether OIL-PCR could be used to further confirm the association between *Romboutsia* and the three beta-lactamases. We focused specifically on *Romboutsia* because of the strong signal in the OIL-PCR results compared to *Agathobacter.* For this experiment, instead of fusing the ARG sequence to the 16S rRNA gene using universal primers, we used primers designed to specifically detect the *Romboutsia* 16S rRNA[25] and fused the 16S gene specifically to bla_{TEM} (Fig. 3a, Supplementary Table T2, T3). In this instance, no amplification is possible unless *Romboutsia* is encased in the same droplet with the bla_{TEM} gene and would negate the possibility of false-positive associations due to chimera formation. Results show amplification and sequencing was only produced from time points T2 and T3

263 with no signal detected at time T1, confirming the presence of *bla*$_{TEM}$ within *Romboutsia* at time points T2

264 and T3 but not T1 (Fig. 3b).

265



266 **Figure 3. *R. timonensis* strains associated with the three beta-lactamase genes appear over the**

267 **patient's time course**

268 a) Depiction of the reverse OIL-PCR reaction in which *Romboutsia*-specific 16S rRNA sequences

269 (blue) are fused with the *bla*$_{TEM}$ sequences (red).

270 b) OIL-PCR read counts of the reaction shown in (a) are plotted.

c) The percent sequence identity of assembled *R. timonensis* marker genes between genes identified in timepoints 2 and 3 (top) and between timepoints 1 and between sequences shared at timepoints 2 and 3 (bottom).

d) RPKM-normalized abundance-values for the assembled marker genes for each strain assembled in time point 1 and the major strain present in timepoints 2 and 3.

We next explored the metagenomic data for clues as to whether the *Romboutsia* strain was present at timepoint T1, but below the detection threshold, or whether the strain linked with the genes was acquired sometime between time T1 and T2. Based on the 16S data from OIL-PCR and metagenomic sequencing, we identified the *Romboutsia* species as *R. timonensis.* Genus-level abundance data showed *R. timonensis* to be present in all three timepoints in patient B335. Due to the overall low abundance of this organism, we were unable to assemble a *Romboutsia* genome from these samples. Instead, we aligned patient B335's three samples to the *R. timonensis* (PRJEB14233) genome from NCBI, assembled the aligned reads and examined similarities between the *R. timonensis* taxonomic markers over the three timepoints (Figure 3c). We found that B335 was colonized by at least two independent strains of *R. timonensis* during the first timepoint, but that only one *R. timonensis* strain persisted during timepoint T2 and T3. One of the *R. timonensis* strains from T1 was identical to the strain from T2 and T3 across 15/30 AMPHORA marker genes, and >99% identical in 24/30 genes (Fig. 3c), suggesting that the strain of *R. timonensis* from time T2 and T3 was also present at time point T1. We found no significant difference in the normalized abundance of *Romboutsia* between time point T1 and time points T2 and T3 (Fig. 3d), albeit our data suggests that the persistent strain is the minor variant at time point 1. Despite the sensitivity of OIL-PCR, which can detect cells at least 0.1% abundant (Fig. 1b), we cannot rule out the possibility that the stochasticity of sampling in OIL-PCR and the low abundance of this particular strain of *R. timonensis* precluded our ability to observe this association at the beginning of the time course.

Discussion

Here we show the ease with which OIL-PCR can identify novel carriers of known resistance markers on extrachromosomal elements within complex bacterial communities. We applied it to a neutropenic patient's microbiome and showed the correct association of three beta-lactamases with *K. pneumoniae*, and also discovered novel associations between these beta-lactamases and two gut commensals, *R. timonensis* and *Agathobacter* spp. Two of the genes, $bla_{\text{CTX-M}}$ and bla_{TEM}, were both found on a large *Klebsiella* plasmid within the metagenome, suggesting the possible transfer of these genes to *R. timonensis* during the time course. Analysis of the plasmid sequence showed that it contains an origin of transfer, but does not have the genes necessary to transfer itself, meaning it would require a second "helper plasmid" to mobilize. Additionally, bla_{SHV} was only found on a contig belonging to the *Klebsiella* genome without any known mobilizable transposons or integrative conjugative elements nearby, severely limiting its transfer potential. An alternative explanation for our results is that *Romboutsia* and *Klebsiella* became physically associated within the gut, and thus consistently emulsified together. Both of these explanations highlight the dynamic nature of the gut microbiome, either through horizontal gene transfer, or novel physical associations between pathogens and commensals, with close physical association being a known activator for conjugal transfer of genes[26]. In future applications of OIL-PCR, primers targeting non-transferrable genes could be used to distinguish between transfer and aggregation when identical genes are associated with different taxa.

Our results illustrate the application of a streamlined, simplified fusion PCR approach to obtain robust and sensitive associations of extrachromosomal DNA with bacterial hosts. It is a practical and transportable protocol with no requirements for specialized equipment nor specialized expertise. We identify improvements in performing single cell analysis on stool, namely the incorporation of a Nycodenz purification step and the incorporation of lysozyme plus heat-induced lysis. Additionally, we increased throughput at least three-fold through primer multiplexing and developed an automated protocol

to process at least 48 samples concurrently, allowing a total of 144 gene-sample association tests per batch.

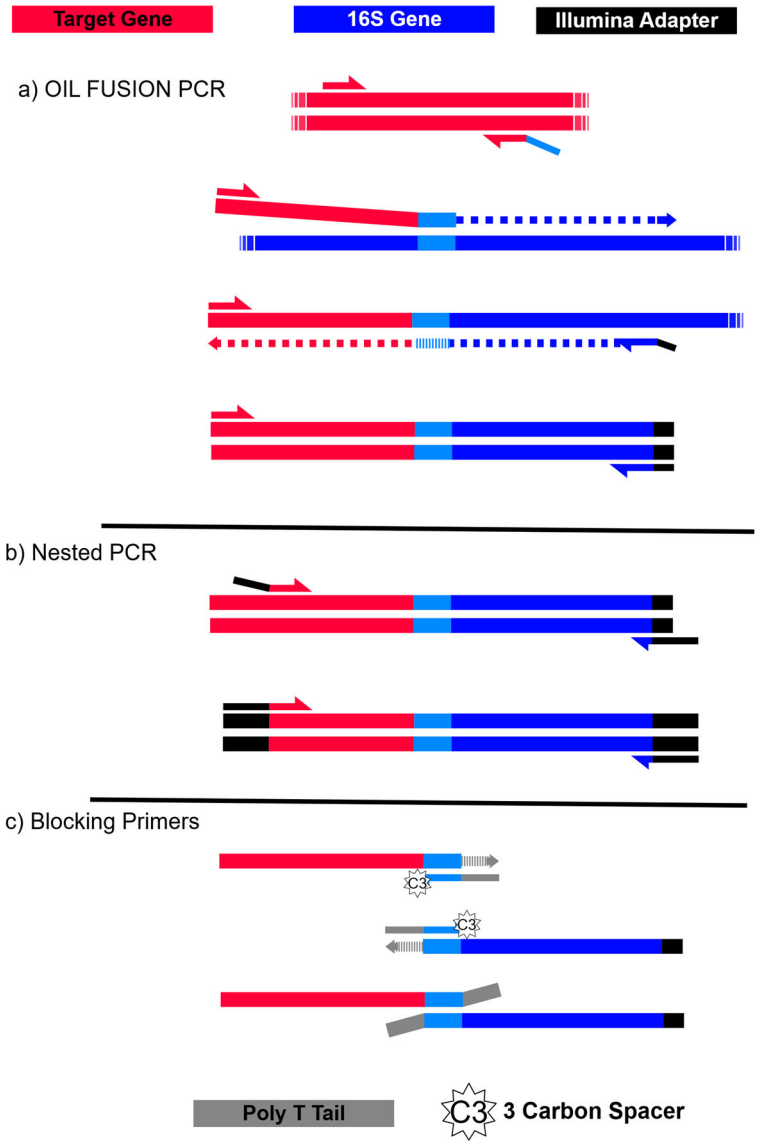
Additional improvements to OIL-PCR could be explored to further increase throughput and sensitivity. Although we tested multiplexing three genes per reaction, this number could likely be increased as we have found no sign of false positives due to multiplexing as demonstrated by associating a novel *bla*$_{\text{TEM}}$ variant with only *Escherichia* in time point T2 of patient B335 (Fig 2b, c). Further, we show that the OIL-PCR master mix facilitates permissive annealing of primers, allowing a mismatch one base from the primer's 3' end as demonstrated when *bla*$_{\text{OXY}}$ was detected in sample B324-2 with *bla*$_{\text{CTX-M}}$ primers. These results could allow for the development of highly degenerate primers to target a broad range of gene variants. Non-specific priming during OIL-PCR is not of concern because the nested PCR reaction specifically filters out undesired fusion products. Lastly, the method described currently allows 40,000 cells total per reaction. However, our TaqMan based qPCR assays suggest that the input concentration could be increased 10-fold by pre-treating cells with dsDNase (Supplementary Fig 4d). Combined with our result showing that OIL-PCR is more accurate when detecting low abundant taxa (Fig 1b), we feel confident that cell input can be increased to improve sensitivity without sacrificing accuracy.

OIL-PCR is a highly versatile platform that could be applied across fields to address a multitude of questions. While we were interested in plasmid-born ARGs in the gut, the method could be used to target any gene of interest that is difficult to associate with a host using metagenomics. As mobile genetic elements are notoriously difficult to assemble due to their promiscuity which complicates de Bruijn graph assembly[27], this method could be applied to find the hosts of integrated and non-integrated mobile elements. Similarly, as metavirome sequencing has revealed a massive number of viral genomes with unknown hosts[28], OIL-PCR may be particularly useful in addressing this gap. Additionally, viral and plasmid host-range is an important determinant for understanding and modeling bacterial ecology of predation and HGT[29]. Further, targeting functional metabolic genes detected in metagenomes, but present at low abundance in bacterial communities, could identify novel bacteria involved in nutrient cycling

345 which has remained a persistent challenge in the field of bacterial ecology[30]. Finally, when combined with
346 microfluidics, direct lysis of bacteria in an emulsion, as shown here, could be used to develop or simplify
347 single-cell genome sequencing or single-cell RNA-seq for bacteria.

348

349 **Supplemental Figures**



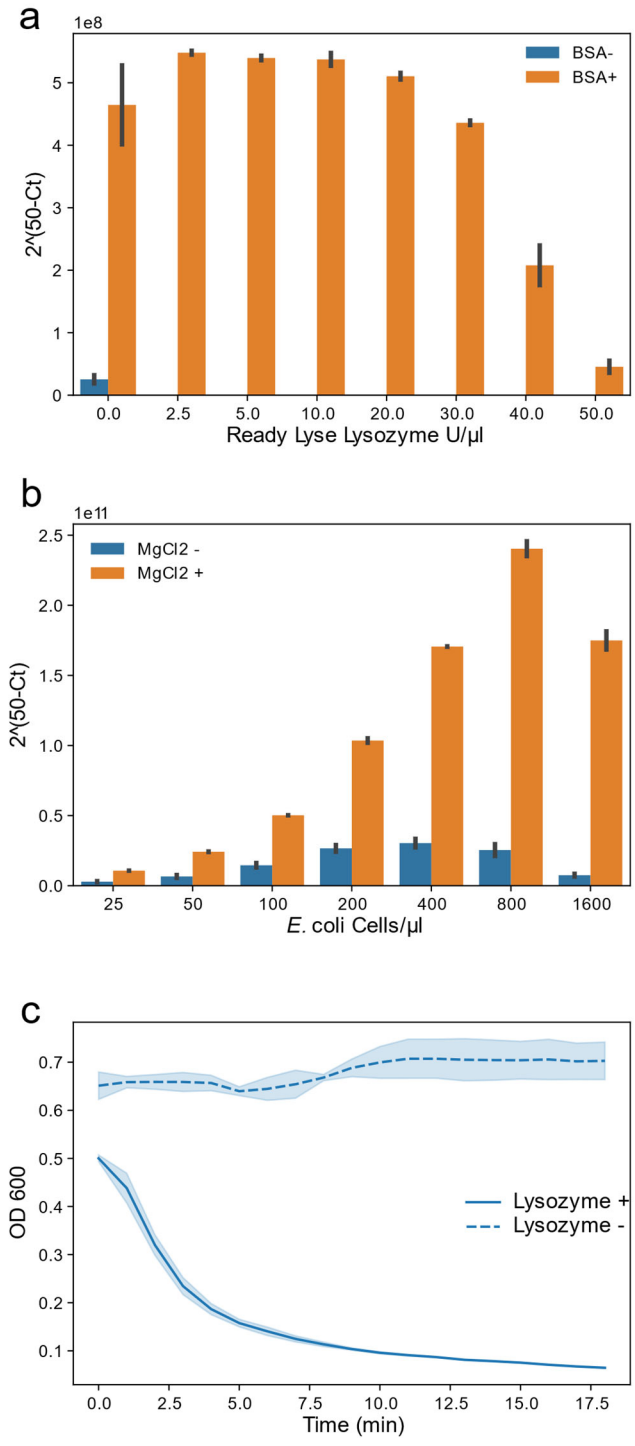
350

351 **Supplemental Figure 1. Depiction of the fusion PCR reaction**

a) PCR is initialized with primers to a target gene (red). The reverse primer contains a 5' overhang complimentary to the universal V4 16S primer 519F (light blue). The product of this first amplification step can act as a forward primer for the 16S rRNA gene (blue). After extension of the full fusion product, the forward target primer can pair with the universal 16S reverse primer 786R (with a portion of an Illumina TruSeq adapter sequence) to amplify the fully fused PCR product.

b) Nested PCR is performed on the fused PCR products from (a) in order to filter out non-specific priming from the fusion PCR. The forward primer anneals within the target gene and has a TruSeq adapter at the 5' end. The reverse primer also has the Illumina adapter sequence at it's 5'; end and anneals to the non-degenerate portion of 786R and the partial Illumina adapter sequence appended in (a).

c) Two blocking primers, both complementary to the 519F priming region, are included in the nested PCR to prevent unfused PCR products from annealing during the nested reaction. Blocking primers have a 3-carbon spacer on the 3' end to prevent extension and a poly-T tail that appends 10 As to the 3' end of any unfused products, thus inactivating them from annealing or extension.



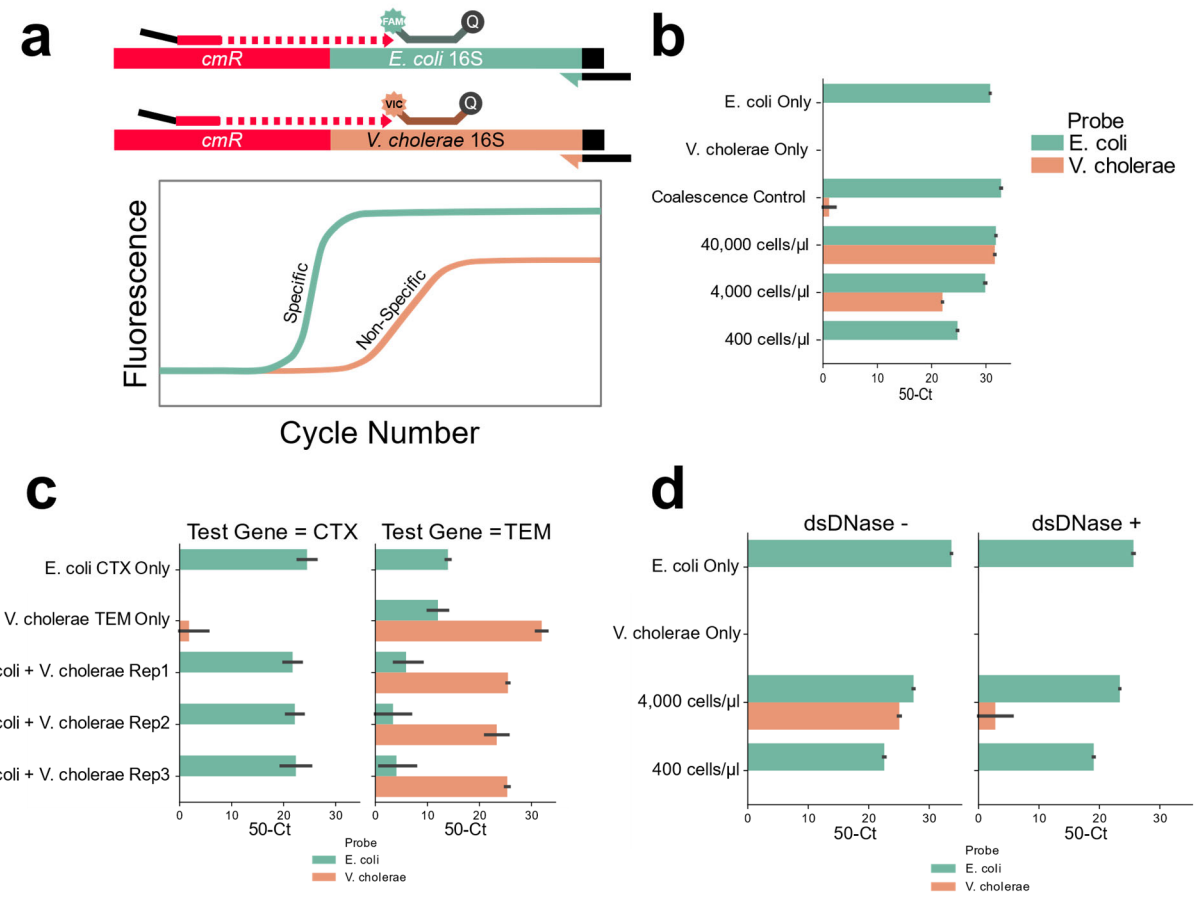
374 **Supplemental Figure 2. BSA and excess $MgCl_2$ improve the efficiency of OIL-PCR and Ready Lyse**

375 **Lysozyme remains active in OIL-PCR master mix**

a) Sybr-based qPCR was performed on the *cmR* gene carried on pBAD33 with varying concentrations of lysozyme in the presence (orange) or absence (blue) of BSA. Higher $2^{(50-Ct)}$ values represent greater amplification.

b) Sybr-based qPCR was performed on the *cmR* gene carried on the pBAD33 plasmid in *E. coli* MG1655 cells at increasing cell concentrations with (orange) and without (blue) additional $MgCl_2$. Higher $2^{(50-Ct)}$ values represent greater amplification.

c) Lysozyme activity against *B. subtilis* suspended in the OIL-PCR optimized reaction mix with (solid line) and without (dashed line) lysozyme.



Supplemental Figure 3. Cell concentration of 400cells/µl, DNase treatment, and multiplexing PCR reactions result in accurate OIL-PCR results

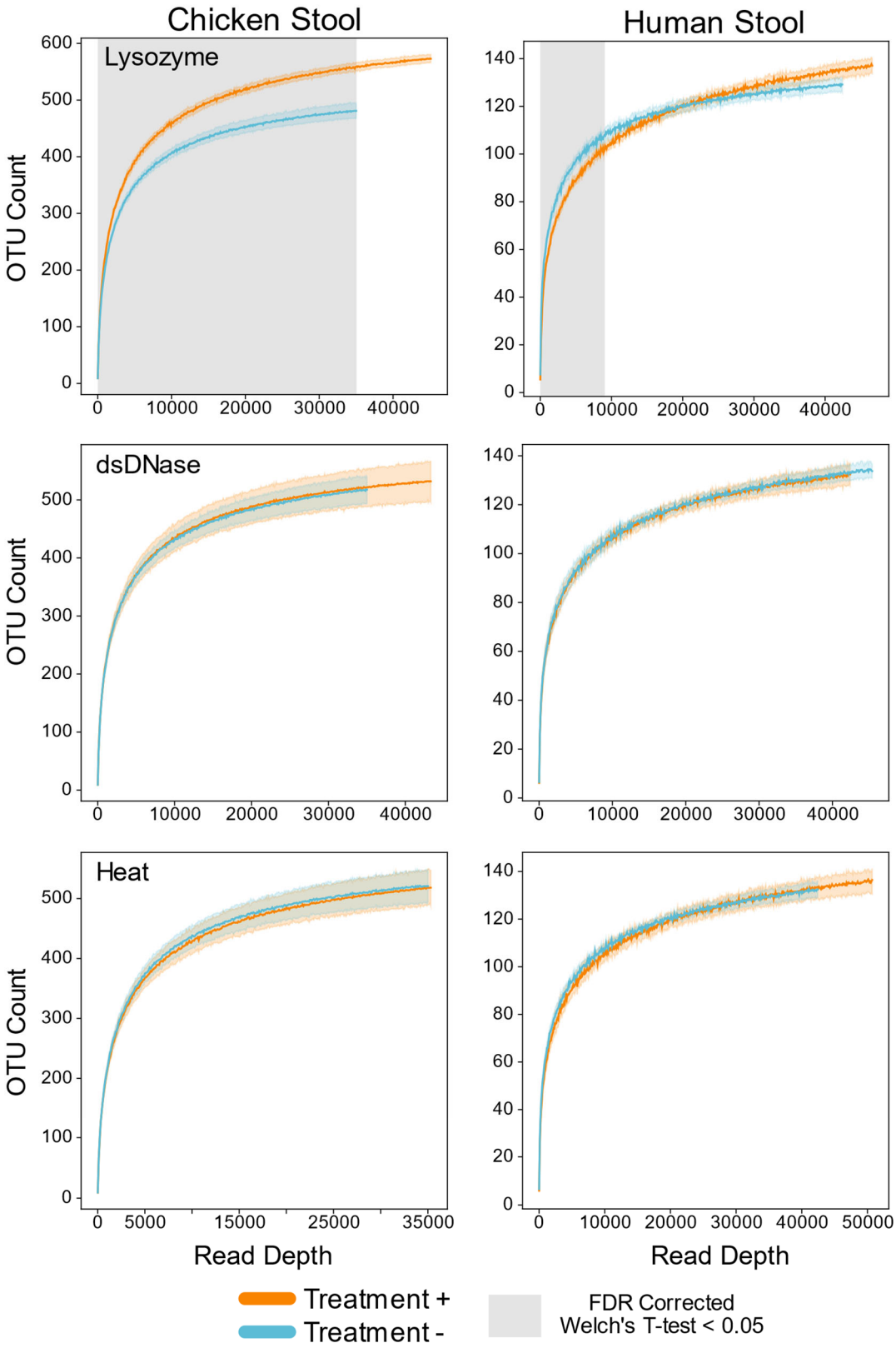
a) Diagram of the Taqman assay used to monitor OIL-PCR results. Briefly, Taqman probes were designed to be complementary for the 16S rRNA genes in either *E. coli* or *V. cholerae*, each with its own fluorophore. OIL-PCR was performed on *E. coli* carrying the *cmR* gene on the pBAD33 plasmid but not present in *V. cholerae.* Fusion PCR products were recovered and nested probe-based qPCR was performed. Upon amplification of the gene, the probe is cleaved by Taq polymerase releasing the fluorophore from the quencher. Specific amplification of the designated region is measured by fluorescence of the expected fusion product vs the non-specific product.

b) OIL-PCR with primers targeting a plasmid-borne *cmR* gene was performed with a 1:1 mix of *cmR* positive *E. coli* and *cmR* negative *V. cholerae* cell suspensions with. A gradient of cell

concentrations was tested (400-40,000 cells/µl), in addition to *E. coli* and *V. cholerae* suspensions alone as positive and negative controls. Control emulsions were mixed 1:1 after emulsification to test for droplet coalescence.

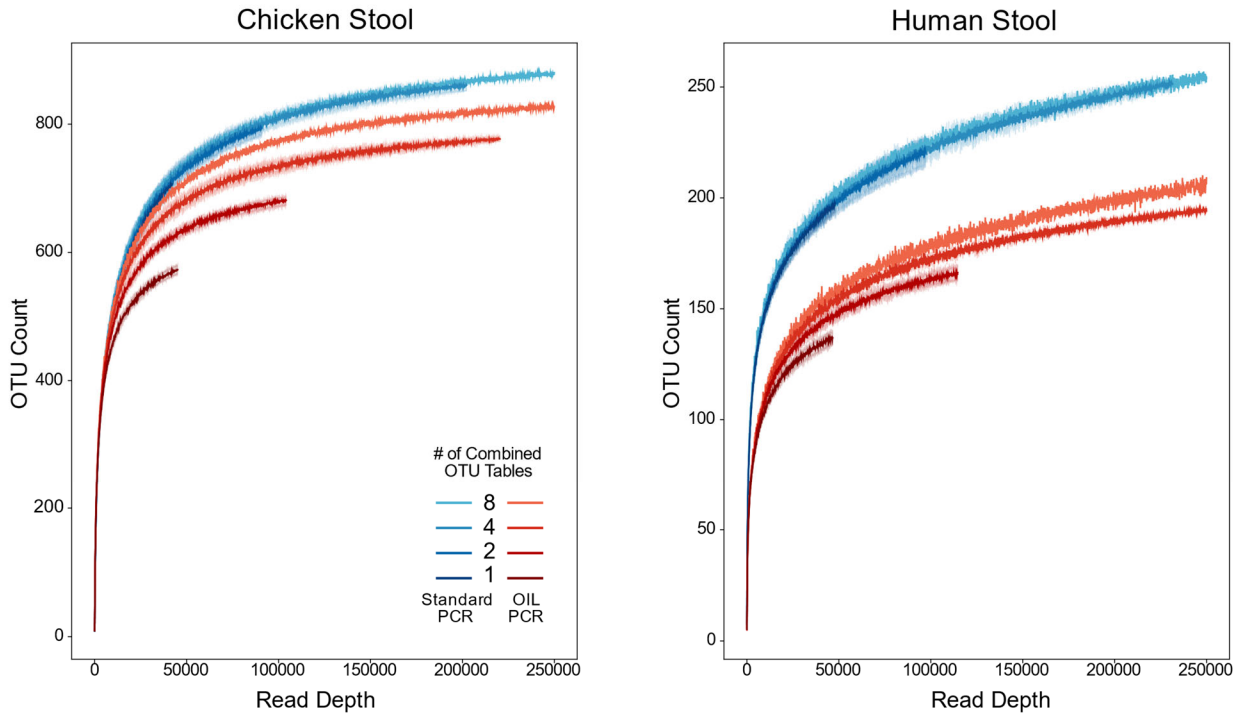
c) Multiplexed OIL-PCR was performed with primer sets targeting a genomic $bla_{CTX\text{-}M}$ gene in *E. coli* and a plasmid-borne bla_{TEM} gene in *V. cholerae.* Experiments were performed in triplicate and on each of the organisms separately. Results are shown for the $bla_{CTX\text{-}M}$ (left) and bla_{TEM} (right).

d) OIL-PCR with primers targeting a plasmid-borne *cmR* was performed after pretreating cells with (right) and without (left) dsDNase at two different 1:1 *E. coli* to *V. cholerae* cell suspension concentrations as well as on the individual bacterial strains for controls.



Supplemental Figure 4. Lysozyme alone improved recovery of species.

Rarefaction analysis of chicken (left) or human (right) gut microbiome samples with (orange) and without (blue) lysozyme (top), dsDNA treatment (middle) and heat (bottom). Grayed regions in the plot represent areas where the curves are significantly different from one another ($p<0.05$), according to an FDR-corrected Welch's t-test.



Supplemental Figure 5. Combining replicates for increased depth improved recovery of species and reduced stochastic sampling bias.

Rarefaction analysis of chicken (left) or human (right) gut microbiome samples performed with OIL-PCR (red) or standard DNA extraction and library prep (blue). OTU tables were repeatedly subsampled in groups of 2, 4, or 8 replicates from 8 total replicate 16S libraries.

Methods

Optimizing OIL-PCR Reaction Buffer for Phusion and RL Lysozyme Compatibility

SYBR-Based qPCR Assay for Phusion Polymerase Activity with Lysozyme

SYBR-based qPCR reactions were set up in duplicate as follows: 25 µl reactions with 20 U/ml Phusion Hot Start Flex DNA polymerase (NEB M0535L), 1X HF Buffer, 200 µM dNTP mix (NEB N0447L), 400 nM of 519F and 786R, 1X SYBR Green (Thermo Fisher S7563), 1X ROX reference dye (Thermo Fisher 12223012), 0.5 mg/ml of BSA (NEB B9000S) when included, 0.01% Triton-X 100, and 1µl of template DNA. Reactions were prepared with 0, 2.5, 5, 10, 20, 30, 40 and 50 U/µl of Ready-Lyse Lysozyme (Lucigen R1810M). qPCR was performed on the Thermo Fisher Quant Studio 3 Real-Time PCR machine with the following parameters: 98 °C for 1 minutes, then 50 cycles of 98 °C for 5 seconds, 54 °C for 30 seconds, and 72 °C for 30 seconds. Proper amplification was confirmed using melt curves: 98 °C for 5 seconds, cool to 60 °C at 1.6 °C/s, and then heat to 95 °C at 0.15 °C /s. Ct values and melt curves were generated with the Quant Studio software V1.4 using the default software settings.

Lysozyme Activity Assay in OIL-PCR Master Mix

Lysozyme testing was performed in a lysozyme test buffer made from the OIL-PCR master mix with dNTPs, primers, and Phusion polymerase replaced with water and 100% glycerol (48ul/ml). Log phase cultures of *B.* subtilis were standardized to an OD_{600} of 2 and suspended in 1x Lysozyme test buffer. Separately, lysozyme was suspended in 1x Lysozyme test buffer at 2x concentration. 100µl of the lysozyme mix was aliquoted into a 96-well, clear, flat-bottomed microtiter plate before adding 100µl of suspended culture. Lysis was monitored using a Spectramax M3 plate reader (Molecular Devices), heated to 37 °C and OD_{600} measured every minute for an hour.

457 Optimizing OIL-PCR Reaction Efficiency in an Emulsion

458 50µl PCR reactions were prepared in 1.5 ml tubes as describe in the Tube-Based OIL-PCR method below
459 with varying concentrations of Phusion polymerase, bovine serum albumin (BSA), Ready Lyse lysozyme,
460 dithiothreitol (DTT), $MgCl_2$, dNTPs, and ammonium sulfate. *E.* coli genomic DNA was used as template
461 and amplified with universal 16S primers 519F and 796R. Reactions were emulsified 25 Hz for 30
462 seconds before aliquoting to PCR tubes and thermocycling. Final PCR products were separated from the
463 emulsion as describe below and amplification efficiency was assessed quantitatively by SYBR-based
464 qPCR or qualitatively by gel image band intensity.

465 Emulsion Stabilization Experiments

466 OIL-PCR test buffer was prepared similarly to the lysis activity assays with ether NEB HF buffer, or
467 Detergent Free Buffer (Thermo Fisher F520L), while omitting bacterial cells or RL Lysozyme. Reactions
468 were emulsified at 25 Hz for 30 seconds on a Retch Mixer Mill MM 400 with adapters 11990 and 11993
469 (Mobio/Qiagen). Emulsion tubes were photographed before and after thermocycling and assayed by eye
470 for coalescence. After confirming a stable emulsion, qPCR and lysis time series experiments were
471 repeated to confirm activity of Phusion DNA polymerase and RL Lysozyme in the DF buffer.

472 **OIL-PCR in Tube Based Format for Master Mix Optimization**

473 Fusion PCR Reaction Setup

474 All steps were performed on ice or in a 4°C centrifuge until after emulsification. 50 µl PCR reactions
475 were prepared in a 1.5 ml microcentrifuge tube with varying experimental conditions. 2 µl of bacterial
476 cells standardized to 10^4 cells/µl were added to 48 µl of master mix and vortexed to evenly disperse cells
477 before adding 300 µl of cold Droplet Generation Oil for Probes (BioRad 1863005). Emulsions were
478 formed immediately after by shaking tubes at 25 Hz for 30 seconds on a Retch Mixer Mill MM 400 with
479 adapters 11990 and 11993 (Mobio/Qiagen). Next, the emulsion mix was divided into four 70 µl aliquots
480 in a PCR strip-tube and thermocycled as follows: 37 °C for 5 minutes, 95 °C for 10 minutes, then 38

cycles of 95 °C for 5 seconds, 54 °C for 30 seconds, and 72 °C for 30 seconds, followed by final extension 72 °C for 2 minutes. After PCR amplification, the aliquots were briefly vortexed and pooled into a clean 1.5 ml microcentrifuge tube. To break the emulsion, 50 µl of TE and 70 µl of Perfluorooctanol (Krackeler Scientific 45-370533-25G) were added and the mixture was vortexed vigorously for 30 seconds. Tubes were centrifuged at 5000 G for 1 minute and the upper aqueous phase was transferred to a new PCR strip tube and purified using AMPure XP beads as described below.

Manual AMPure XP Bead cleanup

AMPure XP beads (Beckman A63880) were added at a ratio of 0.8 µl beads per 1 µl of recovered DNA, vortexed, and incubated for 5 minutes for DNA binding. PCR strip tubes were transferred to a 96-Well magnet (Eppendorf Magnum FLX) to pull down beads for 5 minutes. Supernatant was removed with a multichannel pipette and the pellet was washed twice with 100 µl of 70% EtOH before drying for 10 minutes at room temperature. The bead pellet was suspended in 20 – 50 µl of TE and incubated for 5 minutes to elute DNA, before returning to the magnet and transferring supernatant to fresh PCR strip tubes. Eluted DNA was either run directly on a gel for qualitative analysis of amplification, or used as template in qPCR assays.

Probe-Based qPCR with TaqMan Probes for Cell Input Optimization and Multiplexing

Standardization of Bacterial Test Strains

For all experiments, bacterial type strains *Escherichia coli* MG1655[19], *Vibrio cholerae* N16961[31], and *Bacillus subtilis* 168[32] were inoculated from frozen glycerol stocks into 5 ml LB and grown at 37 °C overnight. Cultures were diluted 1:100 in 5ml fresh LB the next day and grown to OD_{600} 0.4-0.8. CFU/µl at OD_{600} was quantified by serial dilution of cells in LB, plating, and colony counting. Count results were used to standardize cell cultures to a stock concentration of 10^6 CFU/µl to be diluted and used as input for OIL-PCR.

505 Optimizing Cell Input Concentration

506 Cultures of WT *V.* cholerae N16961[31] and E. *coli* MG1655[19] carrying plasmid pBAD33[20] with *cmR* were
507 standardized to 10^4, 10^5, and 10^6 CFU/µl in LB. The two strains were mixed 1:1 at each of the three
508 concentrations, and 2 µl of cells was used as template in 50 µl tube-based OIL-PCR reactions with fusion
509 primers targeting the *cmR* gene. Reactions with each strain emulsified individually were run as controls.
510 Droplet coalescence was assayed by mixing individual control reactions after emulsification, thereby
511 ensuring the two strains were not encapsulated together. Reactions were thermocycled and recovered
512 DNA was used as template in the probe-based qPCR to quantify specific vs non-specific fusion products.

513 Probe-Based qPCR Assay

514 Probes were designed to target unique regions of *E. coli* and *V. cholerae* 16S ribosomal rRNA gene. Both
515 probes hybridized to the antisense strand and can only be cleaved when the polymerase extended from the
516 nested *cmR* primer, across the fusion junction, and into the 16S gene, thus distinguishing actual fused
517 PCR products from stray fragments of 16S DNA. Probes were verified to only target their specified
518 strain, with the *V. cholerae* probe having a VIC/NFQ MGB reporter probe and *E. coli* a FAM/NFG MGB
519 probe. 20 µl qPCR reactions were prepared in duplicate as follows: 1x Luna Universal Probe qPCR
520 Master Mix (NEB M3004L), 300 nM of forward and reverse primer, 3.2 µM of forward and reverse
521 blocking primers, 200 nM of *E. coli* and *V. cholerae* TaqMan probes, and 2-5 µl of recovered OIL-PCR
522 amplicons. Reactions were amplified under the following conditions: 95 °C for 1 minutes, then 50 cycles
523 of 95 °C for 20 seconds, 55 °C for 20 seconds, and 60 °C for 20 seconds. For analysis, Ct values were
524 subtracted from the total number of cycles for easier interpretation.

525 Primer Multiplexing Validation

526 Four strains of bacteria were used to test primer multiplexing: *V. cholerae* N16961[31] carrying *ampR* on
527 RP4 plasmid[33], WT *V. cholerae* N16961, *E. coli* 0006 (CDC & FDA Antibiotic Resistance Isolate Bank)
528 carrying $bla_{\text{CTX-M-15}}$, and WT *E. coli* MG1655[19], all mixed at a ratio of 1:49:10:40 with a final

concentration of 10^4 cells/µl. This mix of cells resulted in 10% of the consortium carrying *bla*$_{\text{CTX-M-15}}$ and 1% carrying *amr*R to provide a more realistic depiction of the abundances of ARGs in natural stool communities. OIL-PCR was performed in a plate-based format with forward and fusion primers for both *amrR* and *bla*$_{\text{CTX-M-15}}$. Each strain was tested individually as controls. Purified fusion products were assayed for correct fusions using the probe-based qPCR assay with nested primers targeting *amrR* or *bla*$_{\text{CTX-M-15}}$ in parallel reactions.

Final OIL-PCR Parameters:

OIL-PCR:

The final, optimized OIL-PCR master mix is as follows: 100 U/µl Phusion Hot Start Flex DNA Polymerase, 1X DF Buffer (Thermo Fisher F520L), 250 µM dNTPs (NEB N0447L), 2 µM universal 16S reverse primer 786R, 1 µM of each target specific forward primer, 0.01 µM of each target specific fusion primer with universal 519F' tail, 1.5 mM additional $MgCl_2$, 5mM Ammonium Sulfate, 5 mM DTT, 4 mg/ml BSA (NEB B9000S), 300 U/µl RL Lysozyme, 400 cells/µl Nycodenz purified cells. 300 µl emulsion oil (BioRad 1863005) was added to 50 µl reactions when performed in individual tubes, or 200 µl of emulsion oil was added to 100 µl OIL-PCR reactions when performed in the 96-well plate format. Tubes were emulsified at 25 Hz for 30 seconds, while plates were sealed with a 50 µm aluminum seal (Axygen PCR-AS-600) and emulsified for 2 rounds of 27.5 Hz for 20 seconds; flipping the plate in between for consistent emulsion across rows.

The lysis and amplification program is: 37 °C for 5 minutes, 95 °C for 10 minutes, then 38 cycles of 95 °C for 5 seconds, 54 °C for 30 seconds, 72 °C for 30 seconds, before final extension of 72 °C for 2 minutes.

dsDNase Treatment and Heat Inactivation in OIL-PCR

dsDNase treatment was not used in the initial OIL-PCR optimization or spike-in experiments. Cells were standardized to 10^4 cells/µl in 100 µl of PBS. 1 µl of stock dsDNase (Thermo Fisher EN0771) was added

to the tube and incubated at room temp for 10 minutes before returning to ice. Treated cells were used directly in OIL-PCR. The enzyme was inactivated immediately after emulsification (optional) by incubating 10 minutes in a water bath set exactly to 50 °C with gentle mixing by hand every 2 minutes.

Nested PCR:

SYBR-based qPCR was performed on purified fusion PCR products to minimize the number of cycles for each reaction with the goal of reducing chimera formation. Amplification was performed in 20 µl reactions using the Luna Universal qPCR Master Mix (NEB M3003L) with 1x PCR master mix, 300 nM forward and reverse primers, and 2-5 µl of purified template. For multiplexed experiments, separate reactions were prepared, with one set of nested primers for each gene assayed. The following thermocycling conditions were used: 95 °C for 2 minutes, 40 cycles of 95 °C for 15 seconds, 55 °C for 15 seconds, 68 °C for 20 seconds, followed by a final extension phase at 68 °C for 1 minutes. Melt curves were measured by heating to 95 °C at 0.15 °C /s. Blocking primers were not included in SYBR-based qPCR reactions because of the strong signal from self-hybridization. Ct values were used to select the cycle number for nested amplification that was equal to the Ct value +/- 2 cycles. Reactions that did not amplify in the qPCR were amplified with the highest number of cycles for that preparation.

Using the qPCR results to select the cycle number, nested PCR reactions were prepared in duplicate 20 µl reactions as follows: 20 U/ml Phusion DNA polymerase, 1x HF Buffer, 2 µM dNTPs, 300 nM target gene specific forward primer and universal reverse primer, 32 µM of each blocking primer, and 2-5 µl of template. Thermocycling was performed with variable number of cycles based on the qPCR as follows: 98 °C for 3 minutes, then variable cycles of 98 °C for 5 seconds, 55 °C for 30 seconds, and 72 °C for 30 seconds, followed by final extension 72 °C for 5 minutes. Duplicate PCR reactions were pooled and purified using automated AMPure XP cleanup.

Illumina Indexing PCR and Library Preparation

Custom indexing primers were designed based on Spencer et al[11]. A set of unique, 9 bp barcodes was generated using Barcode Generator V2.8[34]. The primers are compatible with the Illumina Truseq primers and the index can be read with 8 bp instead of 9 to make them compatible with other libraries.

Indexing PCR was performed with 25 µl reactions as follows: 20 U/ml Phusion DNA polymerase, 1x HF Buffer, 2 µM dNTPs, 100 nM of unique forward and reverse indexing primers, and 2 µl of purified nested PCR template. Cycling was performed as follows: 98 °C for 1 minutes, then 20 cycles of 98 °C for 15 seconds, 56 °C for 30 seconds, and 72 °C for 45 seconds, followed by final extension 72 °C for 2 minutes. PCR reactions were purified using automated AMPure XP cleanup.

Indexed PCR libraries were quantified using QUANT-IT pico green dsDNA assay kit (Invitrogen P7589) and measured on the Spectramax M3 plate reader. Wells were pooled based on the measured concentration using the Eppendorf epMotion 5075vtc robot and the final pool quantified using the Qubit Broad Range Assay Kit (Thermo Fisher Q32853). Pools were run on a gel to confirm clean DNA before sequencing with MiSeq 2x250 V2 chemistry.

Plate-Based OIL-PCR With Robotic Automation:

Reaction Setup

96 µl of the final OIL-PCR master mix was aliquoted into a 500 µl deep well plate (Eppendorf 00.0 501.101). Nycodenz purified stool cells were diluted in PBS to 10^4 cells/µl in an 8-well PCR strip for multichannel pipetting. 4 µl of cells was quickly added to the reactions with a 10 µl 8-channel pipette before sealing with an extra-thick foil seal (Axygen PCR-AS-600) and vortexed to mix. The reactions were briefly centrifuged to return liquid to the bottom of the plate, and then placed on an orbital microplate shaker (VWR 12620-926) at 1200 rpm for 30 seconds to further mix the cells while keeping the mix at the bottom of the wells. After mixing, the foil seal was carefully removed and 200 µl of cold emulsion oil was added using a multichannel pipette. The plate was then sealed with a fresh foil seal and

shaken at 27.5 Hz for 20 seconds on the Retch shaker MM 400 with plate adapter (#11990). The plate was removed and turned over to shake an additional 20 seconds providing an even emulsion across the plate. After emulsifying, each reaction was aliquoted into 4 wells of a PCR plate (Eppendorf 0030 128.648) using the robot for consistency. The plates were sealed and run on the OIL-PCR fusion program described earlier.

DNA Recovery from Emulsion

After amplification, the robot was used to purify the OIL-PCR Products. In short, replicate reactions were pooled into a fresh 500 µl deep well plate, and 60 µl of TE and 70 µl of Perfluorooctanol (Krackeler Scientific 45-370533-25G) were added to each well. The plate was sealed and shaken on the Retch at 30 Hz for 40 seconds to thoroughly disrupt the emulsion. The plate was then centrifuged in a swing bucket rotor at 5000 Gs for 1 minute to separate the phases and returned to the robot. 80 µl of the upper phase was aspirated from a defined height into a fresh 500 µl deep well plate for automated Ampure XP bead purification.

Automated AMPure XP Bead Purification

85 µl of AMPure XP beads (Beckman A63880) was added to the deep well plate containing the recovered OIL-PCR fusion products. The reactions were mixed at 1200 rpm for 1 minute and incubated for 2 minutes for DNA binding, before transferring to the magnet (Eppendorf Magnum FLX) for 3 minutes. After pulldown, the supernatant was discarded and the wells were washed twice with 200 µl of 70% EtOH. After discarding the second wash, the plate was removed from the magnet and dried at room temp for 10 minutes before adding 50 µl of TE buffer. The plate was shaken at 1200 rpm for 1 minute and incubated 2 minutes to elute the DNA. Finally, the plate was returned to the magnet and for 2 minutes and

48 µl of purified DNA was transferred to a fresh 96-well PCR plate (Eppendorf 0030 128.648).

OIL-PCR on Natural Stool Communities

Nycodenz Purification of Stool Cells

All steps were performed on ice and in a 4c refrigerated centrifuge unless otherwise noted. Stool samples were collected in PBS + 20% glycerol + 0.1% L-cysteine and frozen at -80c until processed. Frozen samples were thawed completely and thoroughly homogenized via vortexing. Samples were diluted at least 1:1 in cold PBS to reduce the sample viscosity and glycerol concentration as viscous samples did not to separate well with the Nycodenz. Samples were vortexed at maximum speed for 5 minutes to release cells from stool particles. 300µl of cold 80% Nycodenz (VWR 100356-726) was aliquoted to the bottom of 2 ml microcentrifuge tubes and 1.6 ml of stool slurry was overlaid on top without mixing the two phases. Tubes were centrifuged at 10,000 G for 40 minutes in a swing bucket rotor to separate cells. After centrifugation, the upper phase was removed with a pipetted and 500µl of cold PBS was used to wash the bacterial cell pellet from the insoluble stool fraction. The suspended cells were removed and the pellet was washed a second time with 500 µl of PBS. Cells were centrifuged at 50g for 1 minute to pellet any large particles that carried over from the Nycodenz purification and the upper phase was passed through a 40 µm nylon mesh screen (Falcon 352235) to remove any residual stool debris or large cell clumps. Samples of each preparation were diluted 1:1 PBS + 20% glycerol for whole cell storage. Lastly, purified cells were diluted and imaged at 100x magnification within a 20 µm counting chamber (VWR 15170-048). Images were analyzed using FIJI/ImageJ 1.52p (Java 1.8.0_172) to manually count cells and calculate cell concentration in glycerol stocks.

Spike-In Experiment

This experiment was performed using the individual tube-based format of OIL-PCR. Nycodenz purified stool and *E. coli* carrying pBAD33[20] with *cmR* was standardized to 10^4 cells/µl. *E. coli* cells were mixed with the stool samples at a ratio of 1:10, 1:100, and 1:1000, and the mixed cultures were added to OIL-PCR containing the *cmR* primer set. Reactions were emulsified, lysed, thermocycled, and fusion products

were purified manually. Nested PCR was performed with the nested *cmR* primer before indexing, pooling, and sequencing.

Lysozyme, dsDNase, Heat Experiment

Nycodenz-purified human and chicken stool cells were standardized to 10^5 cells/µl and incubated with or without dsDNase at room temp for 10 minutes. OIL-PCR master mix was prepared with and without Lysozyme using universal 16S rRNA primers i519F and i786R. Cells were added to the OIL-PCR reaction and emulsified. Emulsions were either incubated at 50 °C or room temperature for 10 minutes before aliquoting to PCR plates and running the OIL-PCR fusion program. Amplicons were purified, indexed, and submitted for Illumina sequencing as described above.

OIL-PCR for Detection of *bla* genes in Neutropenic Patients

Sample collection and Metagenomic Assemblies

Samples were collected, sequenced, and metagenomic assemblies were prepared as described in Kent et al.[2] Briefly, serial stool samples were collected from consenting individuals receiving a hematopoietic stem cell transplant at NewYork-Presbyterian Hospital/Weill Cornell Medical Center in accordance with IRB protocols for Weill Cornell Medical College (#1504016114) and Cornell University (#1609006586). Samples were either frozen “as is” (for metagenomic sequencing) or homogenized in phosphate-buffered saline (PBS) + 20% glycerol before freezing (for OIL-PCR). DNA was isolated from samples destined for metagenomic sequencing using the PowerSoil DNA Isolation Kit (Qiagen) with additional proteinase K treatment and freeze/thaw cycles recommended by the manufacturer for difficult-to-lyse cells. Extractions were further purified using 1.8 volumes of Agencourt AMPure XP bead solution (Beckman Coulter). DNA was diluted to 0.2 ng/µL in nuclease-free water and processed for sequencing using the Nextera XT DNA Library Prep Kit (Illumina).

671 Design and Validation of OIL-PCR fusion primers

672 ARG variants for the three *bla* genes were downloaded from the CARD database[35] and aligned in
673 Snapgene using default MUSCLE parameters. Conserved regions were identified manually and
674 degenerate primers were designed to capture as many variants of the genes as possible. Primers were
675 selected for GC content between 40 – 60% and an annealing temperature of 58 °C based on the Snapgene
676 calculation. Degenerate bases were limited to 3 per primer and no less than 5 bp from the 5' end.

677 Strains acquired through the CDC & FDA Antibiotic Resistance Isolate Bank carrying multiple variants
678 of each gene (Supplementary Table T1)) were used as template for testing *bla* primers. At least 3 sets of
679 primers were designed and tested in every possible combination using the OIL-PCR master mix without
680 emulsion to find a set of three primers that provided clean fusion amplification. Lastly, working primer
681 sets were tested in an emulsion on whole cells to confirm amplification in OIL-PCR.

682 Fusion Primers targeting Tn3 transposon genes were designed using scaffolds from the metagenomic
683 assemblies and tested on *Klebsiella* isolate DNA from patient B335.

684 OIL-PCR on Neutropenic Patients

685 All OIL-PCR reactions were performed with the plate-based protocol including dsDNase treatment and
686 heat inactivation. Whole bacterial cells were purified with Nycodenz, quantified and standardized to 10^4
687 cells/µl in PBS before treating with dsDNase. For multiplexed experiments, reactions were prepared in
688 quadruplicate with three sets of primers targeting the three *bla* genes in each reaction. The singleplex
689 reactions were prepared in triplicate with only one primer set per reaction. In all cases, the reactions
690 followed the standard plate-based protocol with automation, including heat inactivation of the dsDNase
691 after emulsification. Nested PCR, indexing, and library preparation was performed as described above.

692

693

694 Romboutsia Specific OIL-PCR

695 CRIB primers[25] were modified to form a fusion product with all three *bla* genes, however, only the *bla*$_{TEM}$
696 primer set amplified when tested. Using only the *bla*$_{TEM}$ primer set, OIL-PCR was performed with
697 dsDNase treatment, in triplicate, using the plate-based format with automation. Nested PCR, indexing,
698 and library preparation was performed as described above.

699 **Computational Methods**

700 Processing 16S rRNA Sequencing

701 Raw reads were merged using usearch[36] (V 11.0.667) -fastq_mergepairs (maxdiffs: 20, pctid: 85,
702 minmergelen: 283, maxmergelen: 293) before trimming primers and quality filtering with usearch -
703 fastq_filter (maxee: 1.0). Unique reads were filtered using usearch -fastx_uniques and OTUs were
704 clustered based on 97% identity with usearch -cluster_otus. OTU tables were generated with usearch -
705 otutab and taxonomy was assigned with RDP classifier implemented in MOTHUR classify.sequs (1.38.1)
706 against silva v132. Rarefaction curves were generated using QIIME1[37] (v1.9) multiple_rarefaction.py (-m
707 10, -x 100000, -s 100, -n 5, -k).

708 Processing OIL-PCR Sequencing

709 Raw reads were merged using usearch[36] (V 11.0.667) -fastq_mergepairs (maxdiffs: 10% of expected
710 overlap, pctid: 85, minmergelen: expected length-15, maxmergelen: expected length +15) before
711 trimming primers and quality filtering with usearch -fastq_filter (maxee: 1.0). Unique reads were filtered
712 using usearch -fastx_uniques. Reads were split at the fusion junction into 16S and target reads using
713 cutadapt V2.1[38] because of its tolerance for PCR errors which are often introduced in the fusion junction
714 of the OIL-PCR amplicons. The 16S reads were clustered based on 97% identity with usearch -
715 cluster_otus, OTU tables were generated with usearch -otutab and taxonomy was assigned with RDP
716 classify implemented in mothur[39] classify.sequs (1.38.1) against SILVA[40] v132. Target reads were
717 identified by blasting against a custom database of expected sequences with blastn[41] (v2.9.0). 16S

718 taxonomy and target read identity were then reassociated using a custom python script to parse the files.
719 Detections were defined by taxa – target associations that make up 0.5% of the total reads across
720 replicates.

721 Strain Level Analysis of *Romboutsia* in Metagenomes

722 Metagenomic reads from each time point were aligned to the R. timonensis reference genome (Refseq
723 accession code: GCF_900106845.1) using BWA mem (v0.7.17, -a)[42]. Reads aligning to the reference
724 genome were then assembled using SPAdes (v3.14.)[43]. To determine the presence and identity of strains
725 from each time point, AMPHORA[44] (v2, marker identification step only) was used to identify the
726 sequences of 30 marker genes within each assembled *R. timonensis* genome. The marker genes identified
727 by AMPHORA were then mapped (Diamond, v2.0.4)[45] to the BLAST[46] nr database for taxonomic
728 annotation (BLAST nr database downloaded 2018). DNA sequences of the marker genes that mapped to
729 *R. timonensis* were retained for further analysis. Genes from time point 2 and time point 2 were aligned to
730 one another (BLAST blastn, v2.9.0)[41], and then sequences from time point 1 were aligned against
731 sequences of the same gene from time point 2, once the sequences at time 2 and time 3 were determined
732 to be the same. To determine how abundant each marker gene, and all of its variants, are at each time
733 point, metagenome reads from each time point were mapped to its own set of marker gene sequences
734 (BWA mem, v0.7.17, -a)[42]. Read counts were normalized for the length of each gene and the total number
735 of reads sequenced per sample (RPKM)[47].

Acknowledgments

We would like to thank the following individuals and organizations for their generosity in providing us with strains and plasmids: Tobias Dörr (STRAINS), Barth Smets (RP4), John Helmann (*B. subtilis*) and the CDC & FDA Antibiotic Resistance (AR) Isolate Bank. We would like to thank Sarah Spencer for technical advice. This study was funded by the Centers for Disease Control (OADS BAA 2016-N-17812) and by the National Sciences Foundation (Awards #1661338 and #1650122). F.N. is a SUNY Diversity Fellow. M.J.S. is funded by the NIAID (K23 AI114994). I.L.B. is funded by the NIH (1DP2HL141007-01) and is a Sloan Foundation Research Fellow, a Packard Fellowship in Science and Engineering, and a Pew Foundation Biomedical Scholar.

Conflict of Interest Statement

M.J.S has received grant funding from BioFire Diagnostics, Allergan, and Merck and has received consulting fees from Shionogi and Achaogen. All other authors declare that they have no conflicts of interest.

References

1. Sommer MO a, Dantas G, Church GM. Functional Characterization of the Antibiotic Resistance Reservoir in the Human Microflora. *Science (80-)*. 2009;325(5944):1128-1131. doi:10.1126/science.1176950

2. Kent AG, Vill AC, Shi Q, Satlin MJ, Brito IL. Widespread transfer of mobile antibiotic resistance genes within individual gut microbiomes revealed through bacterial Hi-C. doi:10.1038/s41467-020-18164-7

3. Beaulaurier J, Zhu S, Deikus G, et al. Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation. *Nat Biotechnol*. 2018;36(1):61-69. doi:10.1038/nbt.4037

4. Xu L, Brito IL, Alm EJ, Blainey PC. Virtual microfluidics for digital quantification and single-cell sequencing. *Nat Methods*. 2016;13(9). doi:10.1038/nmeth.3955

5. Lan F, Demaree B, Ahmed N, Abate AR. Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nat Biotechnol*. 2017;35(7):640-646. doi:10.1038/nbt.3880

6. Zou Y, Xue W, Luo G, et al. 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. *Nat Biotechnol*. 2019;37(2):179-185. doi:10.1038/s41587-018-0008-8

7. Poyet M, Groussin M, Gibbons SM, et al. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat Med*. 2019;25(9):1442-1452. doi:10.1038/s41591-019-0559-3

8. Ottesen EA, Hong JW, Quake SR, Leadbetter JR. *Microfluidic Digital PCR Enables Multigene Analysis of Individual Environmental Bacteria*. http://science.sciencemag.org/. Accessed April 9, 2020.

9. Zeng Y, Novak R, Shuga J, Smith MT, Mathies RA. High-Performance Single Cell Genetic Analysis Using Microfluidic Emulsion Generator Arrays. *Proc Natl Acad Sci USA*. 2008;105(4):3183-3190. doi:10.1021/ac902683t

10. Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R. Probing individual environmental bacteria for viruses by using microfluidic digital PCR. 2011;333(6038):58-62.

11. Spencer SJ, Tamminen M V, Preheim SP, et al. Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers SUPPLIMENT. *ISME J*. 2015;(September):1-10. doi:10.1038/ismej.2015.124

12. Liu Y, Schulze-Makuch D, de Vera JP, et al. The development of an effective bacterial single-cell lysis method suitable for whole genome amplification in microfluidic platforms. *Micromachines*. 2018;9(8). doi:10.3390/mi9080367

13. Tamminen M V., Virta MPJ. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Front Microbiol*. 2015;6(MAR):1-10. doi:10.3389/fmicb.2015.00195

14. Finn TE, Nunez AC, Sunde M, Easterbrook-Smith SB. Serum albumin prevents protein aggregation and amyloid formation and retains chaperone-like activity in the presence of physiological ligands. *J Biol Chem*. 2012;287(25):21530-21540. doi:10.1074/jbc.M112.372961

15. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. 2015;161(5):1202-1214. doi:10.1016/j.cell.2015.05.002

16. Holmsgaard PN, Norman A, Chr S, et al. Bias in bacterial diversity as a result of Nycodenz extraction from bulk soil. *Soil Biol Biochem*. 2011;43(10):2152-2159. doi:10.1016/j.soilbio.2011.06.019

17. Hevia A, Delgado S, Margolles A, Sanchez B. Application of density gradient for the isolation of the fecal microbial stool component and the potential use thereof. *Sci Rep*. 2015;5(1):16807. doi:10.1038/srep16807

18. Vorkapic D, Pressler K, Schild S. Multifaceted roles of extracellular DNA in bacterial physiology. *Curr Genet*. 2016;62(1):71-79. doi:10.1007/s00294-015-0514-x

19. Blattner FR, Plunkett G, Bloch CA, et al. The complete genome sequence of Escherichia coli K-12. *Science (80-)*. 1997;277(5331):1453-1462. doi:10.1126/science.277.5331.1453

20. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. *J Bacteriol*. 1995;177(14):4121-4130. doi:10.1128/jb.177.14.4121-4130.1995

21. Cantón R, González-Alba JM, Galán JC. CTX-M enzymes: Origin and diffusion. *Front Microbiol*. 2012;3(APR):110. doi:10.3389/fmicb.2012.00110

22. Woerther PL, Burdet C, Chachaty E, Andremont A. Trends in human fecal carriage of extended-spectrum β-lactamases in the community: Toward the globalization of CTX-M. *Clin Microbiol Rev*. 2013;26(4):744-758. doi:10.1128/CMR.00023-13

23. Poirel L, Naas T, Nordmann P. Genetic support of extended-spectrum β-lactamases. *Clin Microbiol Infect*. 2008;14(SUPPL. 1):75-81. doi:10.1111/j.1469-0691.2007.01865.x

24. Anou M. Somboro, John Osei Sekyere, Daniel G. Amoako, Sabiha Y. Essack LAB. Diversity and Proliferation of Metallo-Beta-Lactamases: a Clarion Call for Clinically Effective Metallo-Beta-Lactamase Inhibitors. *Appl Environ Microbiol*. 2018;84(18):1-20.

25. Gerritsen J, Fuentes S, Grievink W, et al. Characterization of Romboutsia ilealis gen. nov., sp. nov., isolated from the gastro-intestinal tract of a rat, and proposal for the reclassification of five closely related members of the genus Clostridium into the genera Romboutsia gen. nov., Intestinib.

Int J Syst Evol Microbiol. 2014;64(Pt 5):1600-1616. doi:10.1099/ijs.0.059543-0

26. Clark RR, Judd J, Lasek-Nesselquist E, et al. Direct cell–cell contact activates SigM to express the ESX-4 secretion system in Mycobacterium smegmatis. *Proc Natl Acad Sci U S A*. 2018;115(28):E6595-E6603. doi:10.1073/pnas.1804227115

27. Antipov D, Raiko M, Lapidus A, Pevzner PA. Plasmid detection and assembly in genomic and metagenomic data sets. *Genome Res*. 2019;29(6):961-968. doi:10.1101/gr.241299.118

28. Shkoporov AN, Hill C. Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. *Cell Host Microbe*. 2019;25(2):195-209. doi:10.1016/j.chom.2019.01.017

29. Flores CO, Meyer JR, Valverde S, Farr L, Weitz JS. Statistical structure of host–phage interactions. *Proc Natl Acad Sci* . 2011;108(28):E288-E297. doi:10.1073/pnas.1101595108

30. Preheim SP, Olesen SW, Spencer SJ, et al. Surveys, simulation and single-cell assays relate function and phylogeny in a lake ecosystem. *Nat Microbiol*. 2016;1(9):1-9. doi:10.1038/nmicrobiol.2016.130

31. Heidelberg JF, Elsen JA, Nelson WC, et al. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature*. 2000;406(6795):477-483. doi:10.1038/35020000

32. Kunst F, Ogasawara N, Moszer I, et al. The complete genome sequence of the gram-positive bacterium Bacillus subtilis. *Nature*. 1997;390(6657):249-256. doi:10.1038/36786

33. Klümper U, Riber L, Dechesne A, et al. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J*. 2015;9(4):934-945. doi:10.1038/ismej.2014.191

34. Comai L, Howell T. Barcode Generator. 2012. http://comailab.genomecenter.ucdavis.edu/index.php/Barcode_generator.

841 35. Alcock BP, Raphenya AR, Lau TTY, et al. CARD 2020: Antibiotic resistome surveillance with
842 the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2020;48(D1):D517-D525.
843 doi:10.1093/nar/gkz935

844 36. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*.
845 2010;26(19):2460-2461. doi:10.1093/bioinformatics/btq461

846 37. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput
847 community sequencing data. *Nat Methods*. 2010;7(5):335-336. doi:10.1038/nmeth.f.303

848 38. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
849 *EMBnet.journal*. 2011;17(1):10. doi:10.14806/ej.17.1.200

850 39. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: Open-source, platform-
851 independent, community-supported software for describing and comparing microbial
852 communities. *Appl Environ Microbiol*. 2009;75(23):7537-7541. doi:10.1128/AEM.01541-09

853 40. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: Improved
854 data processing and web-based tools. *Nucleic Acids Res*. 2013;41(D1):D590.
855 doi:10.1093/nar/gks1219

856 41. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC*
857 *Bioinformatics*. 2009;10(1):421. doi:10.1186/1471-2105-10-421

858 42. Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*.; 2013.

859 43. Nurk S, Bankevich A, Antipov D, et al. Assembling Single-Cell Genomes and Mini-Metagenomes
860 From Chimeric MDA Products. *J Comput Biol*. 2013;20(10):714-737. doi:10.1089/cmb.2013.0084

861 44. Wu M, Scott AJ. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2.
862 *Bioinformatics*. 2012;28(7):1033-1034. doi:10.1093/bioinformatics/bts079

863 45. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat*
864 *Methods*. 2015;12(1):59-60. doi:10.1038/nmeth.3176

865 46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol*
866 *Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

867 47. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying
868 mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621-628.
869 doi:10.1038/nmeth.1226

870