Step 1 Proposal

Antimicrobial Resistance Diagnostics Challenge

Universal Microbial Diagnostics

Pallavi Bugga, Vishu Asthana, Amirali Aghazadeh,
Rebekah A. Drezek, Richard G. Baraniuk

Rice University

Executive Summary

The accurate, efficient, and rapid identification of bacteria and other microorganisms is of mounting importance in the fields of health care, disease management, and environmental monitoring (*1*–*3*). In the absence of a rapid identification method, patients are often treated with broad-spectrum antibiotics, which can produce significant, undesirable side-effects, and promote long-term antibiotic resistance. Treatment with such antibiotics is expensive and generates exorbitant expenditures for the both hospital and the patient.

Currently, the predominant method for bacterial identification in the clinic relies on culturing (a process which can take anywhere from 24-72 hours), generally followed by gram-stain analysis and antibiotic susceptibility testing. While alternative methods of clinical bacterial identification have been developed (e.g., RT-PCR and DNA microarray based diagnostics), these methods are insufficient for the clinic. They either require *a priori* knowledge of the microbial agent (species-specific), have limited sensitivity, or need a long time to produce reliable results which often worsen patient prognosis during the detection/identification period. These limitations reveal a great need to develop a rapid and accurate clinical bacterial identification system.

Here we propose a new microbial diagnostic platform that satisfies the above desiderata. In common with microarrays and PCR-based techniques, our Universal Microbial Diagnostics (UMD) platform exposes a microbial sample (which may contain more than one genus/species) to a set of random, target-agnostic DNA probes. By measuring the degree to which the sample hybridizes with the collection of random probes, we set up a statistical inverse problem to detect the *presence* and *estimate the concentrations* of the various bacteria in the sample. Using signal recovery techniques from the recently developed theory of compressive sensing (*4, 5*), we show below that it is possible to stably solve this inverse problem even when the number of probes is significantly smaller than the size of the library of possible bacteria of interest. The proposed UMD platform is not only *universal*, but also, inexpensive, rapid, and phylogenetically informative (as random probes bind to arbitrary spots on the genome). Moreover, due to the universal nature of its probe design, UMD can classify not only known organisms, but also novel mutants with their closest known relatives, such as clinical strains.

We believe the proposed diagnostic can significantly improve clinical treatment decisions for bacterial infections in in-patient settings, and promote more appropriate use of antibiotics for such cases. Our promising preliminary results have been recently published in *Science Advances* journal[1].

Proposed *in vitro* Diagnostic

A. Description

In our proposed UMD diagnostic, the genomic DNA of an infectious sample is extracted and exposed to a small number *M* of DNA probes, which hybridize to the genomic DNA at various locations; this hybridization is experimentally quantified, producing a probe-binding (or hybridization) vector **y** whose entries correspond to the hybridization binding level of each probe with the microbial sample (Fig. **1A**). *A priori*, the hybridization binding level of each probe to a reference database of *N* bacterial genomes is obtained and stored in an *M*

[1] A. Aghazadeh, et al. "Universal microbial diagnostics using random DNA probes." *Science Advances* 2.9 (2016): e1600025.

× *N* hybridization affinity matrix (Fig. **1B**). The hybridization affinity matrix can be measured either experimentally *in vitro* or predicted computationally *in silico*. To speed up the probe design and prove the concept of UMD, our initial studies predicted the affinity matrix using a thermodynamic model *in silico*.

First, probes of length 38 base pairs and various nucleotide sequences were randomly generated, and their hybridization affinities tested against the genomes of 40 common clinically infectious bacteria with known sequence. Probes whch generated the greatest statistical variance in the number of stable binding were chosen as candidates for experimentation. To compute the entry φ_{ij} in the matrix $\mathbf{\Phi}$, the hybridization binding level of probe *i* to genome *j*, we first performed a rapid thermodynamic alignment of the sequence of the probe to the sequence of the genome using the alignment model described in (*6*).

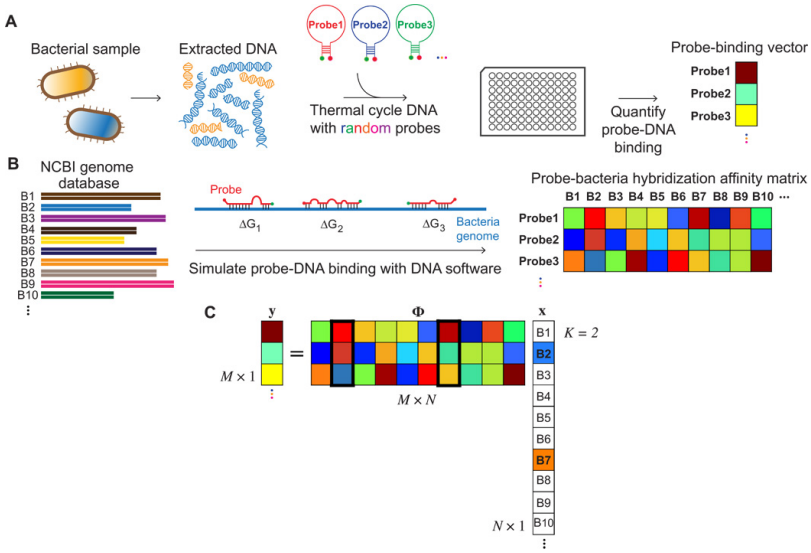

Fig. 1: Schematic of UMD platform. A) Genomic DNA is extracted from a bacterial sample and thermal-cycled with *M* random DNA probes. The genome-probe binding is quantified, producing a probe-binding vector **y**; in this study, the random probes are in the form of MBs, and the DNA-probe binding is quantified by the ratio of open/hybridized to closed/nonhybridized MBs. (**B**) The hybridization binding level of each probe to a potentially large reference database of *N* bacterial genomes (B1, B2, …, BN) is predicted using a thermodynamic model and stored in an *M* × *N* hybridization affinity matrix **Φ**. NCBI, National Center for Biotechnology Information. (**C**) Assuming that *K* bacterial species comprise the sample, the probe-binding vector **y** is a sparse linear combination of the corresponding *K* columns of the matrix **Φ** weighted by the bacterial concentrations **x**, that is, **y** = **Φx** + **n**, where the vector **n** accounts for noise and modeling errors. When *K* is small enough and *M* is large enough, **Φ** can be effectively inverted using techniques from compressive sensing, yielding the estimate for the microbial makeup of the sample **x**; in this illustration, the *K* = 2 bacteria-labeled B2 and B7 are present in the sample.

Next, we extracted sequence fragments from the genome sequence, which contain a significant hybridization affinity with the probe sequence. The fragment-probe mixture was then fed into a thermodynamics-based hybridization model (*6*). This model predicts all possible stable probe-bacteria fragment bindings along with their resulting concentrations for a given set of experimental conditions (Fig. **2B**, **2C**). The overall hybridization affinity φ_{ij} was then computed by summing the concentrations of all predicted and stable probe-fragment bindings for a unit concentration of bacterial genome.

Due to an excess concentration of probes as compared to sample DNA, the probe-binding vector **y** can be closely approximated as a linear combination of the predicted hybridization affinities of the species in the reference genome database (the columns of the matrix) weighted by their concentrations **x**; i.e., **y** = **Φx** + **n**, where the vector **n** accounts for noise and modeling errors (Fig. **1C**). As stated previously, the two key capabilities of the proposed diagnostic platform are to (i) detect the presence of and (ii) estimate the concentrations (**x**) of a potentially large number *N* of reference bacterial genomes in an infectious sample, given only a small number *M* of probe-binding measurements **y**. While solving for **x** would normally be tedious, we can apply the principle of compressive sensing to simplify this problem. The use of compressive sensing in our system is dependent on two requirements: 1) that the clinical sample be sparse, containing only a few bacterial species (or unique genomes) at a time, and 2) that the probes be incoherent, such that their binding profiles to all bacterial genomes in our library be sufficiently dissimilar. Having fulfilled these two requirments, we find that **x** can be easily solved, and the respective concentrations of each of the bacterial species in the clinical sample determined.

B. Development of Approach

To test and validate the efficacy of our compressive sensing universal probe system, we recovered pathogenic bacteria using random probes in the form of mismatch-tolerant sloppy Molecular Beacons (Sloppy MBs) (*7*). Previous studies that have utilized MBs in bacterial detection (*8*) have designed the loop

sequence specific to certain regions (e.g., 16S rDNA) within a single bacterium (*9*) or multiple bacteria (*7*). In our system, the loop sequence is selected as a unique, random, target-agnostic sequence of length 38 nucleotides (Fig. **2A**). The 4-nucleotide long stem sequence is, however, consistent across all MB probes. The unusually long loop and short stem enable our random probes to form hybrids with several base pair mismatches across the entire bacterial genome and compensate for the lower signal intensity in the absence of DNA amplification methods such as PCR (Fig. **2C**).

In our first *in vitro* experiment, five UMD MBs designed *in silico* (as shown in Fig. **2A**, with GC-contents 50, 56.5, 60.8, 50, and 52.7%, identical melting temperature of 40°C, and concentration 1*m*M) were separately mixed with genomic DNA isolated from nine human infectious bacterial strains grouped into three categories: I. Exact sequence known (*Escherichia coli*, *Francisella tularensis*, *Staphylococcus aureus*, *Campylobacter jejuni*, and *Proteus mirabilis*), II. Exact sequence unknown (*Cupriavidus metallidurans* and *Micrococcus luteus*), and III. Clinical isolates, whose exact sequence is unknown (*Bacteroides fragilis* and *Enterobacter aerogenes*) (The identification of *Pseudomonas aeruginosa* and *Bifidobacterium dentium* strains were tested using four random probes). For bacteria in groups II and III, the DNA sequences in the database might not exactly match the sequences present in the bacterial samples. The five probes were conjugated on either end to Cy3 and Cy5. In the closed state, the MBs function as a Cy3/Cy5 donor-acceptor FRET pair, with the resultant fluorescent emission produced principally from Cy5. In the open state, however, the fluorophore pair spatially separates, and thus only produces fluorescence from Cy3 emissions. Using these properties, the number of MB probe bindings to any given bacterial genome can be determined via quantification of Cy3 and Cy5 fluorescent intensities, followed by subsequent calculation of FRET ratios (Cy5/(C3+Cy5)). The protocol for using MB probes in our experiments was as follows: first, genomic DNA was extracted and purified from a sample containing bacteria. Once the genomic DNA had been extracted and residual RNA degraded, the target agnostic MB probes were added into solution and allowed to hybridize to regions of complementarity on the bacterial genome overnight. Following hybridization, the fluorescent intensities of Cy3 and Cy5 were measured via a fluorometer, and FRET ratios calculated. The FRET ratios from binding of the nine bacteria to the MBs are depicted in Fig. **3A**.

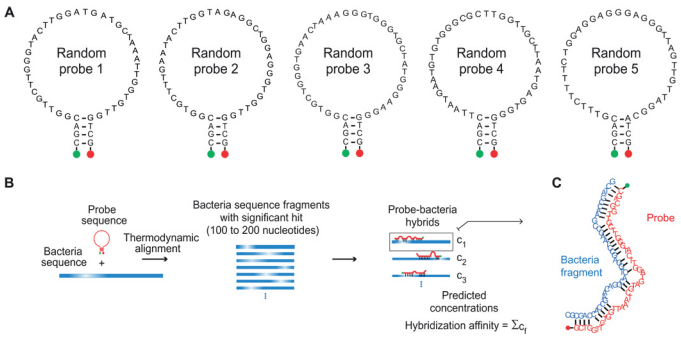

Fig. 2: Random probe design and hybridization affinity computation process. (**A**) DNA sequence structure of five test random DNA probes. (**B**) Both strands of the bacterial genome (blue lines) are first thermodynamically aligned with the probe sequence. The sequence of the bacteria is segmented into fragments of roughly equal length (~100 to 200 nt), each containing a significant hybridization affinity with the probe. Then, all of the bacterial fragments and probe sequences along with the experimental conditions are fed into the DNA software (*18*) to predict all stable probe-bacteria complexes and concentrations. These concentrations, in aggregate, determine the concentration of opened MBs, which is defined as the hybridization affinity of the probes to the bacterial genome. (**C**) Example of a predicted probe-bacteria fragment binding with many base pair mismatches.

To estimate the bacterial concentrations in physical units, we translated the FRET ratio of each bacterium-MB pair into a hybridization affinity, represented in units of molarity. For this, we experimentally obtained and fitted FRET ratios for each of the five MBs as a function of the concentration of their exact probe complements, using an optimization method described in (*10*) (See fit curve methods in Appendix Ic). The R^2 values for the fits ranged from 0.97 to 0.99, suggesting a satisfactory fit. Based on the fit equations, the hybridization affinities corresponding to the FRET ratios for all bacteria were calculated. We refer to these measurements as the measured *hybridization affinity vectors* and show them in Fig. **3B**.

Our challenge here was to decode the experimentally measured affinities of the bacterial species samples reacting to UMD MBs using compressive sensing recovery techniques. With the predicted hybridization affinities of $N = 9$ bacteria to $M = 5$ random probes stored in the computationally obtained $\mathbf{\Phi}_{5\times9}$, we used a variant of the Orthogonal Matching Pursuit (OMP) algorithm (*11*) to successfully identify the species present in each of the samples (Fig. **3**), and estimate the relative concentrations of each of the species within the sample. Using this algorithm, we were able to determine the concentrations of each of the nine bacterial species with an average error of 11.5%.

To further assess the power of our diagnostic, we attempted to detect the same nine test species from a more comprehensive list of pathogens using our five DNA probes. To do this, we expanded the reference genome database to contain 40 genera (i.e., $M = 5 \ll N = 40$), including bacterial pathogens listed by the Centers for Disease Control and Prevention (CDC) as the most common notifiable human diseases (*12*) (Appendix II) including the 18 bacteria of the highest-concern in Table 3 of the "National Action Plan for Combating Antibiotic Resistant Bacteria" booklet. With the most common pathogens' genomes in the database, the detection performance remained above AUC = 0.84, suggesting a high recovery rate with only five random probes.

Our initial *in vitro* experiments demonstrated successful identification of eleven infectious strains using a fixed set of five randomly selected test probes. To further validate the proposed diagnostic, we ran additional studies *in silico.* Through these simulation studies, we numerically demonstrate that, if a sufficient number of probes is used, then any group of randomly selected probes will detect the presence of one ($K = 1$) or a mixture of several ($K = 2, 3, \ldots$) pathogenic organisms in a sample out of a database of 40 pathogenic organisms. Additive white Gaussian noise was introduced to the simulated hybridization affinity vectors to capture the variance in the hybridization affinities among the independent test trials in Fig. **3.** The noise levels were extrapolated from the above eleven test bacteria experiments, with the noise variance set to $\sigma_0 = 2.4 \times 10^{-8}$ M. To control for differences in the genome size of each organism, we normalized numerical simulations to unit weight of bacterial DNA.

In Fig. **4A**, we first demonstrate the detection performance of UMD in identifying a single bacterium ($K = 1$) among the pathogen database at different noise levels. As the ROC curves suggest, UMD's detection performance improves when the noise variance decreases. With only a five-fold decrease in the noise variance, UMD identifies all 40 bacteria in the database almost perfectly (AUC > 0.95) using only five randomly selected probes. Further improvement in screening performance is observed when the number of random MB probes used is increased. Fig. **4B** demonstrates that UMD identifies all 40 bacteria in the CDC database almost perfectly (AUC = 0.95) with *any* $M = 15$ randomly selected MBs when the noise variance is similar to that measured experimentally (Fig. **3**).

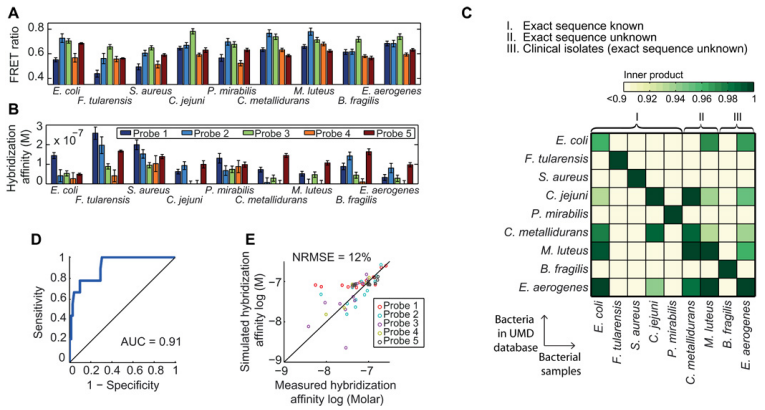

Fig. 3: Binding patterns of five random probes correctly identify the bacteria present in nine diverse bacterial samples. (**A**) Experimentally measured FRET ratios to quantify hybridization between bacterial DNA and probes 1 to 5. (**B**) Hybridization affinity between DNA samples and probes converted from FRET ratio through the probe characteristic curve fit equations **C**) Heat map of normalized inner products between the experimentally obtained hybridization affinity and predicted hybridization affinities (by thermodynamic model) for nine DNA samples as a measure of the similarity of the probe measurements to the bacteria in the data set. DNA samples are clustered into three groups: (i) exact sequence known, (ii) exact sequence unknown, and (iii) clinical isolates (whose exact sequence is unknown). UMD correctly recovers the diagonally highlighted bacterium (with inner product >0.9). (**D**) The average ROC curve of UMD in detecting nine bacteria, assuming the independence of the different experiments. Each point on the curve corresponds to a threshold value between [−1,1]. UMD achieves high values of the AUC (AUC > 0.9). (**E**) Correlation of measured and simulated hybridization affinities and the NRMSE of the prediction (straight line corresponds to maximum correlation). All experiments were performed in triplicate, and the results shown here average over the trials with the error bars representing SEM.

UMD has the unique advantage that it can recover more than a single ($K > 1$) organism in an infectious sample. To evaluate the minimum number of probes M required for this task, we used the Basis Pursuit De-Noising (BPDN) algorithm to identify the composition of a sample containing $K = \{2,3\}$ equi-concentration bacterial species (**Fig. 4C**). We found that any set of $M = 15$ randomly selected probes will recover all possible mixtures of $K = \{2,3\}$ pathogenic species in the CDC database. This result confirms that the incoherence requirement for compressive sensing is empirically satisfied for the pathogenic strains in the CDC database and thus that UMD is capable of screening for pathogenic bacteria at the genus-level.

To evaluate the UMD's performance at the species level, we analyzed 24 species of *Staphylococcus* genus and 23 species of *Vibrio* genus *in silico* (Appendix II). Through our system, we were able to identify the composition of samples containing *Staphylococcus* species using 11 random probes (fig. **5A**) and the

composition of samples containing *Vibrio* species using 18 random probes (fig. **5B**) with high sensitivity and specificity (AUC > 0.95). This underscores UMD's potential to differentiate pathogens at high taxonomic resolution.

Using the UMD platform, one can tradeoff between universality (detecting species outside of the library) and cost efficiency (number of probes). That is, it is possible to select a set of probes that achieves better detection performance in terms of specificity and sensitivity than the average performance of random probe sets at the cost of universality. Given a set of *P* random probes however, finding the set of *M* probes with the best detection performance in terms of sensitivity and specificity is an extremely challenging problem. A brute-force search would require one to search among all $\binom{M}{P}$ possible combinations of *M* probes to find the optimal probe set. This combinatorial search algorithm grows quadratic with *P* and thus becomes computationally intractable when the number of probes grows. We thus developed a rapid probe selection method that we dub GPS. With a small sacrifice in sensitivity, GPS finds the best performance probe in a few seconds: exponentially faster than the naïve search method. The algorithm in each iteration finds the probe that maximizes a detection performance criterion (here, the maximum pairwise correlation of bacteria) and adds it to the list of probes picked from the previous iterations. GPS stops when the maximum desired number of probes is reached. Using this algorithm, we created a set of 3 and 10 probes to test against the reference database of 40 genomes. As shown in Figures **4D** and **4E**, the false positive rate drops for all of the bacteria in the database as the number of probes increases from *M* = 3 (AUC > 0.95) to *M* = 10 (AUC > 0.99). While the performance detection is high (AUC > 0.99), the confusion matrix shows few cases where the inner product values for possible species (e.g., *Coxiella*, *Aeromonas*, and *Proteus* when the actual sample contains *Coxiella*) are only slightly separated. Greater separation between inner product values for candidate bacterial species can be achieved by using a larger number of UMD probes. This can increase the robustness of the UMD system and ensure a lower false positive rate for noisier environments.

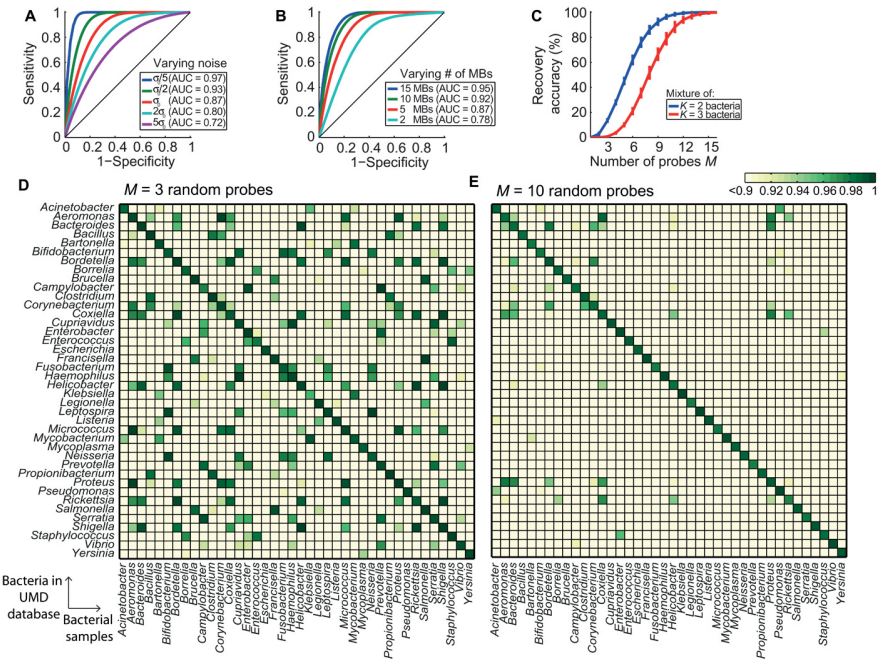

Fig. 4: Performance of UMD platform in genus-level recovery of 40 species listed as the most common human infectious genera by CDC with different number of random probes *M* and noise variance σ.(**A**) The ROC curve in detecting single bacterium (*K* = 1) with different noise levels. $\sigma_0 = 2.4 \times 10^{-8}$ M denotes the variance of the additive white Gaussian noise used in the simulation. This value is obtained from the experiments in Fig. 3 by calculating the propagated variance of measured FRET ratios. UMD performs more accurately with lower noise variance. The detection is almost perfect (AUC > 0.95) under noise variance $\sigma = \sigma_0/5$. (**B**) The average ROC curve in detecting single bacterium using different number of random probes *M* and fixed noise variance $\sigma = \sigma_0$. The detection performance universally improves over all the 40 species by increasing the number of random probes. With 15 random probes, UMD achieves almost perfect detection performance (AUC > 0.95). (**C**) The percentage of simulated trials, where *K* bacteria present in the samples were recovered correctly with zero false positives, among all possible $\binom{40}{K}$ bacteria mixtures (blue and red curves corresponding to *K* = 2 and *K* = 3 bacteria, respectively). Simulations were repeated 1000 times with randomly selected MBs, and error bars represent SD. (**D** and **E**) Confusion matrices illustrating the detection result of UMD using *M* = 3 and *M* = 10 probes selected by the GPS algorithm.

While mainly intended to rapidly screen for pathogens at higher taxonomy levels, the UMD platform is also capable of identifying bacteria at the strain-level when using GPS MB probes. In Fig. **6** we demonstrate that GPS selects UMD probes that differentiate among 9 strains of *E. coli* (8 pathogenic and one nonpathogenic) with high detection accuracy (AUC > 0.95) *in silico*.

C. Challenges

As exhibited by the results of our previous studies, the proposed *in vitro* diagnostic demonstrates exquisite accuracy in identifying one, or a mixture, of unique bacterial genomes in a given clinical sample. As the number of unique species in the clinical sample increases however (or as **x** becomes less sparse), the predictive accuracy of the system begins to drop off, as demonstrated by Fig. **4C**. While accuracy can to some degree be recovered by the inclusion of more MB probes, the system eventually reaches a predictive threshold (around 3 equi-concentration bacterial species), beyond which it is unable to accurately identify all species in the clinical sample. This limitation is not necessarily restrictive for most infections encountered in the clinic, but poses a difficult challenge for the diagnosis of many gastro-intestinal (GI) and anaerobic infections. It should be noted that the presence of additional endogenous species in a sample is not necessarily confounding if the concentrations of the primary infectious bacterial species are markedly greater than the concentration of those endogenous species. Given that this is not the case for GI and anaerobic infections however, since multiple bacterial species (>3) of equal concentration are present in the isolated sample, it is clear that our system would struggle in accurately identifying the infectious source in patients suffering from these conditions. While we are currently working on improving our system to allow for better screening of such infections, there is still strong evidence to support the success of our proposed diagnostic in dealing with the clear majority of other infections that do not exhibit this type of presentation.

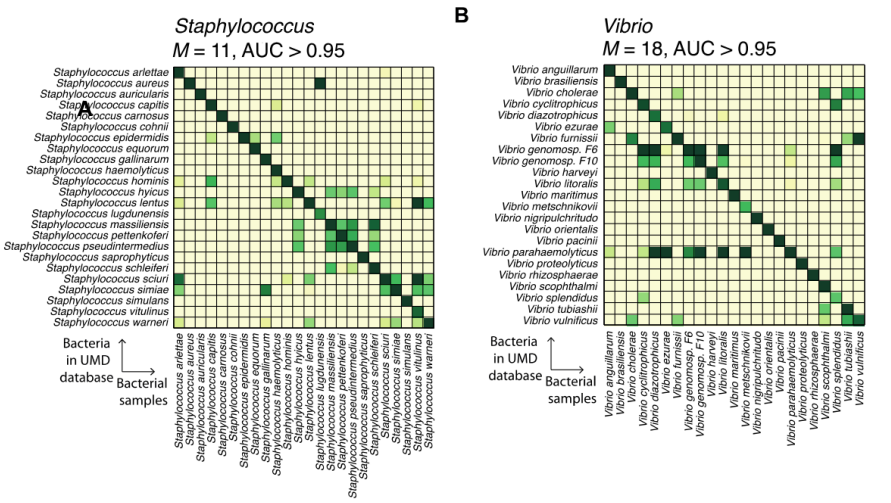

Fig.5: Performance of UMD in species-level recovery of 24 strains of Staphylococcus and 23 strains of Vibrio. (A) UMD's confusion matrix in identifying 24 strains of Staphylococcus. With M = 11 random probes, UMD identifies all the species of Staphylococcus with AUC > 0.95. **(B)** UMD's confusion matrix in identifying 23 strains of Vibrio. With M = 18 random probes, UMD identifies all the species of Vibrio with AUC > 0.95.

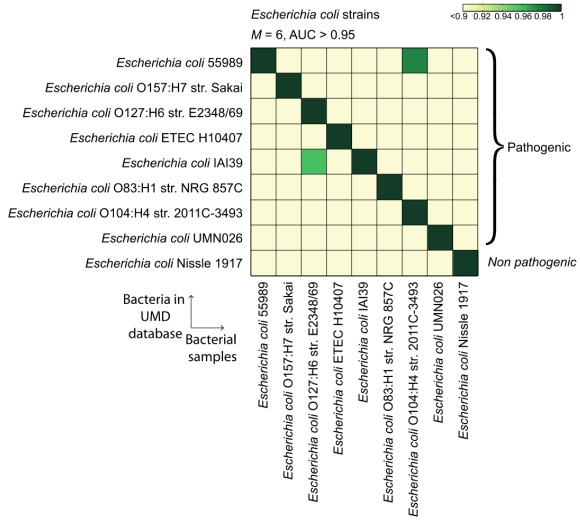

Fig. 6: Performance of UMD in identifying eight pathogenic and one nonpathogenic *E.coli* strains using GPS probes. GPS selects 6 UMD probes that differentiate between eight pathogenic and one nonpathogenic *E.coli* strains in silico (AUC >0.95).

An additional challenge that we are currently attempting to address involves improving the resolution of our diagnostic to allow for identification of bacteria at concentrations that are clinically relevant. In our initial experiments, the concentration of bacteria used was 1 micro-molar—a concentration that is markedly higher than the clinical threshold for a large majority of infections. Thus, we are currently working towards improving our system to allow for the identification of bacteria at lower limits of detection.

D. Risk

The *in vitro* diagnostic we have proposed here is highly accurate and maintains only a minimal degree of error (Fig. 3D, 3E, 4). As shown in Fig. 4B and 4C, this percent error decreases as the number of probes used increases. Thus, the risk of mis-informing physicians as to what the optimal choice of antibiotic may

be for a given infection remains low. As far as risk to the patient is concerned, the diagnostic we have proposed here only requires manipulation of the isolated microbial sample, and thus bears no additional danger to the patient (when compared to the current gold standard).

"State of the Art" Statement

(1) Conventional strategies for microbial detection are based on microbe-specific genomic or proteomic markers and protocols. Polymerase chain reaction (PCR)-based approaches rely on the binding of specific capture probes with unique genomic identifiers, such as the 16S ribosomal DNA subunit in bacteria. While these methods show promise as highly specific tools for microbial identification (*9*, *13*), they have limitations in clinical, industrial, and defense settings (*14*, *15*). In the case of an epidemic, the detection of a newly mutated species using current PCR methods would require entirely new capture probes to be manufactured, introducing additional costs and delays. For bacterial detection, blood cultures typically require 48 to 72 hours to produce reliable results (*16*–*19*). During this waiting period, administration of broad-spectrum antibiotics breeds further threats of bacterial resistance and missed coverage (*20*). DNA microarrays also require many target-specific probes to detect multiple pathogens and lie dormant against unknown organisms. Whole genome sequencing (WGS), currently the most complete and accurate technique, is not yet conducive to point-of-care diagnostics; it requires millions of expensive sequencing reads to assemble or align with genomic identifiers. It follows that there is a critical need for a new means of microbial detection: a *universal* (i.e., works for bacteria outside of the target library), *inexpensive* (i.e., requires minimal resources for acquisition such as DNA probes and sequencing reads, etc.), and *rapid* sensing platform capable of identifying known and novel species with high phylogenetic power.
(2) We propose on the design and validation of a new microbial diagnostic platform that satisfies the above desiderata. In common with microarrays and PCR-based techniques, our Universal Microbial Diagnostics (UMD) platform exposes a microbial sample (which may contain more than one genus/species) to a collection of DNA probes. In sharp contrast to conventional methods, however, the probes are randomly generated (and hence target-agnostic) permutations of nucleotides (nts) that freely hybridize to different spots and to different extents on different bacterial genomes. By measuring the degree to which the sample hybridizes with the collection of random probes, we set up a statistical inverse problem to detect the presence and estimate the concentrations of the various bacteria in the sample. Using signal recovery techniques from the recently developed theory of *compressive sensing* (*4*, *5*), we show below that it is possible to stably solve this inverse problem *even when the number of probes is significantly smaller than the size of the library of possible bacteria of interest*.

The UMD platform possesses numerous advantages over other currently available microbial diagnostics. First, the UMD probes used are universal in the sense that a fixed set of probes captures the salient information required to distinguish between members of a large and growing database of species (*21*). This gives UMD a potentially important future proof property: a fixed set of measurement probes can be used to detect and estimate the concentration of newly sequenced species not yet present in the library. To detect a new organism, the software merely has to be adjusted to take into account how the new organism will react to the existing probe set. Thus, the design of new capture probes is not required. Moreover, since the number of probes grows only logarithmically in the size of the library, the UMD platform naturally contends with the data deluge (*22*) from new microbial species being discovered and sequenced every day. Second, the UMD platform is largely inexpensive compared to other microbial systems. Unlike PCR and microarrays, which require specific primers for each unique bacterial species, the UMD system requires only five to ten probes to search through the whole database of clinically relevant bacterial species with known sequence. Third, the proposed diagnostic is significantly more rapid compared to other universal systems (i.e. culturing), requiring only 24 hours of processing time from beginning to end. It can potentially be implemented on a microfluidic device with a DSP chip that does the signal processing computation. The proposed lab-on-a-chip structure can produce reliable results potentially in few minutes. Lastly, the proposed *in vitro* platform is *phylogenetically informative*. Given its universality, this diagnostic can also be adapted for the characterization and taxonomic classification of new pathogenic strains—which arise either through natural evolutionary processes or scientific manipulation (bio warfare)—by analyzing similarities (via nearest neighbor algorithms) in probe binding patterns to those of known genomes. To the best of our knowledge, UMD is the only technique that enables a unified representation of bacterial organisms in a low-dimensional geometric space. The theory of compressive sensing provides rigorous recovery and suggests algorithms to leverage this geometry to both detect bacteria and estimate their concentrations efficiently.

Our successful implementation of UMD confirms that a small number of random DNA probes satisfy the incoherency requirements of compressive sensing theory and can be used for rapid, universal microbial sensing.

(3) The rapid and accurate identification of infectious bacterial organisms in the clinic remains a significant challenge for physicians and researchers. Due to the inability to rapidly identify these infectious agents, patients are often treated with broad-spectrum antibiotics, which can produce significant, undesirable side-effects, and promote long-term antibiotic resistance. Treatment with broad-spectrum antibiotics is also expensive, generating exorbitant expenditures for the both hospital and the patient. Currently, the predominant method for bacterial identification in the clinic relies on culturing (a process which can take anywhere from 48-72 hours), generally followed by gram-stain analysis and antibiotic susceptibility testing. While alternative methods of clinical bacterial identification have been developed (i.e. RT-PCR and DNA microarray based diagnostics), these methods are insufficient in that they either require *a priori* knowledge of the microbial agent, are species specific, limited in sensitivity, or too time consuming, often leading to a worsening of patient prognosis during the detection/identification period. The UMD platform we are proposing here is a universal system that can identify clinically infectious bacterial organisms within a 24-hour period (and potentially a few minutes with better implementations in the Step 2 of the challenge). Thus, it has the potential to more rapidly inform physicians of appropriate management protocols (specifically in terms of optimal antibiotic treatments), and expedite improvement of the patient's condition, while subsequently minimizing side-effects borne from broad-spectrum antibiotic administration, and the risk of long-term antibiotic resistance. In this way, our proposed *in vitro* diagnostic can provide physicians with useful direction, and facilitate more rapid and effective clinical decision-making for the treatment of bacterial infections.

(4) Perhaps the most significant quantifiable improvement afforded by our system (over existing diagnostics) is processing time. As stated previously, the current gold standard for universal microbial detection/identification is culturing (a process which, on its own, can take up to 72 hours), followed by various biochemical/susceptibility assays. The diagnostic we are proposing here requires only 24 hours for identification, representing a 66% reduction in processing time. This reduction in identification time is crucial, as it prevents further worsening of patient prognosis, and allows for the development of more effective treatment protocols.

One additional advantage of our proposed system is that it provides the physician or clinician with a metric quantifying the confidence of each identification event. As seen in our initial experiments, the UMD quantitatively measures (and subsequently outputs) the distances between the predicted hybridization affinity vectors and experimental hybridization affinity vectors for each unique bacterium in the clinical sample. Using this information, the user can easily characterize the performance of the UMD diagnostic in terms of false positives and false negatives, and thus gage the overall sensitivity and specificity of the system. In this way, the physician is provided with a "confidence metric" that takes into account the overall accuracy of each bacterial identification event, and allows him or her to rapidly screen through a list of probabilities for each species (in the database) being present in the clinical sample. This information is, to our knowledge, not provided by any other currently available microbial diagnostic.

Lastly, numerous studies conducted by the World Health Organization and CDC have shown that one of the biggest causes of antibiotic resistance is inappropriate administration of antibiotics, either via over-prescribing or mis-prescribing. In approximately 40% of cases (taken from both in- and out-patient), the choice of antibiotic, or duration of antibiotic therapy is incorrect (*23*). One method of addressing this issue is by developing better diagnostic protocols for microbial identification. Our proposed diagnostic aims to do just that—by allowing for more rapid and universal identification of infectious bacteria, the physician can make a more informed decision regarding appropriate antibiotic therapy. Thus, we believe incorporation of our proposed system will produce a potentially quantifiable reduction in inappropriate antibiotic use, and promote more effective clinical decisions for bacterial infection treatment.

Plan to Complete Step 2

To assess the ability of the proposed *in vitro* device to meet the target product profile, we are aiming to run a series of experimental tests. **First, the limit of detection of the device will need to be quantified**. Our initial experiments were conducted using a concentration of ~1 μM of bacteria. This concentration is, however, above the average clinical threshold for clinical infections. To determine the limit of detection, we will test our five probes against decreasing concentrations of *S. aureus* (a BSL-2 bacteria), ranging from 10

pM to 1 μM. **Second, we will need to determine the minimum allowable "dose" (or concentration) of probe** needed to achieve sufficient diagnostic resolution at the limit of detection. Various probe concentrations (ranging from .1 to 5 times the lowest concentration of bacteria) will be tested against *S. aureus*, and analyzed**. Lastly, we will need to calculate the minimum time of analysis** required for accurate identification. Hybridization times (of probe to genome) will be modulated between 3 hours to 16 hours, and again tested in *S. aureus.* For each of the conditions stated above, confusion matrices and ROC curves will be generated to quantitatively assess diagnostic performance. Having established these diagnostic parameters, we believe our proposed device will be ready for prototype development.

The estimated time frame to complete the above testing is approximately 6 months, as two months will be required to run all experiments and post-analysis for each condition. The experimental methodology that will be utilized to conduct these additional tests has already been established and validated in our initial UMD experiments. Further, aside from the purchasing of *S. aureus*, there are no additional materials or pieces of equipment that are needed to complete these studies. Taken all together, we believe that this plan will be easily executable in the given timeframe.

Protection of Human Subjects

No human specimens, data or material will be used in this study. All bacteria used in Step 2 will be purchased through American Type Culture Collection (ATCC).

Inclusion of Women, Children, and Minorities

As stated above, no human specimens, data, or material will be used in this study.

Compliance with Policies Related to Vertebrate Animals

Given that no vertebrate animals have or will be used in any part of this study, compliance with this policy is automatically fulfilled.

Compliance with Biosafety Issues

All bacteria purchased will be classified as BSL-2 or lower. Our lab is currently classified as a BSL-2 lab, and has (and will continue to follow) applicable BSL-2 safety regulations.

Compliance with Patents or Intellectual Property Protection

The diagnostic platform described in this LOI is currently in the early stages of acquiring intellectual property protection. As of the writing of this proposal, Rice University plans to file a provisional patent for this diagnostic, and a disclosure is currently in the process of being prepared.

Appendix:

(I) *Materials and Methods*

a) Random DNA probe construction and preparation

To implement the Universal Microbial Diagnostics (UMD) platform, we obtained DNA oligonucleotides for the random DNA probes and their exact complements from Integrated DNA Technologies (Coralville, IA). The sequences are provided below. $MgCl_2$, KCl, and sterile nuclease-free water for making the molecular beacon (MB) buffer were purchased from Fisher Scientific (Waltham, MA). 1M Tris-HCl solution (pH 8.3) and Tris-EDTA buffer (TE buffer; 10 mM Tris-HCl, 0.1 mM EDTA pH 8) were obtained from Teknova (Hollister, CA). To prevent nuclease contamination, all work surfaces and materials were routinely cleaned with RNAse OFF decontamination solution (Takara, Japan).

Random probe 1: 5'- /5Cy5/CGA CGG TTG CTT GGG TAC TTG GAT GAT GCT AAA TTG GTG TTG GTC G/3Cy3Sp/ -3', Random probe 2: 5'- /5Cy5/CGA CGG TGC TTT GAA TAC TTG GTA GAG GCT GGA GGG TGG TTG GTC G/ 3Cy3Sp/ -3', Random probe 3: 5'- /5Cy5/CGA CGG TGC TGG GTG AAC TAA AGG GTG GGT GCT ATG GGA AGG GTC G/3Cy3Sp/ -3', Random probe 4: 5'/5Cy5/CGA CTT AAT GAA TGT GTG GGC GCT TGG TTG CTT AAT GAG TGG GTC G/3Cy3Sp/ -3', and Random probe 5: 5'- /5Cy5/CGA CGT TTC TTT TCT GGA GGA GGG AGG GTT AGT TGT TAG GCA GTC G/3Cy3Sp/ -3'.

Random probe complement 1: 5'- CGA CCA ACA CCA ATT TAG CAT CAT CCA AGT ACC CAA GCA ACC GTC G -3', Random probe complement 2: 5'- CGA CCA ACC ACC CTC CAG CCT CTA CCA AGT ATT CAA AGC ACC GTC G -3', Random probe complement 3: 5'- CGA CCC TTC CCA TAG CAC CCA CCC TTT AGT TCA CCC AGC ACC GTC G-3', Random probe complement 4: 5'- CGA CCC ACT CAT TAA GCA ACC AAG CGC CCA CAC ATT CAT TAA GTC G-3', and Random probe complement 5: 5'- CGA CTG CCT AAC AAC TAA CCC TCC CTC CTC CAG AAA AGA AAC GTC G-3'.

b) Random DNA probe design

In the design of molecular beacons (MBs) (*20*) for random DNA probes, the length and GC content (ratio of G+C to other nucleotides) of the probe loop and stem sequences were considered to strike a balance between two factors: fluorescence signal level and probe stability. Signal intensity was especially important in our detection scheme, since no DNA amplification method (such as PCR) was utilized. Similar to sloppy molecular beacons (sloppy MBs) (19), we selected the random probe loop sequence to be longer that traditional MBs. In addition, we made the stem sequence one nucleotide shorter to introduce additional sloppiness (i.e., hybridization in presence of more base-pair mismatches).

Our challenge was to find probes that maintain the MB's signature hairpin structure over a wide range of temperatures (4-50 °C) after introducing additional sloppiness. To produce random MBs we followed the following procedure: We first generated one million random sequences of length 46 nucleotides with fixed stem sequences on both ends (Fig. **2A**). Then, we used a package in the DNA software (Visual OMP DE) to generate all the possible stable and secondary structures of the sequences in the experimental thermodynamic conditions. We parsed the output of the DNA software and filtered out the probes with undesired secondary structures or melting temperatures. By no means is this the only method to generate random probes for a UMD platform; any method that produces probes with a stable hairpin structure and uniform melting temperature while providing the required signal intensity can be utilized.

c) Generation of random probe characteristic curves

The experimentally measured fluorescence resonance energy transfer (FRET), defined as the ratio of Cy5 intensity over total fluorescence intensity (Cy3 + Cy5), is a function of the concentration of open random probes in the solution, i.e., the probe-target hybridization affinity. To discern the hybridization affinity between a probe and target in units of molarity rather than as a FRET ratio, a characteristic curve was constructed for each probe. These curves presented the FRET ratio as a function of the concentration of open probes in molarity.

To obtain the characteristic curves, random probes were diluted to 1μM in 1 X MB buffer (4 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl, pH = 8, in sterile RNAse free water). DNA oligonucleotides perfectly complementary to the random probes were diluted using 1 X TE buffer to 10^{-5}, 10^{-6}, 8×10^{-7}, 6×10^{-7}, 4×10^{-7}, 2×10^{-7}, 10^{-7}, 8×10^{-8}, 6×10^{-8}, 4×10^{-8}, 2×10^{-8}, 10^{-8}, 10^{-9}, 10^{-10}, or 10^{-11} M concentration. 25 μL of 1 μM random DNA probes (diluted in MB buffer) were added to 25 μL of perfect complement DNA of various concentrations, or to the TE buffer-only control.

The DNA mixture was briefly centrifuged with a mini centrifuge (VWR) to collect all DNA to the bottom of the tube. Then the DNA was hybridized using a MyCycler Thermal Cycler (Bio-Rad) under the following conditions: 95 °C for 5 minutes, 50 °C for 2 minutes, 30 °C for 1 minute, 20 °C for 1 minute, and 4 °C for 2 minutes. 45 μL of each thermal cycled mixture was added to 155 μL 1 X MB buffer in a black flat bottom 96

well plate (Corning) and kept at 4 °C overnight. A non-linear optimization algorithm (22) was utilized to fit the parameters *a*, *b*, *n* and $FRET_0$ to the characteristic curve: $FRET(c) = FRET_0 + a / (1 + b (10^{-6} - c)^{-n})$.

d) Bacterial DNA extraction

Overnight cultures of *S. aureus* USA 300 and *E. coli* MG1655 were used to inoculate fresh cultures grown in 50-100 mL 2xTY or Luria-Bertani (LB) broth, respectively. *F. tularensis* LVS was obtained from Dynport Vaccine Company LLC (derived from NDBR101 Lot 4) and grown in modified Mueller-Hinton cation-adjusted (MHII) broth (Becton Dickinson) supplemented with sterile 0.1% glucose, sterile 0.025% ferric pyrophosphate, and 2% reconstituted IsoVitaleX (Becton Dickinston). Cultures were pelleted and washed three times with sterile PBS. To release chromosomal DNA cells were resuspended in TE buffer and mixed with 10% sodium dodecyl sulfate (SDS) and Proteinase K at 65 °C overnight. DNA was isolated using phenol: chloroform and precipitated via ethanol precipitation (protocol adapted from elsewhere (30)). DNA pellets were resuspended in 50 *μ*L TE buffer and stored at -20 °C.

Bacterial strains, *C. jejuni*, *P. mirabilis*, *C. metallidurans*, *M. luteus*, *B. dentium*, *E. aerogenes*, *B. fragilis*, and *P. aeruginosa* were grown overnight in 30 mL Brain Heart Infusion (BHI) media (BD) at 37 °C. Bacterial cells were pelleted, washed two times with sterile 1X PBS, and resuspended in TE buffer. Proteinase K (1 mg/mL) (Sigma) and 0.5% SDS were added to the bacterial cells, which were then incubated overnight at 55 °C on an orbital shaker. The samples were then mixed with phenol-chloroform (Invitrogen) and centrifuged; supernatants were transferred to a fresh tube. This aqueous phase was then mixed with an equal volume of chloroform and centrifuged (and repeated). Finally, 1/10 volume of 2 M sodium chloride and an equal volume of isopropanol was added to precipitate the DNA. This mixture was incubated at -20 °C for 30 min and centrifuged. The pellets were rinsed with 70% ethanol, air dried, and resuspended in TE buffer.

e) Random probe and bacterial DNA hybridization

Bacterial DNA was diluted to approximately 500 ng/*μ*L using TE buffer and kept at -20 °C until use. The random MB probes were diluted to 1 *μ*M in MB buffer prior to use. 25 *μ*L of 1 *μ*M random probes (diluted in MB buffer) was added to 25 *μ*L of TE buffer control and *E. coli*, *F. tularensis*, *S. aureus*, *C. jejuni*, *P. mirabilis*, *C. metallidurans*, *M. luteus*, *B. dentium*, *E. aerogenes*, *B. fragilis*, or *P. aeruginosa* DNA. The DNA mixture was briefly centrifuged to collect DNA and then hybridized using a MyCycler Thermal Cycler (Bio-Rad) under the following conditions: 95 °C for 5 minutes, 50 °C for 2 minutes, 30 °C for 1 minute, 20 °C for 1 minute, and 4 °C for 2 minutes. 45 *μ*L of each thermal cycled mixture was added to 155 *μ*L 1 X MB buffer in a black flat bottom 96 well plate (Corning) and kept at 4 °C overnight.

f) Measuring FRET ratio through fluorescence as indicator of random probe-bacteria hybridization

The FRET ratio for the genomic DNA samples following hybridization with each of the random probes was determined by reading the Cy3 and Cy5 fluorescence using a Fluorolog-3 spectrofluorometer (Jobin Yvon Horiba, Edison, NJ) coupled with a MicroMax 384 MicroWell Plate Reader and water-cooled PMT detector. Samples were excited at 545 nm, and single point fluorescence measurements were taken with at 562 nm and 677 nm emission (optimal wavelengths determined through excitation-emission matrix analysis) to measure the Cy3 and Cy5 fluorescence, respectively. The FRET ratio was calculated as Cy5 / (Cy5 + Cy3).

g) Determining DNA hybridization affinity via SantaLucia thermodynamic models

A comprehensive thermodynamic model by SantaLucia et al. (*18*) was utilized to predict the hybridization of probes to bacterial genomes. The SantaLucia model incorporates thermodynamic parameters for mis-hybridizations between two DNA sequences. We utilized two software packages: ThermoBlast DE, which performs fast alignment of sequences against large genome databases to discover thermodynamically stable hybridizations, and Visual OMP DE, which simulates hybridization experiments with detailed solution conditions and generates results for melting temperature (Tm), Gibbs free energy (DG), and the percentage-based concentration of each resultant species post-experiment. The secondary structure of each monomer, homodimer, and heterodimer species formed from the constituent probes and target fragments can also be visualized.

To calculate the hybridization affinity of a genome to a probe, we first used the ThermoBlast package and thermodynamically aligned the sequence of the random probe to both complement strands of the target genome. We extracted all the sequence fragments of the genome (100-200 nucleotides) that aligned with the probe sequence with a predicted melting temperature within approximately 35 °C of the melting temperature of the sequence genome. We then used Visual OMP DE to simulate the hybridization between the probe and the target genome, using the target fragments (Fig. **2**). Every simulation contained information on the probe sequence, the target fragment sequences, and conditions for the experiment, including probe concentration (1 *m*M), unit target concentration (500 ng/*m*L for all bacteria), assay

temperature (4 °C), hybridization buffer composition (4 mM Mg++, 50mM Na+, 0 M Glycerol, 0 M DMSO, 0 M Formamide, 0 M TMAC, 0 M Betaine), and pH (= 8). This procedure was repeated for each probe-target genome pair. We used the percentage of probe-target heterodimer structures formed, i.e., the percentage of probes that are bound to target fragments, as an estimate for the hybridization affinity of the probe to each target (Fig. **2B**).

h) Linearity assumption considerations in UMD

In the UMD platform, the probe concentration (1×10^{-6} M) is in far excess of the target concentrations (~1×10^{-10} M); therefore, we are able to linearly combine the hybridization affinity signatures that we measure for individual targets using the hybridization model. Due to the flooding of excessive number of probes, each target fragment has its choice of binding/not binding to the probes, and thus we can safely sum together multiple target interactions of the same probe, assuming them to be independent.

i) Receiver-operator curve (ROC) analysis

Receiver-operator curve (ROC) analysis was performed by plotting a ROC curve showing the sensitivity and (1-specificity) for 1000 threshold values ranging from -1 to 1. For each threshold value, the following procedure was performed on the data matrix of normalized inner products between the experimentally obtained hybridization affinity and predicted hybridization affinities (by thermodynamic model) for the nine independent bacterial DNA samples (Fig. **3C**): Each entry in the inner product data matrix was compared with the threshold value to determine the number of true positives, false positives, true negatives, and false negatives. True positives were identified when values in the diagonal entries of the inner product data matrix were greater than the threshold value since diagonal entries represent the correct classification of the bacterial sample with its corresponding genus in the database. False positives were identified as off-diagonal values that were greater than the threshold value. True negatives were identified as off diagonal values that were less than or equal to the threshold value. False negatives were identified as diagonal values that were less than or equal to the threshold value. For each threshold value, sensitivity was defined as (# true positives / (# true positives + # false negatives)) and specificity was defined as (# true negatives / (# true negatives + # false positives)).

j) Greedy probe selection

Given a set of *P* random probes, finding the set of *M* probes with the best detection performance in terms of sensitivity and specificity is an extremely challenging problem. A brute force search would require one to search among all possible combination of *M* probes to find the optimal probe set. This combinatorial search algorithm grows quadratic with *P* and thus becomes computationally intractable when the number of probes grows. We thus developed a rapid probe selection method that we dub Greedy Probe Selection (GPS). With a small sacrifice in sensitivity, GPS finds the best performance probe in a few seconds: exponentially faster than the naïve search method. The algorithm in each iteration finds the probe that maximizes a detection performance criterion (here the maximum pairwise correlation of bacteria) and adds it to the list of probes picked from the previous iterations. GPS stops when the maximum desired number of probes is reached.

(II) *Complete list of bacterial strains used in UMD simulations*

To evaluate the UMD platform for genus level bacterial detection, we selected 40 species from 40 different genera that are listed among most commonly pathogenic to humans by the Center for Disease Control and Prevention (CDC) (24). The genome sequences of the following strains were obtained from the NCBI website:

Acinetobacter baumannii ATCC 17978,
Aeromonas salmonicida subsp. salmonicida A449,
Bacteroides fragilis 638R,
Bacillus cereus ATCC 14579,
Bartonella henselae str. Houston-1,
Bifidobacterium dentium Bd1,
Bordetella pertussis Tohama,
Borrelia burgdorferi B31,
Brucella abortus S19,
Campylobacter jejuni subsp. jejuni 81116,
Clostridium botulinum B1 str. Okra,
Corynebacterium jeikeium K411,
Coxiella burnetii RSA 331,
Cupriavidus metallidurans CH34,
Enterobacter aerogenes EA1509E,
Enterococcus faecalis V583,
Escherichia coli str. K-12 substr. MG1655,
Francisella tularensis subsp. holarctica LVS,
Fusobacterium nucleatum subsp. nucleatum ATCC 25586,
Haemophilus influenzae F3047,
Helicobacter pylori B38,
Klebsiella pneumoniae 342,
Legionella pneumophila str. Corby,
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130,
Listeria monocytogenes 08-5578,

Micrococcus luteus NCTC 2665,
Mycobacterium leprae TN,
Mycoplasma pneumoniae M129,
Neisseria meningitidis MC58,
Prevotella melaninogenica ATCC 25845,
Propionibacterium acnes KPA171202,
Proteus mirabilis HI4320,
Pseudomonas aeruginosa LESB58,
Rickettsia rickettsii str. Iowa,
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC 9150,
Serratia proteamaculans 568,
Shigella sonnei Ss046,
Staphylococcus aureus subsp. aureus USA300 FPR3757,
Vibrio cholerae MJ-1236,
and Yersinia pestis CO92.

For species-level bacterial detection, the following 24 different strains from the Staphylococcus genus were selected to perform the simulations:
Staphylococcus arlettae CVD059 SARL-c1,
Staphylococcus aureus subsp. aureus NCTC 8325,
Staphylococcus auricularis strain DSM 20609,
Staphylococcus capitis subsp. capitis strain AYP1020,
Staphylococcus carnosus subsp. carnosus TM300,
Staphylococcus cohnii subsp. cohnii strain 532,
Staphylococcus epidermidis ATCC 12228,
Staphylococcus equorum strain KS1039,
Staphylococcus gallinarum strain DSM 20610,
Staphylococcus haemolyticus JCSC1435,
Staphylococcus hominis subsp. hominis C80,
Staphylococcus hyicus strain ATCC 11249,
Staphylococcus lentus F1142 s6-trimmed-contig-1,
Staphylococcus lugdunensis HKU09-01,
Staphylococcus massiliensis CCUG 55927,
Staphylococcus pettenkoferi VCU012,
Staphylococcus pseudintermedius HKU10-03,
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305,
Staphylococcus schleiferi strain 2317-03,
Staphylococcus sciuri subsp. sciuri strain DSM 20345,
Staphylococcus simiae CCM 7213 contig00565,
Staphylococcus simulans ACS-120-V-Sch1,
Staphylococcus vitulinus F1028 S-vitulinus-F1028-0001,
and Staphylococcus warneri SG1.

For species-level bacterial detection, the following 23 different strains from the Vibrio genus were selected to perform the simulations:
Vibrio anguillarum 775,
Vibrio brasiliensis LMG 20546 VIBR0546-99,
Vibriocholerae O1 biovar El Tor str. N16961,
Vibrio cyclitrophicus FF75 Ctg1,
Vibrio diazotrophicus NBRC 103148,
Vibrio ezurae NBRC 102218,
Vibrio furnissii NCTC 11218,
Vibrio genomosp. F6 str. FF-238,
Vibrio genomosp. F10 str. ZF-129,
Vibrio harveyi ATCC BAA-1116,
Vibrio litoralis DSM 17657,
Vibrio maritimus strain: JCM 19235,
Vibrio metschnikovii CIP 69.14 VIB.Contig153,
Vibrio nigripulchritudo str. SFn1,
Vibrio orientalis CIP 102891 ATCC 33934 strain CIP 102891,
Vibrio pacinii DSM 19139 BS19DRAFT-scaffold00001.1-C,
Vibrio parahaemolyticus RIMD 2210633,
Vibrio proteolyticus NBRC 13287,
Vibrio rhizosphaerae DSM 18581,
Vibrio scophthalmi LMG 19158 VIS19158-99,
Vibrio splendidus LGP32,
Vibrio tubiashii ATCC 19109,
and Vibrio vulnificus YJ016.

References and Notes

1. M. Klompas, D. S. Yokoe, Automated surveillance of health care-associated infections. *Clin. Infect. Dis.* **48**, 1268-1275 (2009).
2. V. N. Pinto, Bioterrorism: Health sector alertness. *J. Nat. Sc. Biol. Med.* **4**, 24-28 (2013).
3. B. V. Dorst, J. Mehta, K. Bekaert, E. Rouah-Martin, W. De Coen, P. Dubruel, R. Blust, J. Robbens, Recent advances in recognition elements of food and environmental biosensors: A review. *Biosensors Bioelectronics* **26**, 1178- 1194 (2010).
4. D. L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289-1306 (2006).
5. R. G. Baraniuk, Compressive sensing. *IEEE Signal Process. Mag.* **24**, 118-121 (2007).
6. J. SantaLucia, D. Hicks, The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415-440 (2004).
7. S. Chakravorty, B. Aladegbami, M. Burday, M. Levi, S. Marras, et al. Rapid universal identification of bacterial pathogens from clinical cultures by using a novel sloppy molecular beacon melting temperature signature technique. *J. Clin. Microbiol.* **48**, 258-267 (2010).
8. S. Tyagi, F. R. Kramer, Molecular beacons: probes that fluoresce upon hybridization. *Nature Biotechnol.* **14**, 303-308 (1996).
9. P. M. Dark, P. Dean, G. Warhurst, Bench-to-bedside review: the promise of rapid infection diagnosis during sepsis using polymerase chain reaction-based pathogen detection. *Crit. Care* **13**, 217 (2009).
10. Ž. Jeričević, Ž. Kušter, Non-linear optimization of parameters in Michaelis-Menten kinetics. *Croat. Chem. Acta* **78**, 519-523 (2005).
11. A. M. Bruckstein, M. Elad, M. Zibulevsky, On the uniqueness of non-negative sparse solutions to underdetermined systems of equations. *IEEE Trans. Inf. Theory* **54**, 4813- 4820 (2008).
12. Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report, Nov. (2013). http://www.cdc.gov/mmwr/
13. A. Pechorsky, Y. Nitzan, T. Lazarovitch, Identification of pathogenic bacteria in blood cultures: comparison between conventional and PCR methods. *J. Microbiol. Methods* **78**, 325-330 (2009).
14. S. Sontakke, M. B. Cadenas, R. G. Maggi, P. V. Diniz, E. B. Breitschwerdt, Use of broad range 16S rDNA PCR in clinical microbiology. *J. Microbiol. Methods* **76**, 217-225 (2009).
15. C. D. Sibley, G. Peirano, D. L. Church, Molecular methods for pathogen and microbial community detection and characterization: current and potential application in diagnostic microbiology. *Infect. Genet. Evol.* **12**, 505-521 (2012).
16. S. Riedel, K. C. Carroll, Blood cultures: key elements for best practices and future directions. *J. Infect. Chemother.* **16**, 301-316 (2010).
17. M. Paolucci, M. P. Landini, V. Sambri, Conventional and molecular techniques for the early diagnosis of bacteremia. *Int. J. Antimicrob. Ag.* **36**, S6-S16 (2010).
18. M. Bauer, K. Reinhart, Molecular diagnostics of sepsis - Where are we today? *Int. J. Med. Microbiol.* **300**, 411-413 (2010).
19. R. P. Peters, M. A. Van Agtmael, S. A. Danner, P. H. Savelkoul, C. M. Vandenbroucke- Grauls, New developments in the diagnosis of bloodstream infections. *Lancet Infect. Dis.* **4**, 751-760 (2004).
20. Centers for Disease Control and Prevention Antibiotic Resistance threats in the United States, Nov. (2013). http://www.cdc.gov/drugresistance/
21. M. A. Davenport, P. T. Boufounos, M. B. Wakin, R. G. Baraniuk, Signal processing with compressive measurements. *J. Sel. Top. Sign. Proces.* **4**, 445-460 (2010).
22. R. G. Baraniuk, More is less: Signal processing and the data deluge. *Science* **331**, 717-719 (2011).
23. C. L. Ventola, The Antibiotic Resistance Crisis: Part 1: Causes and Threats. *Pharmacy and Therapeutics* **40**, 277-283 (2015).
24. J. Sambrook, F. Fritsch, T. Maniatis, *Molecular cloning* (Cold Spring Harbor Lab. Press, New York, 2001), Appendix 8. [third edition]