Supporting Information
Expanded Multiplexing on Sensor Constrained Microfluidic Partitioning Systems

Pavan K. Kota[1], Hoang-Anh Vu[1], Daniel LeJeune[2], Margaret Han[3], Saamiya Syed[4], Richard G. Baraniuk[2], and Rebekah A. Drezek*[1]

[1]Department of Bioengineering, Rice University
[2]Department of Electrical and Computer Engineering, Rice University
[3]Department of Biosciences, Rice University
[4]Department of Engineering Technology, University of Houston
*Corresponding author: drezek@rice.edu

S1 Probe Design

We used ThermoBLAST from DNA Software (Plymouth, MI) to align the 16S primers (27F and 1492R) against bacterial genomes and find amplicons. We passed these amplicons to a custom Matlab script to design probes. We chose a sequence length of 11 nucleotides with 5-8 GC nucleotides, without four consecutive G's or C's, and without a G on the 5' end to avoid self-quenching of the fluorophore. We used Smith-Waterman alignment in Matlab to pre-screen for probes that self-hybridize and to assess cross-hybridization of probes amongst the evolving candidate set. To assist in achieving near binary measurements, we considered perfect matches on all 16S copies to be "1" for a genome, and for imperfect homology, we filtered for sequences that had neither nine consecutive matches nor a single G-T mismatch. This latter filtering is a proxy for ensuring that probes have weak, negligible interactions against 16S sequences where they do not have perfect complementarity. The former filtering for positive hits was intended to avoid the issue of mixtures of barcodes for any particular bacteria for simplicity in our initial demonstration.

Given a set of filtered, candidate probes, we used a coordinate ascent strategy to iteratively optimize a set. We hypothesized that barcoding the full-length 16S gene with probes could achieve genus level resolution, as sequencing the full gene achieves a mix of genus and species resolution. As a result, we encouraged similarity of the three *Staphylococcus* species and the two *Streptococcous* species. Define $\mathcal{S}$ as a set of pairs of bacteria (b_i, b_j) within a genus that are similar. The complementary set $\mathcal{D}$ includes all other bacterial pairs. Let $\mathbf{k}_{p,i}$ represent the 11-mer barcode of bacteria i with probes indexed by p. Coordinate ascent sought to solve:

$$\arg\max_{p} \left[\sum_{(b_i,b_j)\in\mathcal{S}} -\|\mathbf{k}_{p,i} - \mathbf{k}_{p,j}\|_2^2 + \log\left(\sum_{(b_i,b_j)\in\mathcal{D}} \|\mathbf{k}_{p,i} - \mathbf{k}_{p,j}\|_2 \right) \right] + \theta \min_{(b_i,b_j)\in\mathcal{D}} \|\mathbf{k}_{p,i} - \mathbf{k}_{p,j}\|_2 \qquad \text{(S1)}$$

The first term is taken from research in metric learning [1], and the second term (with a weight of $\theta = 10$) highly rewards some nonzero separation between all bacterial pairs that are intended to be discriminated between. We chose an initial random set of probes that passed our cross-hybridization check. We iteratively cycled through a shuffled order of the candidate set of probes, evaluating one probe at a time for replacement with any of the other probes that passed the initial filtering step. If replacing a probe improved the objective function, the probe set was updated and the search continued. The algorithm terminated when all probe sequences had been evaluated for replacement but not replaced. For the chosen set of sequences, we evaluated the alignment against 16S genes with imperfect homology (the zeroes in the barcodes). As much as possible, we positioned LNAs at mismatch sites to improve the thermodynamic discrimination against these sequences. We evaluated all T_m's in IDT's OligoAnalyzer, positioning additional LNAs as necessary to reach a sufficient probe T_m.

S2 Theory: Identifiability with Common Types of Sensors

For estimating $\boldsymbol{\lambda}$, the property of *identifiability* means that there is a one-to-one correspondence between each realizable distribution of measurements and the Poisson rates $\boldsymbol{\lambda}$: if $p(\mathbf{y}|\boldsymbol{\lambda}) = p(\mathbf{y}|\boldsymbol{\lambda}')$ for all $\mathbf{y}$, then $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$. From an optimization perspective, identifiability implies that the $\boldsymbol{\lambda}^*$ is the unique global optimum to the likelihood function if we have infinite measurements. Therefore, identifiability is a necessary condition for our method to work.

S2.1 Notation

We use bold face upper and lower case letters for matrices and vectors, respectively. Non-bold, lower case letters represent scalars. We denote the vector of all zeros as $\mathbf{0}$ with its dimensionality dependent on context. We use script letters ($\mathcal{A}$, $\mathcal{B}$, etc.) to denote sets. We denote $\mathbf{e}_j$ as the standard basis vector with $\mathbf{e}_j = 1$ and $\mathbf{e}_i = 0$ for all $i \neq j$. Let $\mathbf{a}$ and $\mathbf{b}$ be two arbitrary vectors of the same dimension, and let a_i and b_i denote their ith elements. We use $\text{supp}(\mathbf{a})$ to denote the support of vector $\mathbf{a}$ defined as the index set where $a_i > 0$ for $i \in \text{supp}(\mathbf{a})$. We use the notation $\mathbf{a} \succeq \mathbf{b}$ to imply that $a_i \geq b_i \ \forall i$, and we use $\mathbf{a} \succ \mathbf{b}$ to further imply the existence of at least one index i where $a_i > b_i$. A set in the subscript of a vector such as $\mathbf{x}_{\mathcal{A}}$ refers to the subvector of $\mathbf{x}$ indexed by the elements of $\mathcal{A}$. We make frequent use of the shorthand $\sum \mathbf{a}$ to denote the summation over elements of a vector $\mathbf{a}$.

S2.2 Definitions and Assumptions

We treat the dataset of measurements from all partitions in a sensor group as samples of a random variable $\mathbf{y}$. The signal (i.e., analyte quantities in a partition) $\mathbf{x}$ is N-dimensional with $\mathbf{x} \sim \text{Poisson}(\boldsymbol{\lambda}^*)$. Each signal is measured by M sensors to yield the observation vector $\mathbf{y}$ (e.g., M fluorescence measurements). We define the function $f : \mathbb{Z}_+^N \rightarrow \mathbb{R}^M$ that is composed of M scalar functions $f_m : \mathbb{Z}_+^N \rightarrow \mathbb{R}$. A particular measurement value y_m is determined by the sensor output $f_m(\mathbf{x})$ plus some additive, zero-mean random noise n_m.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_M(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix} \tag{S2}$$

Note that the sensor functions f_m are group-dependent. For example, each group may have different probes.

We assume that our sensors are *monotonic* and that our analytes obey *responsiveness* and *fingerprint equivalence*. Given these properties, we prove sufficient conditions for identifiability. Without loss of generality, we will say that all M sensor functions are monotonic.

Definition S2.1 (Monotonic Sensors). *A sensor function $f_m : \mathbb{Z}_+^N \rightarrow \mathbb{R}$ is monotonic increasing if $\mathbf{a} \succeq \mathbf{b} \Rightarrow f_m(\mathbf{a}) \geq f_m(\mathbf{b})$ and monotonic decreasing if $\mathbf{a} \succeq \mathbf{b} \Rightarrow f_m(\mathbf{a}) \leq f_m(\mathbf{b})$.*

Monotonic functions are very common and natural; for instance, many sensing modalities have a monotonic increasing sigmoidal response to their input. Any time a Lemma or Theorem relies on monotonicity, its proof will assume all M sensors are monotonic increasing without loss of generality. Next, we define the *responsiveness* property of analytes:

Definition S2.2 (Responsiveness). *If $f(\mathbf{e}_i) \neq f(\mathbf{0})$, the analyte indexed by i is said to be responsive. If $f(\mathbf{e}_i) = f(\mathbf{0})$, then the analyte indexed by i is nonresponsive. If we let $\mathcal{B}$ define the set of indices for all such nonresponsive analytes ($i \in \mathcal{B}$), then for any two signals $\mathbf{x}$ and $\mathbf{x}'$, $f(\mathbf{x}) = f(\mathbf{x}')$ if $x_n = x'_n$ for all $n \notin \mathcal{B}$.*

In other words, a nonresponsive analyte does not influence the sensor output regardless of its quantity. An analyte is considered "responsive" if a single copy yields a different measurement than the null signal (e.g., an empty microfluidic partition).

We define a final intuitive condition on our system called *fingerprint equivalence*. The *fingerprint* of analyte n is the measurement yielded by an isolated copy of the analyte, or $f(\mathbf{e}_n)$. Among analytes with identical fingerprint responses within a sensor group, the total number of occurrences of these analytes dictates the output response. In other words, the sensors treat these analytes as interchangeable copies of each other.

Definition S2.3 (Fingerprint Equivalence). *Let $\mathcal{X} \subseteq \{1, ..., N\}$ be an index set of analytes with identical fingerprints, i.e. $f(\mathbf{e}_i)$ is fixed for all $i \in \mathcal{X}$. A system has the fingerprint equivalence property if for any pair of vectors $\mathbf{x}$ and $\mathbf{x}'$ with $\text{supp}(\mathbf{x})$, $\text{supp}(\mathbf{x}') \subseteq \mathcal{X}$ and $\sum_n x_n = \sum_n x'_n$, we have $f(\mathbf{x}) = f(\mathbf{x}')$.*

Note that even if all analyte fingerprints are distinct, multiple signals can still map to the same measurement vector since we allow for cases of multi-analyte capture in the same partition. We define these signals as members of a *collision set.*

Definition S2.4 (Collision Sets)**.** *The collision set* $\mathcal{C}_{\mathbf{x}}$ *for signal* $\mathbf{x}$ *is the set of all signals* $\mathbf{x}'$ *that satisfy* $f(\mathbf{x}') = f(\mathbf{x})$*.*

We define $\mathcal{U}$ as the set of unique collision sets. In 2-channel ddPCR with binarized measurements, there are four collision sets in each sensor group ($\{0,1\}^2$). It will soon be clear that observations $\mathbf{y}$ are drawn from a mixture model. We can define each *mixture element* as follows:

Definition S2.5 (Mixture Element)**.** *The mixture element* $\mathcal{E}_{\mathbf{x}}$ *for signal* $\mathbf{x}$ *is the set of all signals* $\mathbf{x}'$ *that satisfy* $p(\mathbf{y}|\mathbf{x}) \sim p(\mathbf{y}|\mathbf{x}')$*.*

Note with any zero-mean noise, $p(\mathbf{y}|\mathbf{x}) \sim p(\mathbf{y}|\mathbf{x}') \Rightarrow f(\mathbf{x}) = f(\mathbf{x}')$ such that $\mathcal{E}_{\mathbf{x}} \subseteq \mathcal{C}_{\mathbf{x}}$. In some cases, such as additive white Gaussian noise, $\mathcal{E}_{\mathbf{x}} = \mathcal{C}_{\mathbf{x}}$. We define $\mathcal{V}$ as the set of unique mixture elements with arbitrary $\mathcal{E}_v \in \mathcal{V}$.

S2.3 Proof of Identifiability

With G different sensor groups indexed by g, we assume that the sensor group applied to a measurement $\mathbf{y}$ is known and deterministic. Each sensor group has a different function f that maps $\mathbf{x}$ to M-dimensional space (e.g., different probes in ddPCR). Identifiability means that $p(\mathbf{y}|\boldsymbol{\lambda}) = p(\mathbf{y}|\boldsymbol{\lambda}')\ \forall \mathbf{y}, g \Rightarrow \boldsymbol{\lambda} = \boldsymbol{\lambda}'$. Each $\boldsymbol{\lambda}$ must yield a unique set of G distributions of measurements.

We will use the notation $\mathcal{C}_u^g$ and $\mathcal{E}_v^g$ to specify the group g when necessary. For an arbitrary group, we can express $p(\mathbf{y}|\boldsymbol{\lambda})$ as:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\lambda}) &= \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}) P(\mathbf{x}|\boldsymbol{\lambda}) \\ &= \sum_{\mathcal{E}_v \in \mathcal{V}} p(\mathbf{y}|\mathbf{x} \in \mathcal{E}_v) P(\mathcal{E}_v|\boldsymbol{\lambda}), \end{aligned} \tag{S3}$$

$$P(\mathcal{E}_v|\boldsymbol{\lambda}) = \sum_{\mathbf{x} \in \mathcal{E}_v} P(\mathbf{x}|\boldsymbol{\lambda}). \tag{S4}$$

If a mixture distribution is identifiable, it means that identical distributions must come from the same set of weights on the mixture elements; in this context, $p(\mathbf{y}|\boldsymbol{\lambda}) \sim p(\mathbf{y}|\boldsymbol{\lambda}') \Rightarrow P(\mathcal{E}_v|\boldsymbol{\lambda}) = P(\mathcal{E}_v|\boldsymbol{\lambda}')\ \forall v$. Many finite mixtures (what we practically have in MMVP) and countably infinite mixtures with common noise distributions are identifiable [2-3], and we assume that the system noise characteristics lend to an identifiable mixture. However, we need to prove the identifiability of MMVP, or that equal mixture weights implies equal Poisson parameters: $P(\mathcal{E}_v^g|\boldsymbol{\lambda}) = P(\mathcal{E}_v^g|\boldsymbol{\lambda}')\ \forall v, g \Rightarrow \boldsymbol{\lambda} = \boldsymbol{\lambda}'$. Note that because $\mathcal{E}_{\mathbf{x}}^g \subseteq \mathcal{C}_{\mathbf{x}}^g$ and unique collision sets are disjoint, $P(\mathcal{E}_v^g|\boldsymbol{\lambda}) = P(\mathcal{E}_v^g|\boldsymbol{\lambda}')\ \forall v, g \Rightarrow P(\mathcal{C}_u^g|\boldsymbol{\lambda}) = P(\mathcal{C}_u^g|\boldsymbol{\lambda}')\ \forall u, g$.

We assume $P(\mathcal{C}_u^g|\boldsymbol{\lambda}) = P(\mathcal{C}_u^g|\boldsymbol{\lambda}')\ \forall u, g$ and prove the implication of $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$ given a set of monotonic sensors and with analytes exhibiting responsiveness and fingerprint equivalence in all G groups. We first focus on what can be concluded from a single, arbitrary sensor group (dropping the g superscript) with analytes potentially having nonunique fingerprints, and then we conclude with how multiple groups can be pooled to achieve identifiability. Again, note that our analysis will focus on monotonic increasing sensors without loss of generality.

Define $\mathcal{A} \subseteq \{1, ..., N\}$ such that analytes indexed by $a \in \mathcal{A}$ are all responsive such that there exists some m such that $f_m(\mathbf{e}_a) > f_m(\mathbf{0})$. Define the complementary set $\mathcal{B}$ with nonresponding analytes indexed by b.

Lemma S2.1. *If* f_m *is monotonic increasing for all* $m \in \{1, ..., M\}$*, and if only the analytes indexed by* $a \in \mathcal{A} \subseteq \{1, ..., N\}$ *are responsive, then* $f(\mathbf{x}) = f(\mathbf{0})$ *if and only if* $\mathbf{x}_{\mathcal{A}} = \mathbf{0}$*.*

Proof. Consider $\mathbf{z}$ such that $\mathbf{z}_{\mathcal{A}} = \mathbf{0}$. Note that $\text{supp}(\mathbf{z}) \subseteq \mathcal{B}$. Because analytes indexed by $b \in \mathcal{B}$ are nonresponding, $f(\mathbf{z}) = f(\mathbf{0})$ by definition. Next, we prove the forward condition, $f(\mathbf{x}) = f(\mathbf{0}) \Rightarrow \mathbf{x}_{\mathcal{A}} = \mathbf{0}$, by contradiction. Say $f(\mathbf{z}) = f(\mathbf{0})$ and let $\mathbf{z}$ satisfy $z_a \geq 1$ for some $a \in \mathcal{A}$. By definition of $\mathcal{A}$, $f(\mathbf{e}_a) > f(\mathbf{0})$, and $\mathbf{z} \succeq \mathbf{e}_a$. With monotonic functions, $f(\mathbf{z}) \succeq f(\mathbf{e}_a) \succ f(\mathbf{0})$ and we have arrived at a contradiction. □

The key concept to carry forward is that values of elements in $\mathbf{x}_\mathcal{B}$ are entirely arbitrary for the analysis of collision sets.

Lemma S2.2. *Let f_m be monotonic increasing for all $m \in \{1, ..., M\}$, and let only the analytes indexed by $a \in \mathcal{A} \subseteq \{1, ..., N\}$ be responsive. If $P(\mathcal{C}_\mathbf{0}|\boldsymbol{\lambda}) = P(\mathcal{C}_\mathbf{0}|\boldsymbol{\lambda}')$, then $\sum \boldsymbol{\lambda}_\mathcal{A} = \sum \boldsymbol{\lambda}'_\mathcal{A}$.*

Proof. By Lemma S2.1, $\mathcal{C}_\mathbf{0}$ contains all $\mathbf{x}$ with $\mathbf{x}_\mathcal{A} = \mathbf{0}$ with arbitrary values on $\mathbf{x}_\mathcal{B}$. Therefore, $P(\mathcal{C}_\mathbf{0}|\lambda) = P(\mathcal{C}_\mathbf{0}|\lambda')$ implies

$$P(\mathbf{x}_\mathcal{A} = \mathbf{0}|\boldsymbol{\lambda}) = P(\mathbf{x}_\mathcal{A} = \mathbf{0}|\boldsymbol{\lambda}') \tag{S5}$$

$$e^{-\sum \boldsymbol{\lambda}_\mathcal{A}} = e^{-\sum \boldsymbol{\lambda}'_\mathcal{A}} \tag{S6}$$

$$\sum \boldsymbol{\lambda}_\mathcal{A} = \sum \boldsymbol{\lambda}'_\mathcal{A}. \tag{S7}$$

□

Lemma S2.3. *Let f_m be monotonic increasing for all $m \in \{1, ..., M\}$, and let only the analytes indexed by $a \in \mathcal{A} \subseteq \{1, ..., N\}$ be responsive. For $a \in \mathcal{A}$, if for all $\mathbf{x} \in \mathcal{C}_{\mathbf{e}_a}$, x_a is the only nonzero value in $\mathbf{x}_\mathcal{A}$, then $\lambda_a = \lambda'_a$.*

Proof. We assume $P(\mathcal{C}_{\mathbf{e}_a}|\boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{e}_a}|\boldsymbol{\lambda}')$. By definition of $\mathcal{A}$, $f(\mathbf{e}_a) \succ f(\mathbf{0})$. If x_a is the only nonzero value of $x_\mathcal{A}$, we have $\mathcal{C}_{\mathbf{e}_a} = \{c\mathbf{e}_a \colon c \in \mathcal{K} \subseteq \{1, 2, \ldots\}\}$. Then, $P(\mathcal{C}_{\mathbf{e}_a}|\boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{e}_a}|\boldsymbol{\lambda}')$ implies

$$e^{-\sum \boldsymbol{\lambda}_\mathcal{A}} \sum_{c \in \mathcal{K}} \frac{\lambda_a^c}{c!} = e^{-\sum \boldsymbol{\lambda}'_\mathcal{A}} \sum_{c \in \mathcal{K}} \frac{{\lambda'_a}^c}{c!}. \tag{S8}$$

Using Lemma S2.2, $\sum_{c \in \mathcal{K}} \frac{\lambda_a^c}{c!} = \sum_{c \in \mathcal{K}} \frac{{\lambda'_a}^c}{c!}$, which implies $\lambda_a = \lambda'_a$ since the function on both sides is monotonic in λ_a. □

From here, we first derive results for the special case where all analytes indexed by $\mathcal{A}$ have unique single-copy fingerprints. Afterwards, we generalize to multiple groups, allowing for equal nonzero fingerprints within a group. The next Lemma guarantees at least one index a to which Lemma S2.3 can be applied.

Lemma S2.4. *Let f_m be monotonic increasing for all $m \in \{1, ..., M\}$, and let only the analytes indexed by $a \in \mathcal{A} \subseteq \{1, ..., N\}$ be responsive. If $f(\mathbf{e}_i) \neq f(\mathbf{e}_j)$ $\forall i, j \in \mathcal{A}$ with $i \neq j$, $\exists a \in \mathcal{A}$ such that all $\mathbf{x} \in \mathcal{C}_{\mathbf{e}_a}$ are nonzero in $\mathbf{x}_\mathcal{A}$ only on index a.*

Proof. First, with unique nonzero fingerprints in $\mathcal{A}$ and monotonic sensors, the fingerprint responses $f_m(\mathbf{e}_a)$ can be sorted. Starting arbitrarily with $m = 1$, we can select the minimal set $\mathcal{M} \subseteq \mathcal{A}$ that minimizes $f_1(\mathbf{e}_a)$ such that $\forall a \in \mathcal{M}, \forall j \in \mathcal{M}^c$, $f_1(\mathbf{e}_a) < f_1(\mathbf{e}_j)$. If $|\mathcal{M}| > 1$, then the process can be repeated with $m = 2$ (and so forth) on the subset $\mathcal{M}$ until there is one unique minimum and its corresponding index a.

For this $\mathbf{e}_a$, all $i \in \mathcal{A} \setminus \{a\}$ satisfy $f_m(\mathbf{e}_i) > f_m(\mathbf{e}_a)$ for at least one m. Therefore, signals in the collision set $\mathbf{x} \in \mathcal{C}_{\mathbf{e}_a}$ must satisfy $\mathbf{x}_i = 0$ for all $i \in \mathcal{A} \setminus \{a\}$. Signals with at least one $\mathbf{x}_i \geq 1$ would have at least one m where $f_m(\mathbf{x}) > f_m(\mathbf{e}_a)$, and therefore not be in the collision set by definition. This conclusion completes the proof by contradiction. □

Next, we show how this result chains to all analytes indexed in $\mathcal{A}$.

Lemma S2.5. *Let f_m be monotonic increasing for all $m \in \{1, ..., M\}$, and let only the analytes indexed by $a \in \mathcal{A} \subseteq \{1, ..., N\}$ be responsive. If $f(\mathbf{e}_i) \neq f(\mathbf{e}_j)$ $\forall i, j \in \mathcal{A}$ with $i \neq j$, $\lambda_a = \lambda'_a$ $\forall a \in \mathcal{A}$.*

Proof. Lemmas S2.3 and S2.4 guarantee at least one a that yields $\lambda_a = \lambda'_a$. Let us call this index a_1 and define the subset $\mathcal{S} \subseteq \mathcal{A}$, the subset of indices for which $\lambda_i = \lambda'_i$ $\forall i \in \mathcal{S}$. At this point, $\mathcal{S} = \{a_1\}$. Repeating the process in the proof of Lemma S2.4, we can find a new index a_2 that satisfies $f_m(\mathbf{e}_n) > f_m(\mathbf{e}_{a_2})$ $\forall n \in \mathcal{S}^c \backslash \{a_2\}$ for at least one m.

For the direct proof of identifiability, we assume $P(\mathcal{C}_{\mathbf{e}_{a_2}}|\boldsymbol{\lambda}) = P(\mathcal{C}_{\mathbf{e}_{a_2}}|\boldsymbol{\lambda}')$, or:

$$\sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{e}_{a_2}}} P(\mathbf{x}|\boldsymbol{\lambda}) = \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{e}_{a_2}}} P(\mathbf{x}|\boldsymbol{\lambda}'). \tag{S9}$$

Among signals $\mathbf{x}$ in $\mathcal{C}_{\mathbf{e}_{a_2}}$, $x_n = 0$ for $n \in \mathcal{S}^c \setminus \{a_2\}$ because $f_m(\mathbf{e}_n) > f(\mathbf{e}_{a_2})$ for some m, and sensors are monotonic. These signals can also be partitioned into those with $x_{a_2} = 0$, and those with $x_{a_2} \geq 1$. If any of the former type exist, then $x_i > 0$ for some of the indices $i \in \mathcal{S}$. For instance, we could have $f(2\mathbf{e}_i) = f(\mathbf{e}_{a_2})$ with $x_i = 2$. These signals' terms in the summation follow the form $\left[\prod_{n\in\mathcal{S}} P(x_n|\lambda_n)\right] e^{-\sum_{i\in\mathcal{S}^c}\lambda_i}$. Note that $\lambda_i = \lambda'_i \ \forall i \in \mathcal{S}$ combined with Lemma S2.2 yields $\sum_{i\in\mathcal{S}^c}\lambda_i = \sum_{i\in\mathcal{S}^c}\lambda'_i$. Because $\lambda_i = \lambda'_i$ for $i \in \mathcal{S}$, the product component is equal on both sides as well. Therefore, these terms can be eliminated in Equation (S9). We will denote the set of remaining $\mathbf{x}$ in the summation as $\mathcal{C}'$.

In $\mathcal{C}'$, we have $x_{a_2} \in \mathcal{K} \subseteq \{1, 2, ...\}$. Therefore:

$$e^{\sum \boldsymbol{\lambda}_{\mathcal{A}}} \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{e}_{a_2}}} \prod_{n\in\mathcal{S}\cup\{a_2\}} \frac{\lambda_n^{x_n}}{x_n!} = e^{\sum \boldsymbol{\lambda}'_{\mathcal{A}}} \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{e}_{a_2}}} \prod_{n\in\mathcal{S}\cup\{a_2\}} \frac{{\lambda'_n}^{x_n}}{x_n!} \tag{S10}$$

$$\sum_{k\in\mathcal{K}} \frac{\lambda_{a_2}^k}{k!} h_k(\boldsymbol{\lambda}_{\mathcal{S}}) = \sum_{k\in\mathcal{K}} \frac{{\lambda'_{a_2}}^k}{k!} h_k(\boldsymbol{\lambda}'_{\mathcal{S}}), \tag{S11}$$

where h_k is the mapping $\boldsymbol{\lambda}_{\mathcal{S}} \mapsto \sum_{\mathbf{x}\in\mathcal{C}_{\mathbf{e}_{a_2}}\,:\,x_{a_2}=k} \prod_{n\in\mathcal{S}} \frac{\lambda_n^{x_n}}{x_n!}$. Because $\boldsymbol{\lambda}_{\mathcal{S}} = \boldsymbol{\lambda}'_{\mathcal{S}}$, we can replace both $h_k(\boldsymbol{\lambda}_{\mathcal{S}})$ and $h_k(\boldsymbol{\lambda}'_{\mathcal{S}})$ by constants H_k. Therefore,

$$\sum_{k\in\mathcal{K}} \frac{\lambda_{a_2}^k}{k!} H_k = \sum_{k\in\mathcal{K}} \frac{{\lambda'_{a_2}}^k}{k!} H_k. \tag{S12}$$

Because $H_k \geq 0$, both sides are monotonic in λ_{a_2} such that $\lambda_{a_2} = \lambda'_{a_2}$. Now, a_2 can be added to $\mathcal{S}$ and the process can be repeated until $\mathcal{S} = \mathcal{A}$ such that $\boldsymbol{\lambda}_{\mathcal{A}} = \boldsymbol{\lambda}'_{\mathcal{A}}$. □

Now, we will extend this result to the case of having equal fingerprints in the same sensor group—i.e., that $f(\mathbf{e}_i) = f(\mathbf{e}_j)$ for some pairs i, j.

Lemma S2.6. *Let f_m be monotonic increasing for all $m \in \{1, ..., M\}$, and let only the analytes indexed by $a \in \mathcal{A} \subseteq \{1, ..., N\}$ be responsive. Define the disjoint sets $\mathcal{A}_1, \mathcal{A}_2, ...\mathcal{A}_C$ indexed by c with $\cup_{c=1}^{C}\mathcal{A}_c = \mathcal{A}$ such that for all $i, j \in \mathcal{A}_c$, $f(\mathbf{e}_i) = f(\mathbf{e}_j)$. Then, $\sum \boldsymbol{\lambda}_{\mathcal{A}_c} = \sum \boldsymbol{\lambda}'_{\mathcal{A}_c}$ $\forall c$.*

Proof. Note that the Poisson distribution has the property that if x_i are each independently drawn from Poisson(λ_i), then $\sum_{i\in\mathcal{A}_c} x_i \sim$ Poisson($\sum \boldsymbol{\lambda}_{\mathcal{A}_c}$). We can then simply define a dummy variables $\mathbf{x}_c^\dagger = \sum_{i\in\mathcal{A}_c} x_i$ and $\boldsymbol{\lambda}_c^\dagger$ such that $\mathbf{x}_c^\dagger \sim$ Poisson($\boldsymbol{\lambda}_c^\dagger$). This dummy variable represents an "analyte" that appears with a distribution governed by the total quantities of analytes with the same fingerprint. However, what matters for identifiability is the sensor functional values of signals, i.e. that $f(\mathbf{a}) = f(\mathbf{b})$ if for all c, $\sum \mathbf{a}_{\mathcal{A}_c} = \sum \mathbf{b}_{\mathcal{A}_c}$. Namely, $\mathbf{x}_c^\dagger \sim$ Poisson($\boldsymbol{\lambda}_c^\dagger$) by fundamental properties of the Poisson distribution, but it is only with the condition of fingerprint equivalence (Definition S2.3) that lets us apply all previous results that are based on collision sets, i.e., sets of signals with equal functional values. These yield $\boldsymbol{\lambda}_c^\dagger = {\boldsymbol{\lambda}'_c}^\dagger$ $\forall c$, or $\sum \boldsymbol{\lambda}_{\mathcal{A}_c} = \sum \boldsymbol{\lambda}'_{\mathcal{A}_c}$ $\forall c$. □

Theorem S2.7. *Let g index G different sensor groups that satisfy fingerprint equivalence and that contain monotonic, saturating sensors. For each group g, define the row vector $\mathbf{z}_c^g$ of zeros and ones with ones in the indices associated with $\mathcal{A}_c$. Define the N-column matrix $\mathbf{Z}_g$ whose C rows are comprised of $\mathbf{z}_c^g$ $\forall c$. Define the N-column matrix $\mathbf{Z}$ as the vertical concatenation $\mathbf{Z}_g$ $\forall g$. If* rank($\mathbf{Z}$) $= N$*, then $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$.*

Proof. This theorem is a formal way of saying that Lemma S2.6 must yield N independent equations when applied to all groups where the sensing and system conditions hold. We can consider the system of equations yielded by Lemma S2.6 and represented by $\mathbf{Z}\boldsymbol{\lambda} = \mathbf{Z}\boldsymbol{\lambda}'$, or $\mathbf{Z}(\boldsymbol{\lambda} - \boldsymbol{\lambda}') = \mathbf{0}$. If rank($\mathbf{Z}$) $= N$, then it follows that $\boldsymbol{\lambda} = \boldsymbol{\lambda}'$. Therefore, we have $P(\mathcal{C}_u^g|\boldsymbol{\lambda}) = P(\mathcal{C}_u^g|\boldsymbol{\lambda}')$ $\forall(u, g) \Rightarrow \boldsymbol{\lambda} = \boldsymbol{\lambda}'$, concluding our proof of identifiability. □

S3 Special Cases of MLE with ddPCR

From eq 2 in the main text that describes the generalized gradient in MLE, we consider two commonly employed special cases. First, if samples are sufficiently dilute such that partitions are either empty ($\mathbf{x}_d = \mathbf{0}$) or have only one analyte, the goal is often to identify each nonzero signal independently with a classification process. In other words, assays must be designed such that $p(\mathbf{y}_d|\mathbf{x}) > 0$ *only* for $\mathbf{x}^*_d$ - the measurements are *unambiguous.* Setting the gradient equal to zero and simplifying leads to $\hat{\boldsymbol{\lambda}}_{MLE} = \frac{1}{D}\sum_{d=1}^{D} \mathbf{x}^*_d$. In practice, clusters of classes must have reliable decision boundaries and concentrations are estimated by totaling the classification results.

The second specialized case is common for ddPCR where for each PCR assay is specific for a target analyte and assigned to a particular channel. With M channels, $N = M$ and each measurement unambiguously determines the presence or absence of each target sequence. Precisely, $p(y_d|\mathbf{x})$ is one or zero, and considerations of each analytes' quantity x_n can be simplified to $x_n = 0$ (absent) or $x_n > 0$ (present). Each analyte n can be inferred independently such that $\lambda_n = -\log \frac{D_{0,n}}{D}$ where $D_{0,n}$ is the number of droplets that do *not* contain analyte n. This formula can be found by applying the above assumptions, setting eq 2 in the main text to zero, and simplifying.

S4 Exact Gradient Computation and $p(\mathbf{y}|\mathbf{x})$ Model for ddPCR

We first focus on the gradient resulting from a single probe group. In a single group, there are only four viable measurements with $\mathbf{y} \in \{0,1\}^2$. Let us define $\mathcal{Y} = \{0,1\}^2$, and $p_{\mathbf{y}}$ as the proportion of the D measurements that equal $\mathbf{y}$. We can then re-express the log-likelihood maximization as:

$$\widehat{\boldsymbol{\lambda}}_{MLE} = \arg\max_{\boldsymbol{\lambda}} \frac{1}{D}\sum_{d=1}^{D} \log \sum_{\mathbf{x}\in\mathbb{Z}_+^N} p(\mathbf{y}_d|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda}) \tag{S13}$$

$$= \arg\max_{\boldsymbol{\lambda}} \sum_{\mathbf{y}\in\mathcal{Y}} p_{\mathbf{y}} \log \sum_{\mathbf{x}\in\mathbb{Z}_+^N} p(\mathbf{y}_d|\mathbf{x})P(\mathbf{x}|\boldsymbol{\lambda}) \tag{S14}$$

Here, we will use $\boldsymbol{\lambda} \equiv \boldsymbol{\lambda}^{(bact)}$ since we will be optimizing over the bacterial concentrations directly. In our case, *E. cloacae* is the only bacteria with a fractional abundance of a probe binding site - approximately 87.5% of its copies interact with probes 1, 3, and 4, and 12.5% interact with only probes 3 and 4 (Figure S3). Similarly to how we defined $\mathbf{C}$ in the Results and Discussion, we can define $\mathbf{C}^{(g)}$ for group g with each bacterium's fractional abundances of genes with a corresponding barcode. Figure S3 shows an example of $\mathbf{C}^{(1)}$, which can be generated with Figure 1 as a reference.

We define $p(\mathbf{y}|\mathbf{x}) = \prod_m p(y_m|\mathbf{x})$. For $p(y_m = 1|\mathbf{x})$, then $p(y_m|\mathbf{x}) = 1$ if the droplet has at least one copy of a gene that interacts with probe m and $p(y_m|\mathbf{x}) = 0$ otherwise. For $p(y_m = 0|\mathbf{x})$, this likelihood is 1 if none of the genes in the droplet interact with probe m and 0 otherwise.

However, with the analyte currently defined as a copy of the nth bacterium's 16S gene, we must be careful. For instance, with index 3 corresponding with *E. cloacae*, if $x_3 = 1$ in $\mathbf{x}$, $p(y_1|\mathbf{x})$ may not be 1 since one copy of *E. cloacae*'s 16S gene is not guaranteed to interact with probe 1. To resolve this, we will temporarily transform the problem to the space of gene barcodes for this group: $\boldsymbol{\lambda}^{(BC_1)} = \mathbf{C}^{(1)}\boldsymbol{\lambda}$. Note $\boldsymbol{\lambda}^{(BC_1)}$ is 4-dimensional. We can define $\mathbf{x}^{(BC_1)} = [x_{00}, x_{01}, x_{10}, x_{11}]^T$ as the vector representing the quantities of 16S genes from any source bacteria that interact with the probes in the pattern noted in the subscript, noting $\mathbf{x}^{(BC_1)} \sim \text{Poisson}(\boldsymbol{\lambda}^{(BC_1)})$. Lastly, let us define $\mathcal{X}_{\mathbf{y}}$ as that set where if $\mathbf{x}^{(BC_1)} \in \mathcal{X}_{\mathbf{y}}$, then $p(\mathbf{y}|\mathbf{x}^{(BC_1)}) = 1$. Now we can rewrite Equation (S14) as:

$$= \sum_{\mathbf{y}\in\mathcal{Y}} p_{\mathbf{y}} \log \sum_{\mathbf{x}^{(BC_1)}\in\mathcal{X}_{\mathbf{y}}} P(\mathbf{x}^{(BC_1)}|\boldsymbol{\lambda}^{(BC_1)}). \tag{S15}$$

The linearity of gradients allow us to treat this one $\mathbf{y}$ at a time, summing the contributions from each $\mathbf{y}$ at the end. In general, treat 00 as short for $[0,0]$, 01 for $[0,1]$, etc. Let $\nabla^{00}_{\boldsymbol{\lambda}}$ be the component of the gradient from $\mathbf{y} = [0,0]$, $\nabla^{01}_{\boldsymbol{\lambda}}$ from $\mathbf{y} = [0,1]$, etc. We will similarly define the mean log likelihood contributions as

ℓ^{00}, ℓ^{01}, etc. Similarly, define the rows of $\mathbf{C}^{(1)}$ as $\mathbf{C}^{(1)}_{00}, \mathbf{C}^{(1)}_{01}$, etc. By convention, $\boldsymbol{\lambda}$ and other vectors should be assumed to be column vectors, but the rows of $\mathbf{C}^{(1)}$ are row vectors. Thus we have:

$$\begin{aligned}
\ell^{00} &= p_{00} \log P(x_{01} = 0 \text{ and } x_{10} = 0 \text{ and } x_{11} = 0) && \text{(S16)}\\
&= p_{00} e^{-x_{01} - x_{10} - x_{11}} && \text{(S17)}\\
&= p_{00}(-\mathbf{C}^{(1)}_{01} - \mathbf{C}^{(1)}_{10} - \mathbf{C}^{(1)}_{11})\boldsymbol{\lambda} && \text{(S18)}
\end{aligned}$$

$$\nabla^{00}_{\boldsymbol{\lambda}} = p_{00}(-\mathbf{C}^{(1)}_{01} - \mathbf{C}^{(1)}_{10} - \mathbf{C}^{(1)}_{11})^T. \tag{S19}$$

In the first line, we define the conditions for $\mathbf{x}^{(BC_1)} \in \mathcal{X}_{00}$ and solve. Genes that interact with either probe cannot be in droplets that yield $\mathbf{y} = [0, 0]$. Next, for the $\mathbf{y} = [0, 1]$ response, at least one gene that interacts with the 2nd (FAM) probe must be present, and genes that interact with the HEX probe must be absent.

$$\begin{aligned}
\ell^{01} &= p_{01} \log P(x_{01} \geq 1 \text{ and } x_{10} = 0 \text{ and } x_{11} = 0) && \text{(S20)}\\
&= p_{01} \log(1 - e^{-x_{01}}) e^{-x_{10} - x_{11}} && \text{(S21)}\\
&= p_{01}\Bigg(\log(1 - e^{-\mathbf{C}^{(1)}_{01}\boldsymbol{\lambda}}) - \mathbf{C}^{(1)}_{10}\boldsymbol{\lambda} - \mathbf{C}^{(1)}_{11}\boldsymbol{\lambda}\Bigg) && \text{(S22)}
\end{aligned}$$

$$\nabla^{01}_{\boldsymbol{\lambda}} = p_{01}\Bigg[\Bigg(\frac{e^{-\mathbf{C}^{(1)}_{01}\boldsymbol{\lambda}}}{1 - e^{-\mathbf{C}^{(1)}_{01}\boldsymbol{\lambda}}}\Bigg)\mathbf{C}^{(1)T}_{01} - \mathbf{C}^{(1)T}_{10} - \mathbf{C}^{(1)T}_{11}\Bigg]. \tag{S23}$$

A virtually identical simplification for $\nabla^{10}_{\boldsymbol{\lambda}}$ is omitted here. Lastly, for $\mathbf{y} = [1, 1]$, we have:

$$\begin{aligned}
\ell^{11} &= p_{11} \log(P(x_{11} \geq 1) + P(x_{11} = 0 \text{ and } x_{01} > 0 \text{ and } x_{10} > 0) && \text{(S24)}\\
&= p_{11} \log\Bigg((1 - e^{-x_{11}}) + e^{-x_{11}}(1 - e^{-x_{01}})(1 - e^{-x_{10}})\Bigg). && \text{(S25)}
\end{aligned}$$

The remaining algebraic steps are omitted, but the final result is

$$\nabla^{11}_{\boldsymbol{\lambda}} = p_{11} e^{-x_{11}} \frac{-\mathbf{C}^{(1)T}_{11} - e^{-x_{01}}(-\mathbf{C}^{(1)T}_{01} - \mathbf{C}^{(1)T}_{11}) - e^{-x_{10}}(-\mathbf{C}^{(1)T}_{10} - \mathbf{C}^{(1)T}_{11}) - e^{-x_{01} - x_{10}}(-\mathbf{C}^{(1)T}_{01} - \mathbf{C}^{(1)T}_{10} - \mathbf{C}^{(1)T}_{11})}{(1 - e^{-x_{11}}) + e^{-x_{11}}(1 - e^{-x_{01}})(1 - e^{-x_{10}})} \tag{S26}$$

where $\mathbf{C}^{(1)}_{ij}\boldsymbol{\lambda}$ can be substituted for any x_{ij}.

We can now say that for group 1, $\nabla^{(1)}_{\boldsymbol{\lambda}} = \nabla^{00}_{\boldsymbol{\lambda}} + \nabla^{01}_{\boldsymbol{\lambda}} + \nabla^{10}_{\boldsymbol{\lambda}} + \nabla^{11}_{\boldsymbol{\lambda}}$. The above process can be repeated for any group g. Therefore, the final gradient vector (arbitrarily scaled) is

$$\nabla_{\boldsymbol{\lambda}} = \sum_g p_g \nabla^{(g)}_{\boldsymbol{\lambda}} \tag{S27}$$

where p_g is the proportion of total droplets that come from group g.

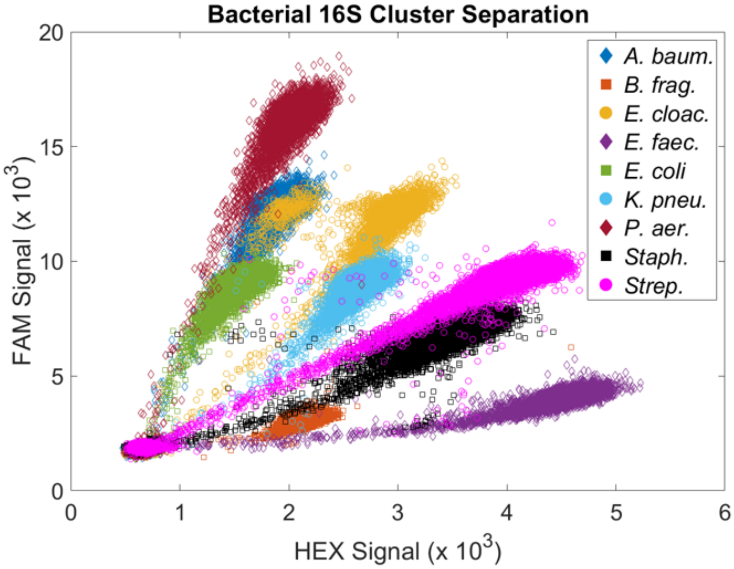
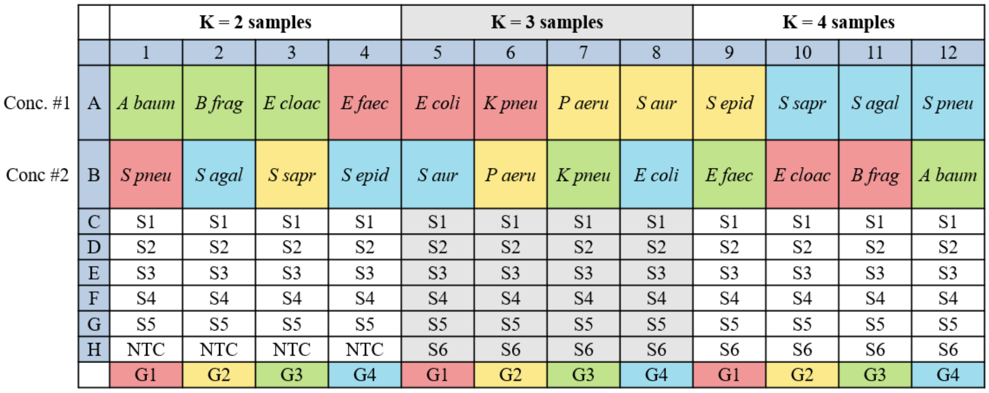


Figure S1: Separation of bacterial barcodes with amplitude multiplexing. Each cluster depicted is from a separate ddPCR reaction with one bacterial species in it. Data from the three *Staphylococcus* bacteria and the two *Streptococcus* bacteria were combined in this plot. Amplitude multiplexing is a technique to resolve more probes than the available number of color channels, but it is typically used with each unique probe participating in an orthogonal assay with its own primer pair. Here, we adjusted probe concentrations to "move" the cluster positions with a single pair of primers. Probes 1 and 5 were tagged with HEX, and Probes 2-4 were tagged with FAM. Probes 1, 2, and 4 at 125 nM, and with Probes 3 and 5 at 250 nM. Based on each 16S gene's barcode, droplets containing that gene will position in clusters whose channel intensities roughly correlate with the total probe concentration tagged with the corresponding fluorophore.

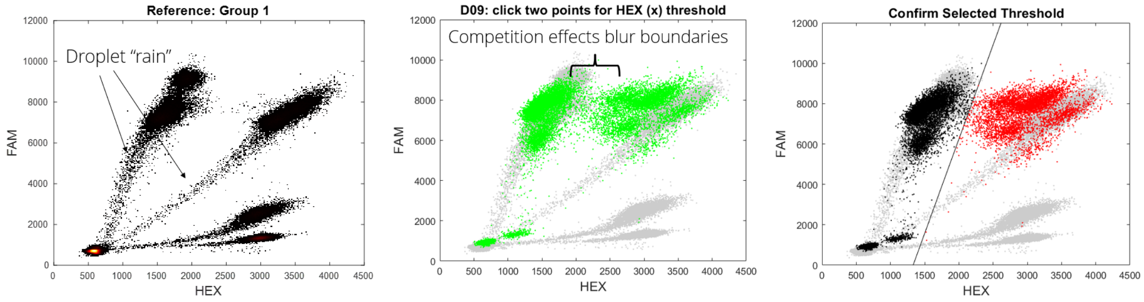


		K = 2 samples				K = 3 samples				K = 4 samples			
		1	2	3	4	5	6	7	8	9	10	11	12
Conc. #1	A	*A baum*	*B frag*	*E cloac*	*E faec*	*E coli*	*K pneu*	*P aeru*	*S aur*	*S epid*	*S sapr*	*S agal*	*S pneu*
Conc #2	B	*S pneu*	*S agal*	*S sapr*	*S epid*	*S aur*	*P aeru*	*K pneu*	*E coli*	*E faec*	*E cloac*	*B frag*	*A baum*
	C	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1	S1
	D	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2
	E	S3	S3	S3	S3	S3	S3	S3	S3	S3	S3	S3	S3
	F	S4	S4	S4	S4	S4	S4	S4	S4	S4	S4	S4	S4
	G	S5	S5	S5	S5	S5	S5	S5	S5	S5	S5	S5	S5
	H	NTC	NTC	NTC	NTC	S6	S6	S6	S6	S6	S6	S6	S6
		G1	G2	G3	G4	G1	G2	G3	G4	G1	G2	G3	G4

Figure S2: Plate layout for ddPCR test samples. The first two rows served as references for ground truth concentration estimation of monomicrobial dilutions and manual thresholding of all wells. The colors of the wells in rows A and B correspond to the probe group applied. Random mixtures of bacteria were distribtued across the rest of the plate with each mixture being applied to four wells, each with a different subset of two probes defining the 16S barcodes.

$$\mathbf{C}^{(1)} = \begin{array}{c|ccccccccc|} & \textit{A. baum.} & \textit{B. frag.} & \textit{E. cloac.} & \textit{E. faec.} & \textit{E. coli} & \textit{K. pneu.} & \textit{P. aeru.} & \textit{Staph.} & \textit{Strep.} \\ \hline \textbf{[0,0]} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \textbf{[0,1]} & 1 & 0 & .125 & 0 & 1 & 0 & 1 & 0 & 1 \\ \textbf{[1,0]} & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \textbf{[1,1]} & 0 & 0 & .875 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline \end{array}$$

Figure S3: Example of partial barcode matrix for Group 1. *E. cloacae*'s amplicons appeared to always interact with Probes 3 and 4, but a small subcluster appeared to lack the HEX response to Probe 1 (Fig. S1a). We used our SPoRe algorithm to estimate the barcode abundances in reference wells A3 and B10 (Fig. S2) which both contained Probe 1. After manual thresholding, the binarized data and "analytes" with barcodes $[0, 1], [1, 0]$, and $[1, 1]$ (ordered as [HEX, FAM]) were passed to SPoRe, and SPoRe estimated the abundances of the amplicons with these responses. In this case, the $[1, 0]$ quantity was nearly zero, consistent with the expectation that *E. cloacae* always interacted with a FAM probe. The fraction of amplicons with the HEX probe was determined by $\lambda_{[1,1]}/(\lambda_{[1,1]} + \lambda_{[0,1]})$. In A3 and B10, these were estimated to be 0.832 and 0.828, respectively. Our sequence analysis found eight 16S copies in the *E. cloacae* genome, so it is possible that one amplicon had a sequencing error such that Probe 1 truly binds to 7/8 copies. Therefore, column 3 of matrix **C** had 0.875 of barcode [1,1] and 0.125 of barcode [0,1] for Groups 1 and 3.



(a) **(b)** **(c)**

Figure S4: Example process for manual thresholding with noted challenges. **(a)** All reference data (rows A and B in Figure S2) from the same probe group was pooled and displayed to serve as a visual reference. Droplet "rain" is evident in each cluster. Due to some mild "lean" and "lift" of the raw ddPCR clusters caused likely by partial probe interactions, we allowed any linear threshold for each fluorescence channel determined by two user-selected points. **(b)** Raw data from a polymicrobial sample was overlaid on the reference data with the same corresponding probe group. An example is shown with the raw data from D09 (Group 1, $k = 4$ sample number 2) overlaid with the Group 1 reference data. PSC effects, as expected, create subpopulations of droplet measurements between the binary clusters. We speculate that the small additional cluster near zero is due to droplets where probes partially interacted with amplicons due to imperfect sequence homology. **(c)** After the user selects two points to define a line for thresholding, the plot is updated to allow the user to visually confirm the results. Red points are assigned the value 1, and black points are assigned 0.

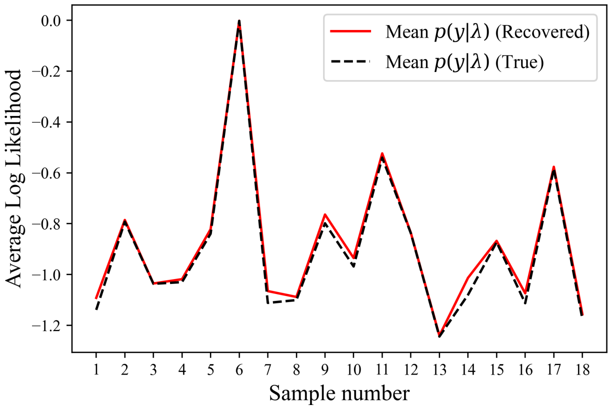


Figure S5: Likelihood comparison of SPoRe's solution against the estimated ground truth. SPoRe's solution exhibits higher average likelihood for the pre-binarized data that it is given.

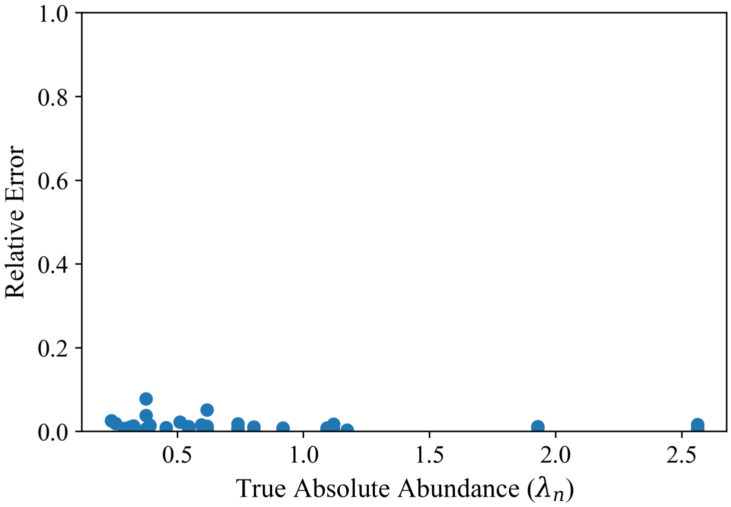


Figure S6: SPoRe's performance on simulations of experimental concentrations. Given the estimated ground truth concentrations ($\boldsymbol{\lambda}^*$), we simulated binary measurement data to pass to SPoRe. SPoRe returns virtually perfect results with mean cosine similarity of 0.9999. Here, we plot the relative error in estimates of the absolute abundance for each bacteria, represented by $|\hat{\lambda}_n - \lambda_n^*|/\lambda_n^*$ when running SPoRe on the simulated measurements.

References

[1] Xing, P. E.; Ng, A. Y.; Jordan, M. I.; Russell, S. Distance Metric Learning, with Application to Clustering with Side-Information. *Adv. Neural Inf. Process. Syst.* **2002**, 521-528

[2] Yakowitz, S. J.; Spragins, J. D. On the Identifiability of Finite Mixtures. *Ann. Math. Statist.* **1968**, *39* (1), 209-214

[3] Yang, L.; Xu, W. A New Sufficient Condition for Identifiability of Countably Infinite Mixtures. *Metrika*. **2013**, *77*, 377-387.