SUPPLEMENTARY NOTE:

Dynamic range of in-gel digital assays

When the input for digital MDA is very low (fewer than 100 per field), molecular counts are significantly inflated by contaminating fluorescence signals (contaminating DNA fragments or particles) that are not differentiated from true counts by our image analysis algorithm. At high target concentrations, the DNA clusters crowd one another, limiting the maximum useful concentration to 10,000 DNA molecules per field of view. For MDA, the smaller cluster sizes observed at higher template concentration (**Supplementary Fig. 4**) benefit assay dynamic range by improving cluster identifiability at the highest template concentrations. Based on our estimate of 10 pg DNA per cluster (**Supplementary Fig. 2**), 10,000 - 100,000 clusters per field of view in our setup approximates typical maximum product concentrations of about 800 ng/μL achieved in conventional liquid MDA reactions. The assay dynamic range can be improved by manipulating the DNA cluster size, increasing the volume of gel imaged (*eg* by combining multiple fields of view), improving image processing methods, and further reducing the number of fluorescent contaminants.

Analysis of reaction extent limitation and local competition among MDA clusters

We noticed smaller average cluster size at higher template concentrations (**Supplementary Fig. 4**). Based on additional studies, we concluded that a global auto-inhibition mechanism limits the growth of MDA clusters (**Supplementary Fig. 5**). We analyzed variability in cluster number and DNA content around large and small reference clusters to test for local reagent competition among WGA reaction centers, finding little evidence for local competition. This observation is consistent with the high diffusion constants for enzymes, primers, and nucleotides measured in PEG hydrogels similar to ours[1]. No specific limiting reagent was identified when reactants were supplemented individually (data not shown). The final reaction pH in our hydrogel reactions was measured to be 6.5 (initial pH = 7.5), which may limit cluster growth due to global loss of polymerase activity. Altogether, these data are consistent with density-dependent average size variation by global auto-inhibition, possibly by pH drop. Variability of cluster size in a single experiment may result from variable initial template conformation, degree of template denaturation, or local inhomogeneities in the hydrogel structure.

Random dispersion model for lab-cultured bacteria

The expected number of punches that have both *E. coli* and *S. aureus* is (based on qPCR analysis of the 80 punches):

$$P_{E.coli} = \frac{14}{80} = 0.175$$

$$P_{S.aureus} = \frac{43}{80} = 0.538$$

$$P_{negative} = \frac{30}{80} = 0.375$$

$$< N_{both\ E.coli\ and\ S.aureus} > = \frac{14}{80} \times \frac{43}{80} \times 80 = 7.52$$

This result is in-line with our qPCR result of seven double positive punches (**Supplementary Table 2**), indicating the likelihood that the distributions of *E. coli* and *S. aureus* across the punch samples are independent as we expected. Furthermore, if we assume a random (Poisson) distribution of microbes in the hydrogel:

$$P_{E.coli}(0,\lambda_E) = e^{-\lambda_E} = \frac{66}{80} = 0.825, \quad \lambda_E = 0.192$$
$$P_{E.coli}(1,\lambda_E) = \lambda_E e^{-\lambda_E} = 0.158$$
$$P_{E.coli}(2,\lambda_E) = \frac{{\lambda_E}^2 e^{-\lambda_E}}{2} = 0.01$$
$$P_{E.coli}\,(Single\ Cell) = \frac{0.158}{1-0.825} = 90.3\%$$

$$P_{S.aureus}(0,\lambda_S) = e^{-\lambda_S} = \frac{37}{80} = 0.462, \quad \lambda_S = 0.772$$
$$P_{S.aureus}(1,\lambda_S) = \lambda_S e^{-\lambda_S} = 0.356$$
$$P_{S.aureus}(2,\lambda_S) = \frac{{\lambda_S}^2 e^{-\lambda_S}}{2} = 0.138$$
$$P_{S.aureus}\,(Single\ Cell) = \frac{0.356}{1-0.462} = 66.2\%$$

The low probability value for the occurrence of single *S. aureus* is calculated based on the high number of hydrogel punches that were identified as *S. aureus* by qPCR. To bring down the value, a more dilute sample of *S. aureus* should be used.

	$P_{E.\ coli}(0) = 0.825$	$P_{E.\ coli}(1) = 0.158$	$P_{E.\ coli}(2) = 0.01$
$P_{S.\ aureus}(0) = 0.463$	0.38	0.073	0.0046
$P_{S.\ aureus}(1) = 0.356$	0.29	0.056	0.0036
$P_{S.\ aureus}(2) = 0.138$	0.114	0.022	0.0014

Genome coverage completeness estimation

Note that some studies in the field report data from quality-filtered ('cherry-picked') cells, which dramatically improves quality statistics such as average coverage. In this study, we report data on complete sets of single-cell MDA reactions.

Potential for amplification bias reduction

Up to 10 pg of DNA is produced by MDA from each template in the hydrogel format using our protocol. Although we re-amplified punch samples in this study to microgram quantities, 10 pg is, in principle, enough product (of order 1000 bacterial genome equivalents) to support deep sequencing directly. Given that we obtain good coverage distribution with our high-yield re-amplification protocol, there seems to be substantial potential for coverage distribution improvement by direct library construction from the 10 pg hydrogel product. Notably, the trendline in the bias/fold amplification curve (**Fig. 2d**) intersects with the non-amplified gDNA value at approximately 1000-fold amplification, promising much improved amplification

uniformity from such limited fold-amplification. Recent advances in ultra-efficient library construction allows library construction from sub-nanogram input levels[2].

Although a number of modified protocols have been proposed to improve coverage distribution in MDA, none has yet been widely adopted, with major single-cell genomics centers continuing to use phi29 DNA polymerase reaction conditions very similar to those originally developed 30 years ago[3]. By contrast, limiting fold-amplification is mathematically guaranteed to reduce coverage bias, since the ratio of maximum possible fold amplification to minimum possible fold amplification is necessarily reduced when the average degree of amplification is reduced. When combined with the cost savings of micro-scaled reactions and increasingly efficient sequence library construction procedures, such an approach is clearly the future of single-cell WGA[4,5].

Today, investigators limit amplification-fold by reducing reaction volumes[6,7] or by limiting reaction time. Although it is currently unknown which approach is more fruitful in bias reduction, both approaches have drawbacks. The hydrogel reaction format offers unique advantages in limited-extent WGA, as the product clusters from each template molecule only reach a few microns in size, even under dilute template conditions. This suggests that one can achieve uniform (limited) reaction extent across single-cell WGA reactions, even when the reactions occur asynchronously. The hydrogel format also enables maintenance of optimal amplification conditions for each template throughout the reaction time course if desired by reagent supplementation, possibly reducing sequence content and template fragment length biases.

Suitability of in-gel amplification format for product cluster labeling

In some applications, the demand for single-cell assay throughput is not driven by the need to amass a large number of single-cell datasets, but rather to access cells that are rare in the population. The hydrogel format is ideally suited for this case as the WGA reaction endpoint is an opportune moment to genotype product clusters using hybridization probes in order to identify products from cells of interest for retrieval and deep sequencing analysis[8,9]. In the post-reaction hydrogel, genomic sequences have been amplified and are not protected by a cell envelope. In addition, the thin gel slab format facilitates the application of reagents for rapid template denaturation, labeling, and de-staining. Once labeled, the desired targets can be selectively retrieved for further analysis by image-guided selection. Sequence-specific labeling also has potential to reduce the number of false-positive background spots that challenged the interacting dye-based approach we used in this study and/or to lend molecular specificity to quantification assays. In fact, sequential FISH could be used to probe for large sets of functional genes within the gel itself, enabling the application of complex selection criteria[10].

BLAST analysis and read assignment for *E. coli* and *S. aureus*

To characterize all samples after quality trimming, each sample (R1 from each read pair) was blasted (task megablast) with the parameters listed in **supplementary figure 6**. The BLAST database for *E. coli* consists of three *E. coli* genomes (strain BL21, MG1655 and W3110). The *S. aureus* database consists of the genomes of strain 8325, TW20 and USA300. Univec, Plasmid and Human genome (GRCH38) databases were downloaded from NCBI. All databases were produced using makeblastdb and blastdb_aliastool. Each read was mapped to all five

databases (*E. coli*, *S. aureus*, Univec, Plasmid, and Human db) and the results were ranked based on bit score, e-value and percent identity. We assigned each read to one of the source databases based on the top hit. Using the filter_fasta.py tool in QIIME[11], we selected reads that did not map to any of the five databases for further analysis. We ran BLAST against the nt database to characterize these reads (**Supplementary Table 6**).

Custom reference generation by *de novo* assembly for *E. coli* and *S. aureus*

The gDNA *E. coli* (BL21) and *S. aureus* (NCTC 8325) positive-control data were assembled and curated to create custom reference genome sequences. Raw sequencing files were quality trimmed using Trimmomatic (**Supplementary Fig. 6**). We blasted and filtered the trimmed files against respective reference database, with the parameters listed in **supplementary figure 6**. The filtered and trimmed files were assembled into unordered contigs with velvet. We mapped (BWA) unordered contigs to their closest NCBI reference genome (NC_012971 and NC_007795.1 respectively). The resulting SAM files were ranked on mapped length in descending order. Using a custom MATLAB script, we created a reference genome backbone consisting of only the null character '-' with the same size as the reference genome and wrote nucleotide sequences on it with the top-scoring SAM-mapped sequence for each contig. We conducted the same assembly process for the genomic DNA (*E. coli* DH10B) data from de Bourcy *et al* 2014 using reference genome NC_010473.1. Assembly Statistics are listed in the **supplementary table 5**.

NGS data analysis for *E. coli* and *S. aureus*

Data quality was first visualized using FastQC (Babraham Bioinformatics). All data were trimmed using TRIMMOMATIC[12] and human reads were filtered out with BLAST and QIIME. Each pair of trimmed and filtered reads was piped into BWA and mapped to the custom reference sequences (unordered contigs). Picard tools SortSam was used to produce BAM files and MarkDuplicates was deployed to mark duplicate reads. The data analysis workflow is illustrated in **supplementary figure 6**. The samples included two positive-control purified genomic DNA samples, seven hydrogel MDA punches identified as *E. coli* only by qPCR, seven punch samples identified as *S. aureus* only by qPCR, and seven punch samples identified by qPCR as double negative. Mapping statistics were obtained using the GAEMR (Broad Institute) get_simple_bam_stats.py tool. Genome coverage was obtained using Bedtools genomecov. Lorenz curve ranks the depth of coverage across the genome sequences in each sample and plots the fraction of total coverage needed to recover a fraction of the genome sequences represented in each dataset (at 20× or maximum depth if sequenced shallow). An area-under-Lorenz-curve (AUC) value of 0.5 represents perfectly uniform representation. Gini Index = 2 × (0.5 − AUC). A Gini index of 0 indicates perfect uniformity and a Gini index of 1 indicates maximal non-uniformity. Lorenz curves were obtained by first processing down-sampled BAM files (duplicates marked) using samtools mpileup, then ranking the ascending coverage per base pair, and plotting with a custom MATLAB script. Single-cell *E. coli* MDA data from de Bourcy *et al* 2014 were downloaded from NCBI Sequence Read Archive (SRA) and analyzed the same way.

Chimera statistics for *E. coli* and *S. aureus*

In order to obtain chimera statistics, we mapped read 1 and read 2 from each sample separately using BWA. To make the chimera statistics comparable, we used *de novo* assembled genome (described below) from all bulk genomic DNA samples. We sorted the SAM file by read index and filtered out repetitive mapping results. We used a custom python code to import pysam in order to pair up the 'mapping position', 'is-reverse', 'read length' and 'read index' information into a .mat file. With a customized MATLAB script, we calculated the insert size for each read pair and checked their relative orientation. We filtered out pairs that were mapped one-ended. The chimera percentage was calculated as the (number of properly orientated read pairs with insert size more than 1000 bp + number of read pairs of wrong orientation)/Total number of read pairs.

Random subsampling of mapped reads for *E. coli* and *S. aureus*

Duplicates-marked BAM files were down-sampled using samtools and bootstrapped with random number seed 0 to 9 for each depth. See **supplementary table 7** for more information.

Pre-processing and assembly of single-cell genomes from stool

First, we removed the adapter sequences from single-cell libraries using TRIMMOMATIC[12] (TRAILING:3 MINLEN:40). To ensure that human DNA was not captured in our single-cell libraries, we screened single-cell amplicons against the human genome (GRCh38 reference) using BMTagger[13] (default). We screened our amplicons against *E. coli* references (BL21 and DH10B) using BMTagger. Overall, the level of contamination was small. We also screened against *Pseudomonas* (PAO1) and *Staphylococcus* (NCTC 8325) genomes, which were sequenced alongside our libraries, to ensure no chimeric reads formed during sample preparation confounded our analyses. Finally, single genome amplicons were quality filtered (Phred score ≥ 3), and filtered for reads that were less than 45 bp. Amplicons were then assembled using SPADES (v3.6.0) (--careful)[14]. We retained genomes where at least 100 kb could be assembled.

Assessing the fidelity of single-cell genomes from stool

To further vet the quality and purity of our assemblies, we used BLAST to assign taxonomies to a set of 31 predetermined core genes that are both phylogenetically conserved and single copy in almost all genomes[15]. Although we could not identify the full set of 31 core genes in any of the assemblies, we were able to easily distinguish cases where two or more cells were sequenced together from those in which there was a single cell. Additional validation of the single-cell assemblies included quantifying the levels of contamination using CheckM[16] and examining the number and taxonomy identified using RNAMMER[17]. CheckM accesses the quality of a genome using a broader set of marker genes specific to its inferred lineage within a reference genome tree and provides estimates of genome completeness and contamination percentages[16]. RNAMMER uses hidden Markov models trained from ribosomal RNA databases to predict the rRNA species[17]. The extent and contiguity of our assemblies was documented by reporting assembled genome size, N50, the number of contigs, CheckM completeness %, CheckM contamination % and notes on RNAMMER classification (**Supplementary Table 9**).

Notably, some microbes can be difficult to isolate from human stool samples due to the cells' tendency to break or aggregate. Some of the punch samples with low numbers of AMPHORA genes could be the result of broken cells containing reduced genomic representation or free genomic DNA fragments, while samples with evidence for multiple taxonomies could have resulted from cell aggregates. Stool samples are also fairly complex and contain a lot particulate matter that complicates sample processing. In principle, genes from samples with sequences of variable taxonomy could arise for several reasons: the products of multiple cells being collected in a single punch, downstream contamination in the second round MDA or library construction steps, informatic demultiplexing, or from taxonomic mis-classification of hard-to-assign sequences.

Analysis of metagenomic shotgun reads from stool

FijiCOMP metagenomic samples, each containing roughly 50 million paired-reads, were profiled using MetaPhlAn[18]. Metagenomic samples were also aligned to the SILVA rRNA database (v.115) to determine the presence of organisms from the Succinivibrionaceae family. Based on alignments to the SILVA rRNA database, we find that organisms within the Succinivibrionaceae family are in fact highly abundant in the FijiCOMP metagenomic data, with average FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values around 26,000.

SUPPLEMENTARY TABLES:

Supplementary Table 1 | A comparison of single-cell isolation technologies

Single-cell isolation technology	Segregation principle	Engineering requirement	Fixed spatial addressing	Reagent addition	Product recovery	Characteristic reaction volume	Max analytes per µL	Max analytes per mm^2
SBS multi-well Plate	Macro-scale container array	Commodity plate & complex robotics	Yes	Liquid handler or manual pipetting		1 - 100 µL	1	0.10
Lab-on-a-chip microfluidics (e.g. Fluidigm)	Individually addressable microfluidic chambers	Specialized microdevice & controller		Automated by microdevice		0.1 - 500 nL	10,000	0.10
Open microfabricated array	Micro-fabricated container array			Liquid handler or manual pipetting (largest volume only)		0.05 pL* - 10 nL	20,000,000	20,000
Monodisperse microdroplets	Multi-phase system		No	Droplet merging	Droplet breakdown	5 - 1000 pL	200,000	10,000
Hydrogel	Selective diffusion restriction	No special equipment needed	Yes	Diffusion into or out of hydrogel	Physical punch or hydrogel breakdown	0.05 - 1 pL**	20,000,000	20,000

* Men et al, Anal. Chem., 2012 used 3.3 micron diameter by 4.2 micron deep wells
** Defined by physical extent of product; reagents likely drawn from a larger volume

Supplementary Table 2 | QPCR characterization of hydrogel punches

Total hydrogel punches	*E. coli* positive punches	*S. aureus* positive punches	Double positive	Double negative
80	7	36	7	30

Supplementary Table 3 | Primer sequences

Complete Primer list	
MDA hexamer	5'-NNNN*N*N-3'
S. aureus G1 F	TGC ACA TTT AAA CCC AGC GG
S. aureus G1 R	ATC GCA TGT GCA ATT CTC GG
S. aureus arc F	TTG ATT CAC CAG CGC GTA TTG TC
S. aureus arc R	AGG TAT CTG CTT CAA TCA GCG
E. coli G2 F	CAA CCA AAT TAT TGC CGC GC
E. coli G2 R	GCC ACG GTA ATT ACT GTC GC
E. coli uspA F	CCG ATA CGC TGC CAA TCA GT
E. coli uspA R	ACG CAG ACC GTA GGC CAG AT
λDNA 780bp F	CGG CAA ACG GGA ATG AAA CGC C
λDNA 780bp R	TGC GGC AAA GAC AGC AAC GG

* Represents phosphorothioated DNA bases
All sequences are listed from 5' to 3'

Supplementary Table 4 | Mapping statistics, *E. coli* and *S. aureus*

	# Raw reads	Median insert length (*E. coli*)	Median insert length (*S. aureus*)	Mapping to *E. coli* (%)	Mapping to *S. aureus* (%)	Coverage of *E. coli* genome	Coverage of *S. aureus* genome	Percent chimera on *E. coli* ref (%)	Percent chimera on *S. aureus* (%)
E. coli genomic DNA	1,888,337	120		89.86%	0.83%	99.99%		0.08%	
S. aureus genomic DNA	987,244		144	3.04%	88.70%		99.99%		0.12%
S1	873,658		150	1.81%	90.89%		64.35%		0.62%
S2	361,857		178	2.07%	89.02%		70.44%		0.84%
S3	594,951		153	5.17%	53.30%		53.85%		0.39%
S4	654,669		160	2.17%	90.16%		67.87%		0.57%
S5	470,077		155	5.32%	69.80%		66.95%		0.61%
S6	803,684		148	5.27%	64.47%		86.46%		0.49%
S7	450,240		166	1.39%	93.04%		47.34%		0.90%
*E1**	*822,345*	137		*9.81%*	*0.59%*	*7.14%*		0.68%	
E2	473,648	134		61.12%	0.64%	23.85%		0.69%	
E3	425,227	135		77.35%	2.38%	47.59%		0.55%	
E4	527,854	142		55.20%	1.01%	30.52%		0.78%	
*E5**	*1,200,678*	121		*9.76%*	*3.36%*	*14.10%*		0.56%	
E6	745,092	124		85.31%	1.56%	16.77%		0.84%	
E7	1,053,059	124		68.43%	0.38%	23.79%		0.52%	
NTC1	972,597			10.89%	16.96%				
NTC2	671,370			0.51%	1.59%				
NTC3	715,428			0.23%	0.20%				
NTC4	508,975			0.12%	0.07%				
NTC5	784,944			0.99%	3.27%				
NTC6	579,921			84.03%	0.06%				
NTC7	605,265			22.64%	0.13%				
Mock	27,429			25.71%	16.81%				

* False-positive *E. coli* single-cell samples

Supplementary Table 5 | *de novo* assembly statistics, *E. coli* and *S. aureus*

Assembly statistics	*E. coli* gDNA	*S. aureus* gDNA	*E. coli* gDNA (de Bourcy *et al*)
# Contigs ≥ 0bp	126	113	115
# Contigs ≥ 1kbp	101	86	91
Total length ≥ 0bp	4,406,278	2,678,216	4,432,657
Total length ≥ 1kbp	4,396,297	2,666,092	4,422,013
Largest contig	295,162	148,028	326,226
Coverage	97%	95%	95%
GC %	50.80	32.69	50.75
N50	75,214	48,471	85,192

Supplementary Table 6 | Sequence read classification ("other reads"), *E. coli* and *S. aureus*

	"Other reads"	% Identified	Cloning/expression vector %	*Other E. coli %*	*Other S. aureus %*	Synthetic construct %	*Propoionibacterium acens %*	Other major categories		Percent listed
E. coli genomic DNA	42941	64.27%	21.30%	6.07%	0.01%	72.33%	0.12%			99.82%
S.aureus genomic DNA	4720	38.24%	1.16%	0.11%	61.83%	3.66%	10.80%	Staphylococcus phage: 349	(19.34%)	96.90%
S1	11887	4.85%	1.22%	0.17%	20.31%	4.34%	19.62%	Assorted bacteria and fungus: 248	(43.06%)	88.72%
S2	1306	16.62%	3.23%		11.52%	1.84%	77.42%			94.01%
S3	34963	0.24%	9.64%		28.92%	15.66%	12.05%	Assorted bacteria and fungus: 28	(33.73%)	100.00%
S4	1800	25.78%	0.65%		61.64%	2.80%	16.16%			81.25%
S5	4571	27.28%	0.88%	0.24%	62.63%		16.84%	Staphylococcus phage: 214	(17.16%)	97.75%
S6	4899	46.87%			18.60%		57.23%	Assorted bacteria and fungus: 506	(22.04%)	97.87%
S7	14074	14.53%	0.29%		77.65%	1.32%		Staphylococcus phage: 361	(17.65%)	96.92%
E1*	352603	11.23%	1.15%	0.01%		3.99%		Assorted bacteria and fungus: 35284	(89.12%)	94.26%
E2	32744	76.79%	25.00%	2.57%		71.89%	0.14%			99.60%
E3	20547	59.69%	27.47%	0.21%		71.67%	0.44%			99.79%
E4	47915	77.10%	13.09%	0.03%		46.17%	0.92%	Malassezia globosa CBS: 13804	(37.36%)	97.57%
E5*	430977	12.60%	5.95%	0.33%		21.91%	15.95%	Assorted bacteria and fungus: 26635	(49.07%)	93.20%
E6	46757	1.64%	1.56%	0.65%	7.16%	5.34%	39.71%	Listeria seeligeri serovar 1/2b str : 245	(31.90%)	86.33%
E7	55490	77.20%	20.42%	0.01%		79.24%				99.67%
NTC1	365936	0.94%	1.51%		3.22%		12.73%	Assorted bacteria and fungus: 2809	(81.47%)	98.93%
NTC2	303923	86.55%			0.08%		2.99%	Human: 247884	(94.24%)	97.31%
NTC3	251727	89.49%					0.26%	Assorted bacteria and fungus: 218140	(96.84%)	97.09%
NTC4	243833	96.26%					99.96%			99.96%
NTC5	213052	90.75%					2.20%	Human: 187178	(96.81%)	99.02%
NTC6	45986	95.44%				0.04%	0.03%	Human: 43806	(99.81%)	99.87%
NTC7	74719	0.54%			37.87%		45.30%			83.17%
Mock	7204	11.30%	8.60%			29.85%	14.74%	Assorted bacteria and fungus: 373	(45.82%)	99.02%

* False-positive *E. coli* single-cell samples
** Values less than 0.01% were omitted for clarity

Supplementary Table 7 | Down sampling on mapped reads from single-cell MDA samples

Filename	Mapped reads	Fraction:	20X	10X	5X	Source
SRR1614004	1,953,388		0.181	0.091	0.045	de Bourcy tube
SRR1614005	124,353		0	0	0.712	de Bourcy tube
SRR1614006	1,552,882		0.228	0.114	0.057	de Bourcy tube
SRR1614007	222,175		0	0.798	0.399	de Bourcy tube
SRR1614011	2,191,740		0.162	0.081	0.040	de Bourcy MF
SRR1614012	2,034,992		0.174	0.087	0.044	de Bourcy MF
SRR1614013	5,476,888		0.065	0.032	0.016	de Bourcy MF
SRR1614014	2,631,391		0.135	0.067	0.034	de Bourcy MF
SRR1614015	11,448,119		0.031	0.015	0.008	de Bourcy MF
SRR1614016	3,833,180		0.092	0.046	0.023	de Bourcy MF+T
SRR1614017	2,720,938		0.13	0.065	0.033	de Bourcy MF+T
SRR1614018	2,377,409		0.149	0.075	0.037	de Bourcy MF+T
SRR1614019	2,217,047		0.16	0.08	0.04	de Bourcy MF+T
SRR1614020	5,331,247		0.066	0.033	0.017	de Bourcy MF+T
LX1	1,690,339		0.208	0.104	0.052	*E.coli* gDNA (No MDA)
LX11	283,308		0	0.62	0.31	*E.coli* hydrogel
LX12	324,651		0	0.54	0.27	*E.coli* hydrogel
LX21	259,540		0	0.68	0.34	*E.coli* hydrogel
LX23	571,435		0.62	0.31	0.15	*E.coli* hydrogel
LX24	708,610		0.5	0.25	0.12	*E.coli* hydrogel
LX3	728,290		0.29	0.15	0.07	*S.aureus* hydrogel
LX4	335,567		0.64	0.32	0.16	*S.aureus* hydrogel
LX5	321,849		0.66	0.33	0.17	*S.aureus* hydrogel
LX6	609,813		0.35	0.18	0.09	*S.aureus* hydrogel
LX7	335,862		0.64	0.32	0.16	*S.aureus* hydrogel
LX8	528,641		0.4	0.2	0.1	*S.aureus* hydrogel
LX9	435,904		0.49	0.25	0.12	*S.aureus* hydrogel

0 means the total mapped depth are less than the target coverage

Supplementary Table 8 | Metagenomic shotgun profiling weighted with single-cell samples

Sample ID	M1.20	W2.21	WL.26	W2.33	M2.41	Weighted average
Prevotellaceae	27.0	46.0	61.9	28.9	52.1	44.0
Succinivibrionaceae	0.0	0.0	0.0	0.0	0.0	0.0
Clostridiaceae	0.6	0.0	0.9	0.0	0.0	0.2
Bacteroidaceae	0.0	0.1	1.3	0.0	4.1	1.2
Veillonellaceae	0.4	3.4	2.9	4.4	9.0	4.6
Firmicute	0.0	0.0	0.0	0.0	0.0	0.0
Enterobacteriaceae	0.1	33.3	1.2	0.3	0.3	6.4
Lachnospiraceae	8.1	4.2	4.7	16.5	4.1	8.4
Eubacteriaceae	19.3	3.4	7.1	29.5	6.6	14.2
Ruminococcaceae	16.7	5.6	2.8	13.5	7.3	8.7
Megasphaera	0.0	0.0	0.0	0.0	0.0	0.0
Acetobacteraceae	0.0	0.0	0.0	0.0	0.0	0.0
Acidaminococcaceae	8.3	0.6	11.0	0.6	0.0	3.2
Clostridiales	0.0	0.0	0.0	0.0	0.0	0.0
Erysipelotrichaceae	16.0	2.0	0.8	3.4	1.3	3.0
Total (%)	96.5	98.5	94.7	97.0	84.7	94.0
Single cell count	8	21	25	37	26	All =117
Single cell percentage	7%	18%	21%	32%	22%	All= 100%

We first acquired the taxonomy ranking from single-cell data including all five samples. In this table we list the taxonomic composition of the samples predicted from standard shotgun metagenomic data. For comparison with the taxonomic profile obtained from our single-cell data, we computed a weighted average taxonomic composition based on the number of single-cell datasets obtained from each sample.

Supplementary Table 9 | Features of 117 FijiCOMP single-cell assemblies
See a separate Excel file

Supplementary References:

1. Weber, L. M., Lopez, C. G. & Anseth, K. S. Effects of PEG hydrogel crosslinking density on protein diffusion and encapsulated islet survival and function. *J Biomed Mater Res A* **90,** 720–729 (2009).
2. White, R. A., Blainey, P. C., Fan, H. C. & Quake, S. R. Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics* **10,** 116 (2009).
3. Blanco, L. & Salas, M. Replication of phage phi 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length phi 29 DNA. *Proc. Natl. Acad. Sci. U.S.A.* **82,** 6404–6408 (1985).
4. Blainey, P. C. The future is now: single-cell genomics of bacteria and archaea - Blainey - FEMS Microbiology Reviews - Wiley Online Library. *FEMS Microbiol. Rev.* **37,** 407–427 (2013).
5. de Bourcy, C. F. A. *et al.* A quantitative comparison of single-cell whole genome amplification methods. *PLoS ONE* **9,** e105585 (2014).
6. Marcy, Y. *et al.* Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet.* **3,** 1702–1708 (2007).
7. Landry, Z. C., Giovanonni, S. J., Quake, S. R. & Blainey, P. C. in *Methods in Enzymology* **531,** 61–90 (Elsevier, 2013).
8. Niki, H. & Hiraga, S. Subcellular distribution of actively partitioning F plasmid during the cell division cycle in E. coli. *Cell* **90,** 951–957 (1997).
9. Yamada, N. A. *et al.* Visualization of Fine-Scale Genomic Structure by Oligonucleotide-Based High-Resolution FISH. *Cytogenet Genome Res* **132,** 248–254 (2011).
10. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat Meth* **11,** 360–361 (2014).
11. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* **7,** 335–336 (2010).
12. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).
13. Rotmistrovsky, K. & Agarwala, R. *BMTagger: Best Match Tagger for removing human reads from metagenomics datasets*. (2011).
14. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).
15. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28,** 1033–1034 (2012).
16. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25,** 1043–1055 (2015).
17. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. (2007).
18. Segata, N., Waldron, L., Ballarini, A. & Narasimhan, V. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature* (2012).