

Supporting Information

Leung et al. 10.1073/pnas.1106752109

SI Text

Microfluidic Device Fabrication. Devices were fabricated using multilayer soft lithography (1). The device uses a three-layer design: The top layer is a “flow layer,” containing channels for reagent manipulation. The middle layer is a “control layer,” containing channels used for pneumatic valves. The bottom layer is a “blank layer,” to which control channels are sealed. All devices were made from polydimethylsiloxane (PDMS) (RTV615; General Electric). Devices were bonded to glass slides after plasma treatment of the bottom of the device and the slide (Harrick Plasma). Photolithography masks were designed by using AutoCAD software (Autodesk) and used to generate high-resolution (20,000 dpi) transparency masks (CAD/Art Services). Molds were fabricated by photolithography on 10.2-cm silicon wafers (Silicon Quest International). The flow layer consisted of three different profiles: 5 μm-high rectangular frits, 12 μm-high rounded channels, and 200 μm-high cylindrical storage chambers. The 5 μm layer was made with SU8-5 negative photoresist (Microchem Corp.), the 12 μm rounded layer was made with SPR220-7 positive photoresist (Microchem Corp.), and the 200 μm layer was made with SU8-100 negative photoresist (Microchem Corp.). The control layer consisted of two different profiles: 25 μm-high rectangular channels used for valves and 5 μm-high features used for sections of control lines passing under flow channels where valving was unwanted. The 5 μm layer was made with SU8-5 negative photoresist and the 25 μm layer was made with SU8-2025 negative photoresist (Microchem Corp.). Resist processing was performed according to the manufacturer’s specifications. Masters machined in Poly(methyl methacrylate) were used to mold the elution nozzle on the chip.

Device Operation. Device operation was automated using custom software written in LabVIEW (National Instruments). On-chip valve actuation was controlled using pneumatic solenoid actuators (Fluidigm) connected to a PCI-6533 digital input/output card (National Instruments). A single LabVIEW program was used to execute user-designed formulation scripts input as text files. Compressed air (5 psi–20 psi) was used to push reagents into the device. Prior to experiments, devices were dead-end filled with carrier fluid which was then flowed through the chamber array at 0.5 μL/min for approximately 1 hr.

Prior to elution of on-chip reactions into microfuge tubes, each tube is filled with light mineral oil, which wets PDMS preferentially over both the fluorinated carrier fluid and the aqueous phase, preventing any aqueous sample from adhering to the nozzle surfaces. The low density of light mineral oil also ensures that the eluted aqueous sample sinks to the bottom of the well, away from the nozzle tip. Between elution of each storage element, the nozzle is rinsed in an isopropanol bath to wash away any aqueous droplets that may remain attached to the nozzle’s exterior that can lead to sample carry-over. After elution into microfuge tubes is complete, additional water or buffer is added to each tube to provide sufficient volume for handling by pipette, and the tubes are spun down to ensure coalescence of the aqueous phase at the bottom of the tube. Sample can then be extracted for further processing by pipetting from the bottom of the tube.

Automated droplet elution was performed by mounting the device on a three-axis robot built from three interconnected precision stages T-LSM025A, T-LSR300D, T-LSR160D (Zaber). The device is vacuum-sealed to the lowering arm using a vacuum pump (Fig. S2B). LabView software was used to coordinate stage control and on-chip valve actuation to automate insertion of the

elution nozzle of the device into selected microfuge tubes during elution.

Calculation of Droplet Wetting Velocity. A droplet flowing down a channel filled with an immiscible fluid is separated from the channel walls by a thin lubricating film (2), the thickness of which is related to droplet velocity by the following equation:

$$b = 0.643r \left(\frac{3\mu U}{\gamma} \right)^{\frac{2}{3}} \quad [S1]$$

where b is the film thickness, r is half the height of the droplet which is approximately 5 μm in the channels of the microfluidic device, μ is the viscosity of the carrier phase which is 3.4 cP (viscosity of FC-40), U is the droplet velocity, and γ is the interfacial tension, which is 14 mJ/m² at the interface between the carrier fluid (FC-40 + 17% 1H,1H, 2H, 2H-perfluoroctanol) and aqueous phase.

If the film thickness is reduced to a critical value, instability arises in which intermolecular forces between the droplet and the channel surface cause the film to spontaneously rupture. This critical thickness is given by the following equation (3) *:

$$h_0 = \left(\frac{AR^2}{\xi_{\max}\gamma} \right)^{\frac{1}{4}} \quad [S2]$$

where h_0 is the critical film thickness, A is the Hamaker constant taken to be 10⁻²⁶ J between PDMS and water (4), R is the radius of the approximated disc of carrier fluid separating the droplet from the channel wall assumed to be 50 μm, and ξ_{\max} is a numerical constant associated with the most unstable mode of a perturbation to the uniform film. The smallest value of ξ_{\max} (approximately 1.7) is used as it leads to the greatest instability of the film (3). h_0 is found to be approximately 8 nm, and this value can be substituted into [Eq. S1] to find the droplet velocity at which the film thickness is reduced to the critical thickness and spontaneous wetting of the droplet to channel surfaces occurs. Solving for U yields the velocity at which droplets will wet channel surfaces. This is found to be approximately 170 μm/s and is in good agreement with experimental estimates made by analysis of videos to determine the lowest velocity achievable before droplets wet to channel walls (100 ± 50 μm/s).

Supplemental Note Regarding Droplet Immobilization. Droplet immobilization at the chamber entrance occurs through a wetting mechanism rather than surface tension forces that act to minimize interfacial energy between the droplet and carrier phase. This is because immobilization by surface tension requires a force opposing droplet motion, which can only occur if the motion of the droplet increases the surface area, and hence the surface energy of the droplet.

Formally, a change in restoring force due to surface tension, δF_s can be expressed as,

$$\delta F_s = \frac{-\partial E_s}{\partial x} \delta x = -\frac{\partial E_s}{\partial A} \frac{\partial A}{\partial x} \delta x$$

Where $E_s = A\Sigma$ is the total interfacial energy, Σ is the surface tension (J/m²), A is the droplet/carrier phase interfacial area,

*See equation 30 in ref. 3

and δx is an increment of displacement. For a positive surface tension ($\Sigma > 0$) this expression is negative only if $\partial A/\partial x > 0$, a condition that is only met when droplet motion causes it to enter a channel constriction. This is not the case in our system where droplets are immobilized after having already undergone an expansion into the cylindrical storage chamber from the inlet channel.

Effect of Surfactants. Few surfactants have been shown to be capable of stabilizing aqueous droplets in fluorocarbon oils (5, 6). Interfacial phenomena are, however, important during the formation of droplets. In the absence of a fluorosurfactant in the carrier fluid, we have observed that droplets wet the channel walls during injection, leading to the formation of satellite droplets and introducing potential for cross-contamination. The inclusion of a fluorosurfactant (17% 1H,1H, 2H, 2H-perfluoroctanol) in the carrier phase (7) was found to suppress wetting during droplet injection for all aqueous reagents tested. We note that this surfactant does not prevent the wetting of stationary droplets to untreated PDMS channels and does not stabilize droplets (5). In addition to surfactants in the carrier phase, it is often desirable to include a second aqueous surfactant to reduce the adsorption of analytes to channel walls or droplet surfaces. We observe that the inclusion of this surfactant partially stabilizes droplets, significantly increasing the time required for coalescence. Thus, when using aqueous surfactants, the robust merging of droplets may require that they be held in contact for an extended time. Aqueous surfactant also appears to enhance wetting of aqueous droplets onto PDMS surfaces (Fig. S12).

Storage Element Design Details. A storage element design with an inlet channel 520 μm in length, with 18 side channels ($30 \mu\text{m} \times 10 \mu\text{m} \times 5 \mu\text{m}$) along each side, was found to allow for deceleration of droplets to the velocity at which they wet the channel walls for incoming droplet velocities at the storage element inlet as high as 3.9 mm/s. We herein refer to this as the “critical velocity” and at this velocity the delivery of droplets to each element in the array takes on average 7 s. When droplets are sent to storage elements at or below the critical velocity, we routinely observe 100% coalescence in 500 events (10 droplets \times 50 chambers) both with and without surfactant in the aqueous phase (0.1% Tween 20). When operating above the critical velocity, droplets are not sufficiently decelerated by the side channels and enter the storage chamber without wetting the channel walls. In this case, the free droplets follow an upward trajectory determined by a combination of laminar flow and buoyancy, ultimately coming to rest at the chamber ceiling where they wet and are immobilized (Movie S4). Provided that the flow rate is constant, each incoming droplet is delivered to the same location and contacts the previously stored droplets. In the absence of a surfactant in the aqueous phase, the droplets coalesce shortly after making contact, thereby allowing for the sequential merging of droplets at maximal flow rates. However, when the droplet contents include a surfactant, the coalescence is delayed, leading to transient droplet contact and unreliable merging during the initial droplet additions. Once the total volume of droplets sent to a storage chamber is large enough to occupy a significant fraction of the chamber volume (approximately 25%), all droplets merge with the stored droplet. Thus, if the final stored droplet volume is sufficiently large and the sequence of droplet merging is unimportant, flow velocities much higher than the critical velocity can be used to achieve faster formulation (Movie S5). Operating in this regime, a storage chamber can be filled with 100 pump increments in approximately 5 s.

Finite Element Simulation. Simulation of fluid flow through the droplet storage element was performed using COMSOL v4.0a (COMSOL) (Fig. S1). A constant flow rate of 3.3 $\mu\text{L}/\text{s}$ was

set at inlet and outlet of the storage element and fluid properties of FC-40 were used.

Microscopy and Image Acquisition. Microfluidic devices were mounted onto a DMIRE2 fluorescent microscope (Leica) or an SMZ1500 stereoscope (Nikon) for imaging. Leica L5 and TX2 filter cubes were used to image green fluorescent protein (GFP) and red fluorescent protein (RFP) fluorescence respectively. Still images of the device were acquired using CCD cameras (Q imaging Retiga 4000R and Canon 50D). Videos were made using an IV-CCAM2 CCD camera (Industrial Vision Source). A confocal scanner (Wellscope, Biomedical Photometrics) was used to acquire confocal fluorescent scans of the device.

Reagents. A 5 : 1 mixture (v/v) of FC-40 or FC-72 (Sigma Aldrich) and 1H,1H, 2H, 2H-perfluoroctanol (Sigma Aldrich) was used as the carrier fluid.

Quasar 670 fluorescent dye was obtained from Biosearch Technologies.

PCR reactions on human gDNA template (Biochain) were performed using the RNase P FAM detection kit (Biorad) and Universal Fast PCR Mix (Biorad), which includes a passive 5-carboxy-X-rhodamine (ROX) fluorescent dye.

On-chip PCR reactions amplifying a fragment of the 16S gene in *Escherichia coli* and *Salmonella typhimurium* bacteria were performed using LC green intercalating dye (Idaho Technology Inc.), Itaq Supermix (Biorad), which includes a passive ROX fluorescent dye, and the following primers: 5'-TCGTGTTGTGA AATGTTGGT-3', 5'- TAAGGGCCATGATGACTTGAC-3' (500 nM each). On-chip PCR reactions amplifying a fragment of the 16S gene specific to *E. coli* were performed as above but with primer sequences from (8). All off-chip PCR reactions on bacterial DNA were performed using the same primers and primer concentrations as on-chip and iQ SYBR Green Supermix (Biorad).

All whole genome amplification (WGA) reactions were performed using the Picoplex WGA Kit for Single Cells (Rubicon Genomics). Additional amplification of eluted on-chip WGA product was performed using only the last PCR step of the full protocol as recommended by the manufacturer.

For all on-chip PCR and WGA experiments, all aqueous solutions were supplemented with 0.1% Tween 20 surfactant to avoid reagent adsorption onto PDMS surfaces of the reagent inlets and droplet interfaces.

Elution Test. Elution of storage chambers was tested by acquiring fluorescent images of a chamber loaded with 5 μM fluorescein-labeled 40-mer oligonucleotides before and after elution with water. A separate chamber filled with water was also imaged for comparison. The mean fluorescent intensity (measured by ImageJ software) of the water-filled chamber was subtracted from that of the eluted chamber and found to be 0.16% of the oligonucleotide-filled chamber before elution, indicating 99.84% sample recovery. Three other water-filled chambers were also imaged to measure the noise of our imaging measurement. The coefficient of variation was found to be 1.7%.

Bacterial Culture. For on-chip bacterial culture, *S. typhimurium* SL1344 transformed with plasmids encoding an ampicillin resistance gene and G/RFP were each first aerobically cultured in 2 mL of LB broth (Sigma Aldrich) with 100 $\mu\text{g}/\text{mL}$ ampicillin for approximately 18 hrs at 37 °C to reach stationary growth phase (approximately 10^9 cells/mL). For each strain, 2 mL of fresh culture media was then inoculated with 6 μL of cell culture and incubated for another 2 h to produce exponential growth phase cultures, which were mixed in a 1 : 1 ratio and diluted using media to a concentration of approximately 1 cell/10 nL for single-cell sorting. The concentrations of the stationary growth phase cul-

tures of both strains were measured by absorbance (OD 600) to be equivalent, and 10x dilutions of these cultures in media were then used to seed on-chip cultures starting with multiple cells. Suspensions of K12 *E. coli* bacteria (ATCC 10798) were cultured and prepared as above, but without ampicillin in the media, and were stained with SYTO 9 DNA stain (Invitrogen) to aid in visualization prior to use of on-chip. For on-chip PCR and WGA experiments, bacterial cultures were resuspended three times in PBS to remove free DNA from the suspension fluid prior to using on-chip.

PCR Protocol. On-chip qPCR was performed using a prototype version of the Biomark microfluidic qPCR instrument (Fluidigm), consisting of a flatbed thermocycler equipped with a CCD camera, fluorescent illumination, and filters. Off-chip qPCR was performed using a Chromo 4 thermocycler (Biorad) and data was analyzed using Opticon Monitor 3 software (Biorad). The thermocycling protocol for RNase P PCR consisted of an initial hotstart at 95 °C for 20 s, followed by 40 cycles of 95 °C for 1 s and 60 °C for 30 s. The thermocycling protocol for PCR reactions on all bacteria consisted of an initial hotstart at 95 °C for 3 min which was also used to lyse cells, followed by 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s.

Image Analysis Software. To measure mean fluorescence intensity of stored droplets containing formulated fluorescent dye concentrations, the fluorescent confocal image of the droplet array was manually analyzed using ImageJ software. Linear fitting of the data was performed using the curve fitting toolbox in MATLAB (Mathworks).

All custom image analysis software described below was written in MATLAB and used functions from the Image Processing Toolbox.

Custom software was written to analyze all on-chip qPCR images. For each cycle, droplets were first segmented using the passive ROX dye images. This dye was included in the PCR reaction mix for all on-chip reactions. Segmentation after each cycle is necessary since the high temperatures that the chip is heated to during PCR cause the positions of the droplets to shift slightly in the storage chambers from cycle to cycle. A pixelwise division of the FAM probe or LC green intercalating dye image by the passive ROX dye image was used to normalize data for variations in illumination across the droplet array and to account for increase in signal due to evaporation. For each droplet, an amplification curve was generated by subtracting the median normalized pixel intensity for each cycle from that of the first cycle, and removing linear components extracted from the pre-exponential phase. Manual thresholding of the amplification curves in the exponential phase was performed to determine the cycle threshold (CT) of each droplet. For the RNase P qPCR experiment, any reactions with a CT greater than 2 standard deviations above the mean CT corresponding to a single molecule were determined to be nonspecific amplifications and were classified as not detected. PCR efficiency was calculated as $-1 + 10^{-1/x}$, where x is the slope of CT vs. $\log_{10}(\text{quantity})$ standard curve.

Custom software was written to analyze fluorescent images acquired from on-chip culture of GFP and RFP-expressing bacteria. As the culture media used was slightly fluorescent in the GFP channel, the first GFP image was used to segment each droplet. The boundary of each droplet was then slightly dilated to generate a new boundary, which was used to identify droplets in all subsequent images. Because the incubation of the chip was performed at a relatively low temperature (25 °C), the droplets did not shift position significantly during the time interval between image acquisitions and this method was able to identify all droplets for all images. To generate a growth curve for each stored droplet, fluorescence intensity was first integrated over each droplet for each image in both GFP and RFP channels and a

moving average filter with a window width of 3 was applied to all datapoints between the third and final images for each droplet to remove noise. When comparing endpoint GFP and RFP fluorescence in each two-strain co-culture, normalization was performed by dividing the integrated fluorescence intensity in each channel from the final image by that of the culture with the highest endpoint integrated fluorescence intensity in each group of co-cultures seeded with the same number of cells.

Preparation of Environmental Samples. Environment 1: Seawater enrichment culture. A modular medium buffered with 10 mM MOPS pH 7.2 containing 1X seawater base amended with salts (10 mM NH4Cl, 1.5 mM KH2PO4), cofactors (12X Vitamins, B12, and trace elements) was used. The carbon source supplied to the medium was 2 mM Na2HCO3. Electron donors included 8 mM Na2S or 30 mM Na2SO3 with 20 mM NaNO3 as alternative electron acceptor. Primary enrichment cultures were generated at 26 °C using modular medium mixed with varying dilutions of Saanich Inlet 135 meter inoculum. 10 µL of the enrichment culture was diluted in 1 mL of PBS prior to on-chip use.

Environment 2: Marine sediments. Aggregates of microbial cells were extracted from marine sediments collected at a depth of 836 m in the Santa Monica Basin (LAT: 33°47'99''N, LONG: 118°38'83''W). Once thawed from their storage conditions (-80 °C), 10 g (wet weight) of a core's 3–6 cm (bsf) layer were suspended in 14 mL of 1X PBS. The slurry was vortexed and sonicated on ice at 180 watts for 20 s with a Sonicator Ultrasonic Processor XL 2020. Percoll density gradients were generated by centrifuging 30 mL of percoll/PBS 1X solution (1:1 vol/vol) at 38,325 × g for 30 min (4 °C). The entire sediment slurry was gently dispensed on top of 8 percoll gradients (approximately 3 mL per tube), which were then centrifuged at 10,300 × g for 15 min (4 °C). A cell suspension was gathered by removing 20 mL from the top of each percoll gradient, and cell aggregates ranging from 3–8 µm in diameter were enriched by filtration through two 8-µm pore-size polycarbonate membranes and onto a 3-µm pore-size membrane. Cells were collected from the 3-µm pore-size membrane with 1 mL of PBS 1X/ethanol (1:1 vol/vol), and stored at -20 °C until processed. Cells were washed and further enriched prior to sorting on-chip by centrifuging 200 µL of the PBS/ethanol suspension at 5,900 × g for 5 min (4 °C). The supernatant was removed and the pellet resuspended in 200 µL of 1X PBS, before the suspension was again centrifuged at 5,900 × g for 10 min (4 °C). The final pellet was resuspended in 10 µL 1X PBS prior to use on-chip.

Environment 3: Human oral swab. The oral biofilm sample was obtained from the mouth of a 31-year-old male by scraping a tooth with a sterile pipette tip and resuspending the accumulated biofilm in 1 mL of PBS for on-chip use.

Whole Genome Amplification (WGA). Thermocycling steps for on-chip WGA was performed by placing the device on a flatbed thermocycler and taping the device to the heating surface to ensure good thermal contact. Recommended thermocycling protocols were used. On-chip WGA reaction product was eluted into 30 or 40 µL of water. To quantify on-chip WGA-amplified *E. coli* DNA, 2 µL of the eluted sample was used in an off-chip qPCR reaction using the K12 *E. coli*-specific 16S gene qPCR assay from (8). CT values were compared to those from a standard curve generated from qPCR reactions on dilutions of purified *E. coli* gDNA (ATCC) with known 16S gene copy number (7 per genome). For on-chip WGA reaction product submitted to a second round of WGA, 5 µL of the eluted sample was used as a template in the second round of WGA.

DNA Sequencing. 16S gene PCR amplicon from on-chip single-cell reactions to be sequenced was first eluted and diluted into 20 μ L of water, 2 μ L of which was used as a template in a second off-chip PCR reaction to increase DNA mass for sequencing. This amplicon was then run on an agarose gel, the band was cut out, and DNA was extracted using a Qiagen Qiaquick Gel Extraction kit. DNA was then sequenced using an Applied Biosystems 3730S 48-capillary DNA Analyzer with POP-7 BigDye Terminator v3.1 sequencing chemistry. Sequencing data was analyzed using CLC Bio Main Workbench software. The expected sequences of the fragment amplified by the 16S gene assay in *E. Coli* and *S. typhimurium* (respectively) are:

TCGTGTGTGAAATGTTGGGTTAAGTCCCGAACGAG
CGCAACCTTATCCTTGTGCCAGCGGTCGGCCGGGA
ACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGT
GGGGATGACGTCAAGTCATCATGGCCCTTA

and

TCGTGTGTGAAATGTTGGGTTAAGTCCCGAACGAG
CGCAACCTTATCCTTGTGCCAGCGGTTAGGCCGGGA
ACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGT
GGGGATGACGTCAAGTCATCATGGCCCTTA

Both amplicons are 144 bp long with mismatches between the 2 sequences at positions 51, 67, 68, and 87.

All sequencing of WGA product was performed on an Illumina Genome Analyzer IIx after library preparation performed according to recommended protocols. *E. coli* samples were sequenced using 75 and 50 bp paired-end reads for one and two rounds of WGA amplification, respectively. All environmental sample data was sequenced using 75 bp paired-end reads.

Sequencing Data Analysis. *E.coli* sequence data was aligned to NCBI reference genome NC_000913 (*E.coli* substrate MG1655) by the BWA program (9).

For environmental samples, raw Illumina reads (65,067,118 paired-end reads, 75 bp in length) were first demultiplexed into 74 fastq files. Each set of reads was trimmed for low quality from both ends and was assembled using Velvet (10) (also see <http://www.ebi.ac.uk/~zerbino/velvet/>) at a range of Kmers (51,55,59,63,67,71) using all the trimmed reads and the default Velvet settings (flags: -exp_cov auto). Assemblies were performed for each sample and contigs greater than 200 bp in length were used for further analysis. Contigs generated by each assembly (six total contig sets), were merged using a combination of in-house Perl scripts. Contigs were then sorted into two pools based on length. In an effort to join minimally overlapping contigs, those smaller than 1,800 bp were assembled together using Newbler (11) in an attempt to generate larger contigs (flags: -tr, -rip, -mi 98, -ml 80). All assembled contigs larger than 1,800 bp, as well as the contigs generated from the final Newbler run, were combined together using minimus2 (12) (also see <http://sourceforge.net/projects/amos>) using an overlap length of 80 bp, a minimum overlap identity cutoff of 98%, and a consensus error of 0.06 for merging contigs. The BWA program was used to map reads back to the final contigs for contig verification and in order to establish contig fold coverage and percent read assembly statistics. A summary of results from sequence assemblies is presented in Table S2.

The number of contigs for each sample varied between environments with ENV1 assemblies yielding the highest average number per sample (mean of 1,998 contigs covering 70% of reads), followed by ENV3 (mean of 659 covering 76% of reads) and ENV2 (mean of 431 contigs covering 70% of reads) (Table S2). This correlated with contig length differences between samples with mean contig lengths of 471, 324, and 424 bp for

ENV1, ENV2, and ENV3, respectively. We note that individual assemblies were limited by sequencing depth and that the higher number of contigs in ENV1 is likely due to reduced sample complexity. NTCs resulted in 7–20 contigs per sample, which covered less than 30% of reads.

Sequence data for all samples was deposited in the NCBI Sequence Read Archive (SRA051477) and is a part of BioProject (PRJNA157923).

Metagenomic Characterization. The taxonomic evenness of each sequenced WGA sample was evaluated by plotting GC content distribution using a kernel density (Gaussian, Fig. S6–S8) as implemented by R package lattice. Data from all metagenomes of each of the three environments was summarized in Fig. 5A.

The taxonomic profile of each metagenomic dataset was first evaluated using a phylogenomic approach, as implemented by MLTreeMap (13). Briefly, this procedure is based on the detection and phylogenetic identification of the 40 most taxonomically informative genes (i.e., COGs). For each metagenome, the overall taxonomic identity of all instances of these genes was compiled using the probabilistic distribution of query sequences among the reference sequences of the 40 COG summary “tree of life.” Reference sequences with less than 5% of the query sequence assignment probabilities were removed for noise reduction, and taxonomic profiles were compiled by summing the taxonomy of all remaining reference sequences at the phylum or subphylum level and normalizing against the total number of assigned genes.

Taxonomic profiles were then further defined using a direct sequence comparison approach. Contigs assembled from each metagenome were searched against the database eggNOG (14) (also see <http://eggnog.embl.de/>) and against NCBI’s RefSeq proteomic database (<http://www.ncbi.nlm.nih.gov/RefSeq/>) using blastx with an E value cutoff of 1e-8. For the search against eggNOG, contigs were taxonomically assigned to the phylum or sub-phylum of the reference sequence producing the best alignment (as determined by bitscores). Dataset-specific taxonomic profiles were then compiled by normalizing against the total number of assigned contigs. For the search against RefSeq, contigs were taxonomically assigned from the blast outputs using MEGAN (15) (also see <http://ab.inf.uni-tuebingen.de/software/megan/>) with default settings. Uninformative categories of MEGAN taxonomic assignment (e.g., not hits, not assigned, unicellular organisms) were removed, and dataset-specific taxonomic profiles were again compiled by normalizing against the total number of assigned contigs.

Information gathered using MLTreeMap, eggNOG, and RefSeq is presented in Fig. S9–S11 for all metagenomes, and for selected datasets from each environment in Fig. 5C.

To compare the taxonomic profiles characterizing the 69 WGA samples of all three environments (no NTCs), the proportional representation of all informative taxonomic categories identified by MEGAN in the blast outputs against RefSeq were compiled in a matrix (Table S3, samples 21 and 93 were removed due to lack of sufficient information). Samples were then submitted to a hierarchical cluster analysis using R package hclust (Fig. 5B).

As mentioned in the main text, two of the 72 submitted samples failed to produce libraries that passed QC. These two libraries were prepared from samples 69 and 70, which were expected to be no cell controls for ENV1. Neither of these was found to have any significant similarity to single-cell samples from ENV1, but did produce significant alignment to unrelated taxa. We believe that these samples were either contaminated during library preparation or mislabeled and thus excluded them from any subsequent analysis.

1. Unger MA, Chou H-P, Thorsen T, Scherer A, Quake SR (2000) Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* 288:113–116.
2. Bretherton FP (1961) The motion of long bubbles in tubes. *J Fluid Mech* 10:166–188.

3. Baldessari F, Homsy GM, Leal LG (2007) Linear stability of a draining film squeezed between two approaching droplets. *J Colloid Interface Sci* 307:188–202.
4. Steinhaus B, Spicer PT, Shen AQ (2006) Droplet size effects on film drainage between droplet and substrate. *Langmuir* 22:5308–5313.

- PNAS
5. Shim J-u, et al. (2007) Control and measurement of the phase behavior of aqueous solutions using microfluidics. *J Am Chem Soc* 129:8825–8835.
 6. Holtze C, et al. (2008) Biocompatible surfactants for water-in-fluorocarbon emulsions. *Lab Chip* 8:1632–1639.
 7. Tice JD, Song H, Lyon AD, Ismagilov RF (2003) Formation of droplets and mixing in multiphase microfluidics at low values of the Reynolds and the capillary numbers. *Langmuir* 19:9127–9133.
 8. Lee C, Lee S, Shin S, Hwang S (2008) Real-time PCR determination of rRNA gene copy number: absolute and relative quantification assays with *Escherichia coli*. *Appl Microbiol Biotechnol* 78:371–376.
 9. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
 10. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.
 11. Chaisson MJ, Pevzner PA (2008) Short read fragment assembly of bacterial genomes. *Genome Res* 18:324–330.
 12. Sommer D, Delcher A, Salzberg S, Pop M (2007) Minimus: A fast, lightweight genome assembler. *BMC Bioinformatics* 8:64.
 13. Stark M, Berger S, Stamatakis A, von Mering C (2010) MLTreeMap—accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* 11:461.
 14. Muller J, et al. (2009) eggNOG v2.0: Extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* 38:D190–D195.
 15. Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17:377–386.

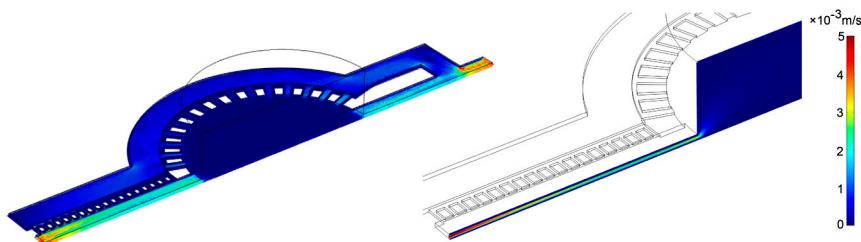


Fig. S1. Finite element simulation of the flow velocity through storage element at a height of 2.5 μm (half of height of side channels) (Left) and on the vertical plane through the center of the storage element (Right).

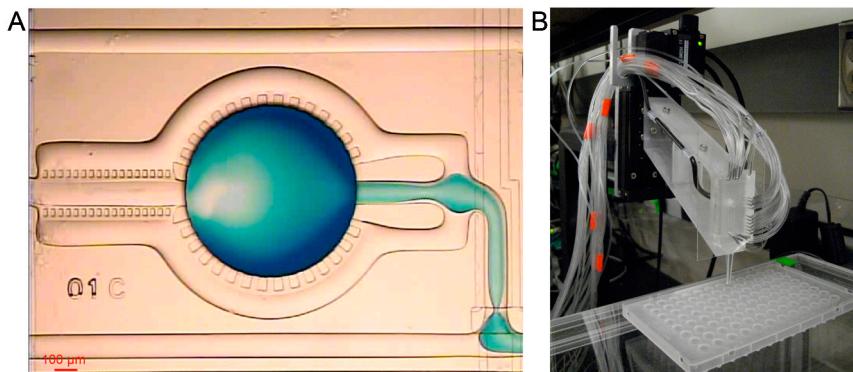


Fig. S2. Automated elution of individual storage elements. (A) Optical micrograph showing an oil-sheathed stream of water flowing into a storage element to perform elution. Separation of the aqueous phase (containing blue dye) from the channel walls is visible at the storage element outlet. (B) During elution, the device is mounted to a 3-axis robotic chip-holder to allow for computer-controlled positioning of the elution nozzle into microfuge tubes.

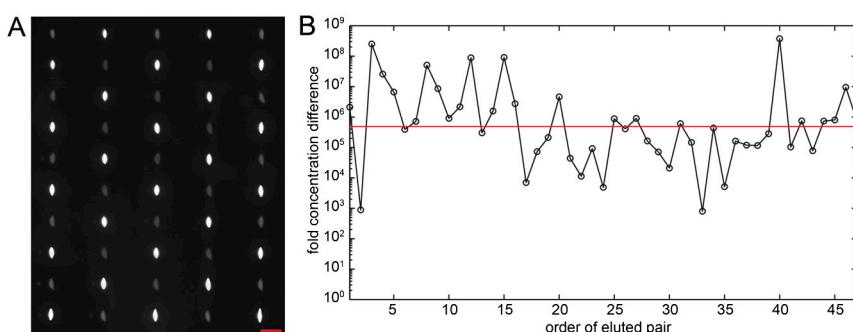


Fig. S3. Low cross-contamination during formulation and elution of stored droplets. (A) Endpoint fluorescent image following 40 cycles of PCR of chambers loaded with either template (1,476 genome copies per droplet) or buffer (NTC) in a checkerboard pattern. The image indicates no detectable cross-contamination between positive (white) and NTC (dark gray) droplets during loading of storage array. Scale bar, 1 mm. (B) Fold concentration difference of template in eluted pairs of droplets containing amplified template and water. Red line denotes mean.

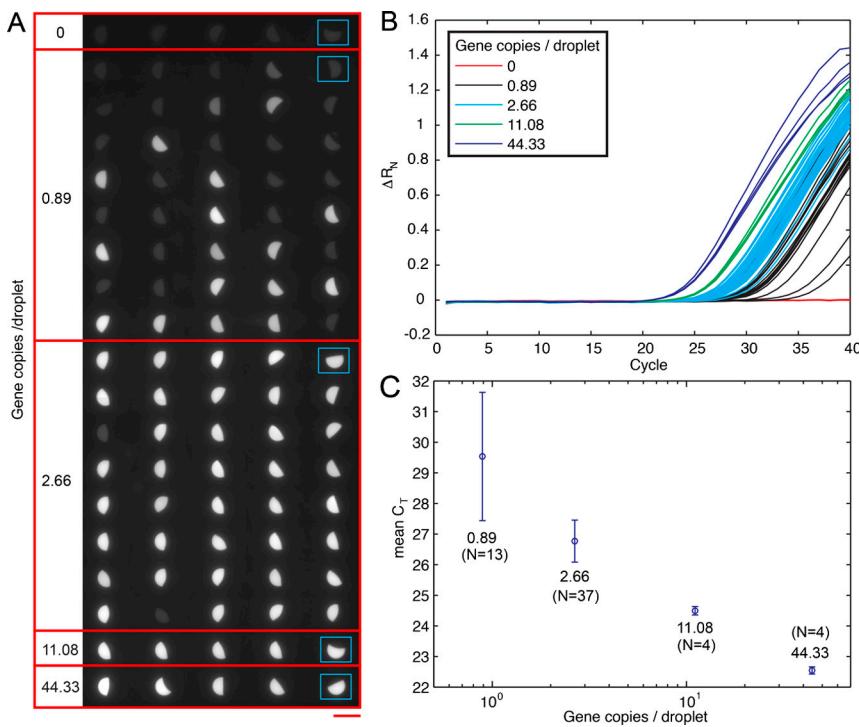


Fig. S4. Efficient on-chip qPCR with single molecule sensitivity. We established the sensitivity and efficiency of on-chip qPCR in stored droplets using serial dilutions of human genomic DNA. Ninety template dilutions were formulated on-chip in which 100 pump increments (approximately 13.3 nL) of either human genomic DNA or water were mixed to achieve concentrations of 44.33 (133 pg, $N = 4$), 11.08 (33.25 pg, $N = 4$), 2.66 (7.98 pg, $N = 39$), 0.89 (2.66 pg, $N = 39$), and 0 ($N = 4$) haploid genome copies per reaction. PCR reactions were then assembled from these dilutions by dispensing equal volumes of PCR master mix to each reaction, including primers and a hydrolysis probe designed for the detection of the RNase P gene, which is found at a single copy per haploid genome. Following reaction assembly, the device was thermocycled on a microfluidic qPCR instrument and fluorescent images were acquired at each cycle. (A) Endpoint fluorescent image of the chamber array following 40 cycles of PCR. Blue rectangles denote control reactions mixed off-chip. Scale bar, 1 mm. At the two lowest dilutions we observed digital patterns of amplification, with 37/39 and 18/39 positive reactions for expected concentrations of 2.66 and 0.89 genome copies per chamber, respectively. These frequencies fall within symmetric 95% binomial confidence intervals constructed around the expected concentrations: 31/39 to 38/39 for 2.66 copies per chamber, and 16/32 to 29/32 for 0.89 copies per chamber. (B) qPCR curves for each stored droplet. (C) Mean CT values from B at each template dilution. Error bars represent the standard deviation across all replicates. CT values for the two highest dilutions were 22.56 ($SD = 0.12$) and 24.49 ($SD = 0.14$) respectively, corresponding to an absolute precision in concentration measurement of 9.5%, which is near the limit of qPCR. The difference in CT corresponding to a fourfold dilution was found to be 1.93 ± 0.18 , indicating a PCR efficiency of 105%.

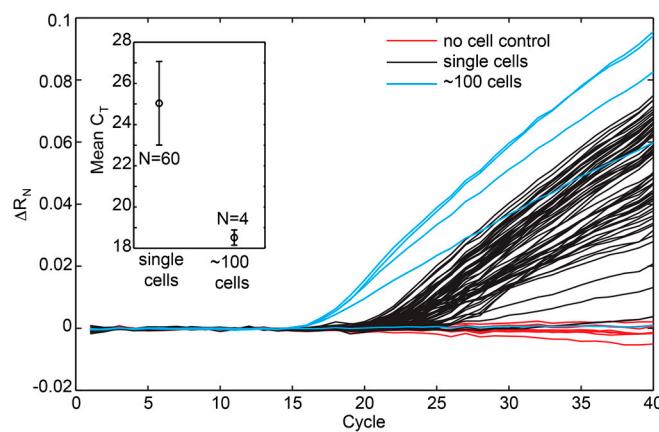


Fig. S5. qPCR on single sorted *E. coli*. qPCR curves and mean CT values (Inset) for single and multiple-cell reactions. Error bars represent standard deviation across all replicates.

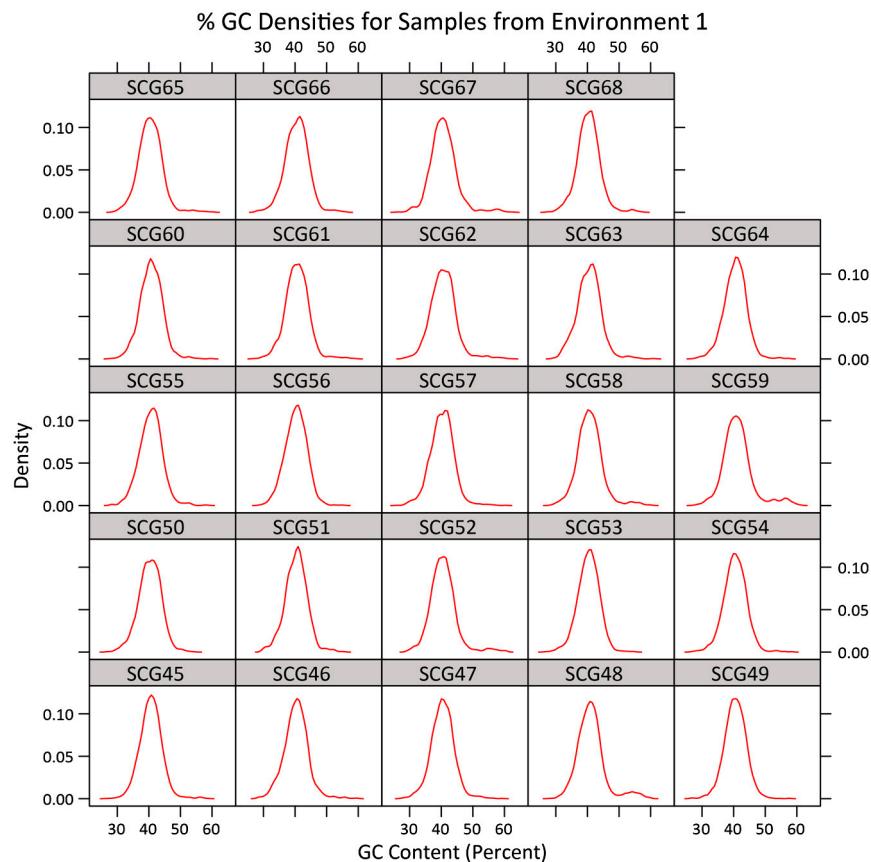


Fig. S6. Kernel density plots of each metagenome associated with cells sorted from environment 1 (seawater enrichment culture).

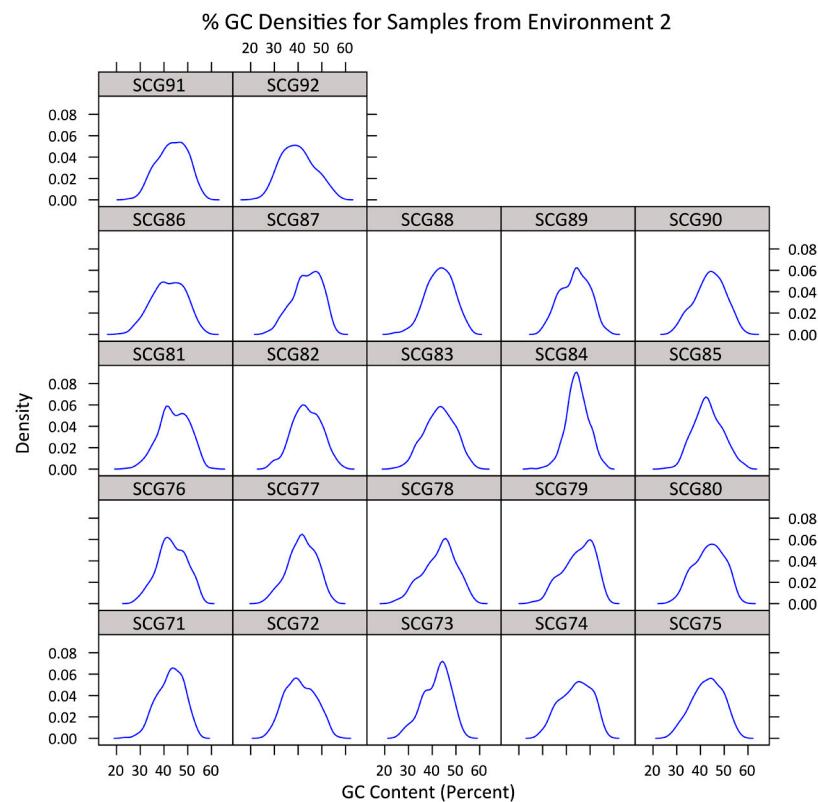


Fig. S7. Kernel density plots of each metagenome associated with cells sorted from environment 2 (marine sediments).

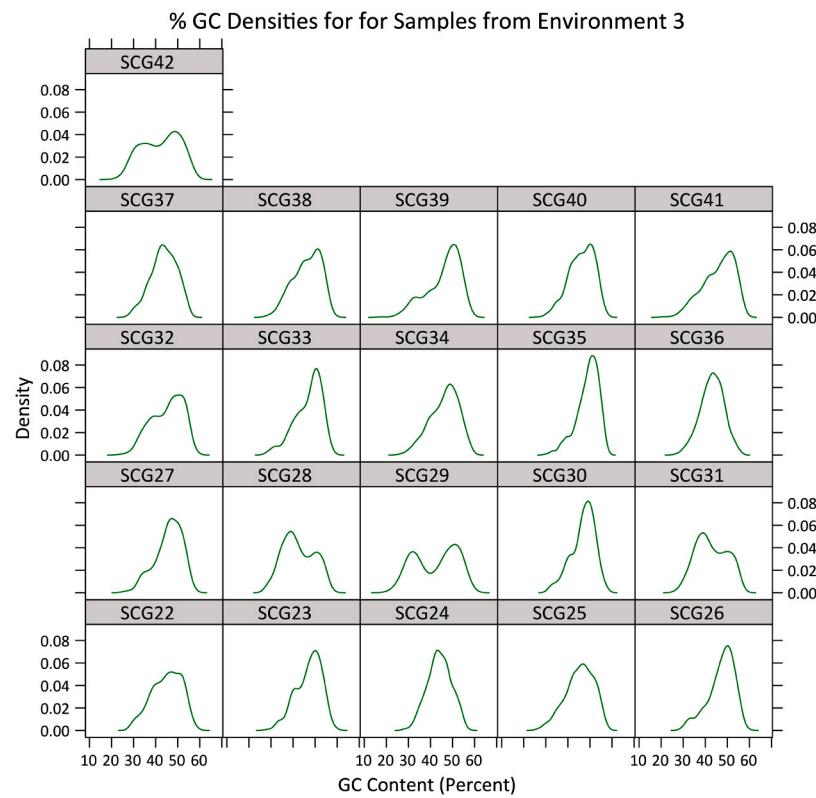


Fig. S8. Kernel density plots of each metagenome associated with cells sorted from environment 3 (human oral swab).

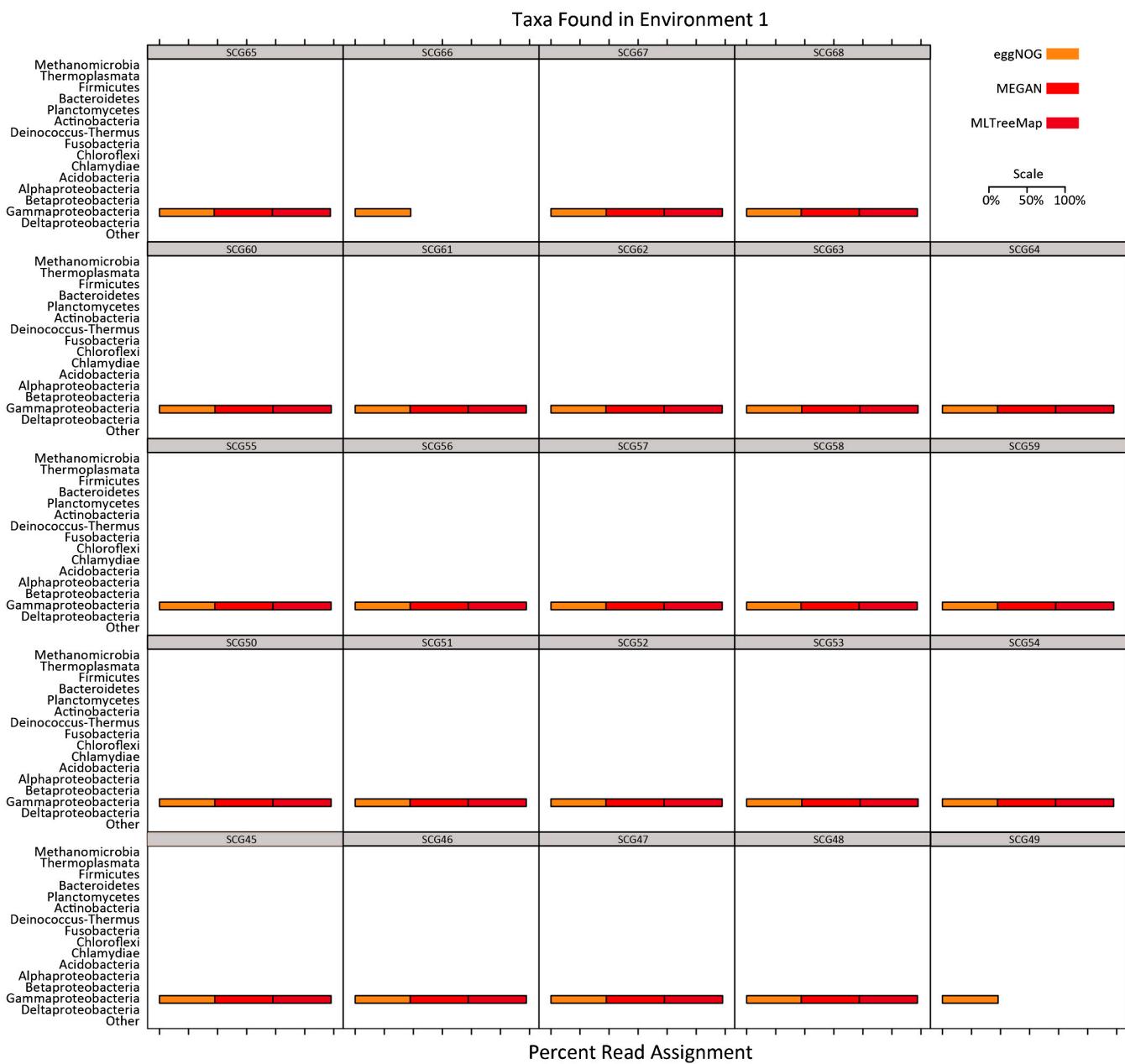


Fig. S9. Taxonomic assignment of metagenomic data associated with cells sorted from environment 1 (seawater enrichment culture), as determined by comparison to databases eggNOG and RefSeq (MEGAN), and by MLTreeMap searches. Representations of taxonomic groups were normalized per sample.

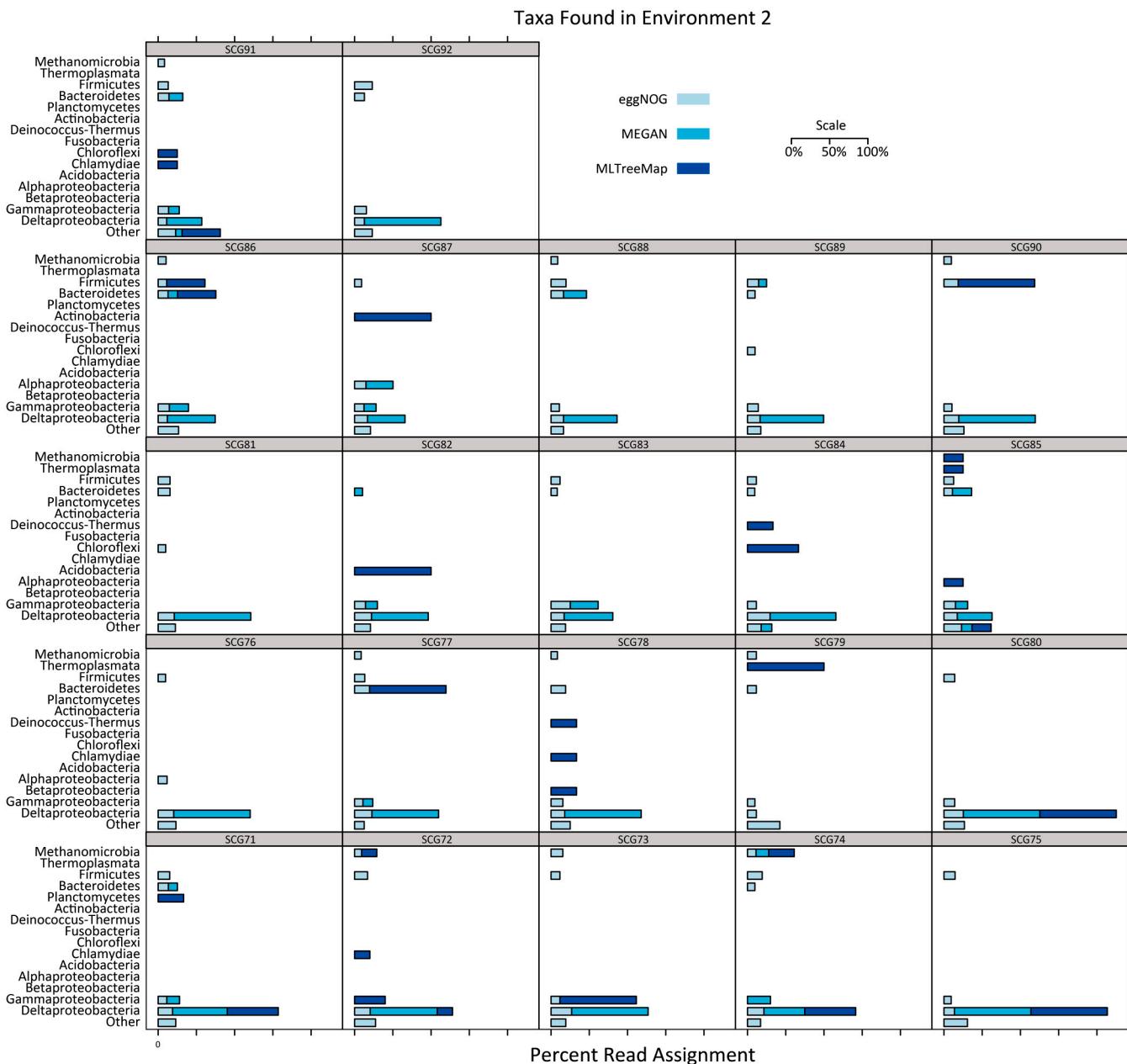


Fig. S10. Taxonomic assignment of metagenomic data associated with cells sorted from environment 2 (marine sediments), as determined by comparison to databases eggNOG and RefSeq (MEGAN), and by MLTreeMap searches. Representations of taxonomic groups were normalized per sample.

Taxa Found in Environment 3



Fig. S11. Taxonomic assignment of metagenomic data associated with cells sorted from environment 3 (human oral swab), as determined by comparison to databases eggNOG and RefSeq (MEGAN), and by MLTreeMap searches. Representations of taxonomic groups were normalized per sample.

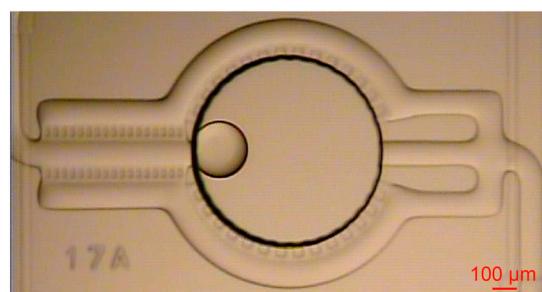
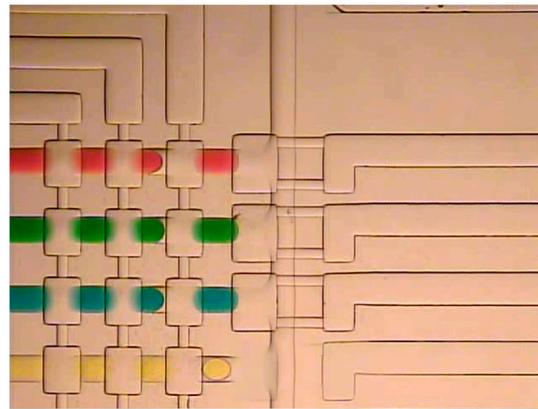
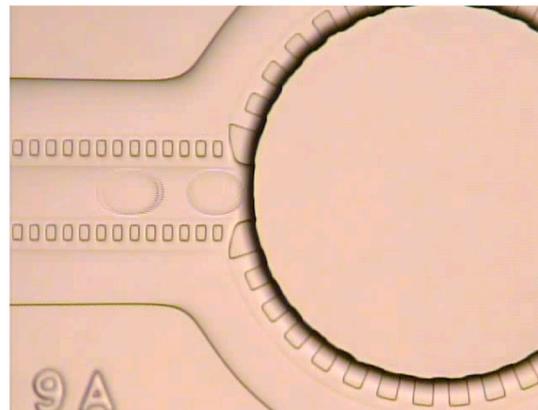


Fig. S12. Stored droplet containing surfactant. The presence of a surfactant in the aqueous phase (0.1% Tween 20) enhances wetting onto the device surface, resulting in an apparent contact angle of approximately 130°.



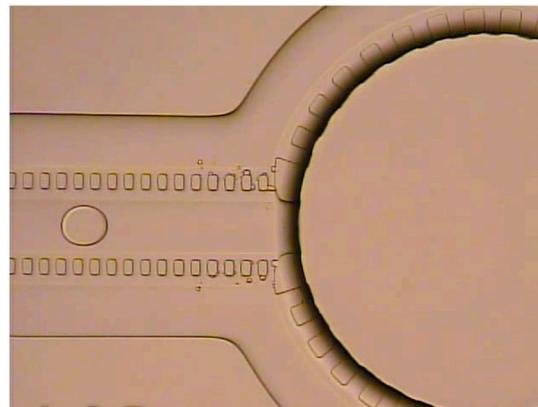
Movie S1. Droplets of food dye being metered from the metering module, with 82.7 kPa applied to the aqueous inlets and a carrier fluid flow rate of 0.24 $\mu\text{L}/\text{min}$.

[Movie S1 \(MOV\)](#)



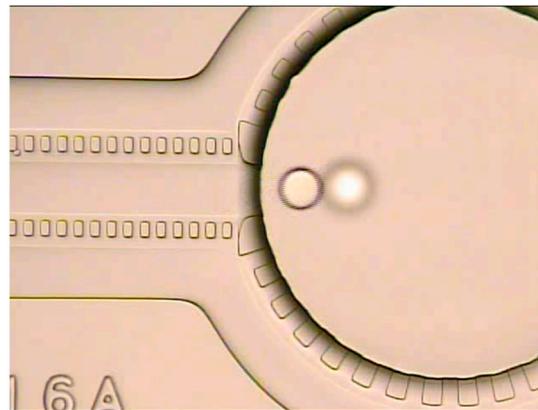
Movie S2. Water droplets entering the storage element at the critical velocity. Mean flow velocity at the storage element inlet is 3.9 mm/s.

[Movie S2 \(MOV\)](#)



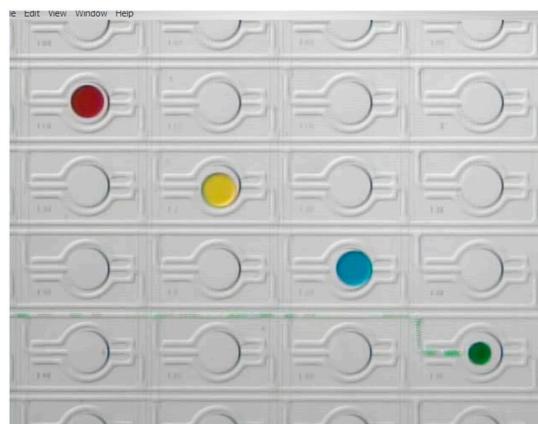
Movie S3. Water droplets entering the storage element below the critical velocity. Mean flow velocity at the storage element inlet is 2.9 mm/s. Multiple droplets merge before the merged droplet reaches the edge of the cylindrical storage chamber and is moved into it by surface tension.

[Movie S3 \(MOV\)](#)



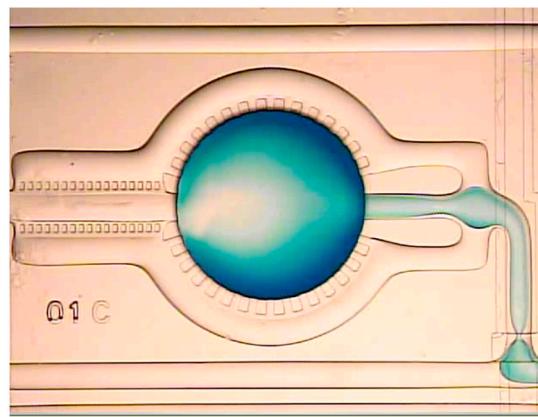
Movie S4. Water droplets entering the storage element above the critical velocity. Mean flow velocity at the storage element inlet is 7.2 mm/s. At the end of the movie, focus is shifted up by approximately 100 μ m, bringing the stored droplet into focus to show that it is positioned on the roof of the cylindrical storage chamber.

[Movie S4 \(MOV\)](#)



Movie S5. Loading of four storage elements above the critical velocity with 100 pump increments of food dye each. Mean flow velocity at the storage element inlet is 31.8 mm/s.

[Movie S5 \(MOV\)](#)



Movie S6. Elution of a storage element filled with food dye. Mean flow velocity at the storage element inlet is 30.1 mm/s.

[Movie S6 \(MOV\)](#)

Table S1. Statistics for single *E. coli* sequencing

	On-chip WGA				On-chip and off-chip WGA			
	75 bp reads	% of reads aligned	% of genome covered at $\geq 1 \times$	% of genome covered at $\geq 10 \times$	50 bp reads	% of reads aligned	% of genome covered at $\geq 1 \times$	% of genome covered at $\geq 10 \times$
No-cell control 1	5,663,384	1.0	7.7	0.1	4,639,808	1.1	5.3	0.1
No-cell control 2	4,341,480	0.2	8.7	0.1	6,677,002	0.1	5.4	0.0
Single cell 1	13,483,184	55.3	64.6	43.1	7,458,890	79.8	62.8	40.0
Single cell 2	9,784,130	47.2	40.6	22.5	6,941,568	80.2	41.9	22.9
Single cell 3	4,708,954	5.7	15.2	4.1	3,304,200	52.2	24.5	8.5
Single cell 4	5,738,682	1.0	18.6	1.6	7,354,470	68.4	42.3	24.8
Single cell 5	10,268,078	48.4	28.2	13.8	6,085,898	56.6	26.3	12.1
Single cell 6	10,644,760	34.5	30.3	14.0	5,373,842	42.6	27.5	12.0
Approximately 1,000 cells					8,074,402	78.8	61.6	34.5
Unamplified gDNA					62,513,866	90.9	99.8	99.7

Other Supporting Information Files**Table S2 (XLSX)**

Source of environmental cells and summary of results obtained from the sequencing and assembly of individual DNA samples.

Table S3 (XLSX)

Taxonomic profiles generated through blastp comparisons of ORF sequences to RefSeq amino acid sequences and analyzed

using MEGAN. Representations of taxonomic groups were normalized per sample and uninformative taxonomic categories were removed.