Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

# Reagent-free bacterial identification using multivariate analysis of transmission spectra

Jennifer M. Smith
Debra E. Huffman
Dayanis Acosta
Yulia Serebrennikova
Luis García-Rubio
German F. Leparc

SPIE

# Reagent-free bacterial identification using multivariate analysis of transmission spectra

**Jennifer M. Smith,[a] Debra E. Huffman,[a] Dayanis Acosta,[a] Yulia Serebrennikova,[a] Luis García-Rubio,[a] and German F. Leparc[b]**
[a]Claro Scientific, LLC, 10100 Dr. Martin Luther King Jr. Street North, St. Petersburg, Florida 33710
[b]Florida Blood Services, A Division of OneBlood, Inc., 10100 Dr. Martin Luther King Jr. Street North, St. Petersburg, Florida 33710

**Abstract.** The identification of bacterial pathogens from culture is critical to the proper administration of antibiotics and patient treatment. Many of the tests currently used in the clinical microbiology laboratory for bacterial identification today can be highly sensitive and specific; however, they have the additional burdens of complexity, cost, and the need for specialized reagents. We present an innovative, reagent-free method for the identification of pathogens from culture. A clinical study has been initiated to evaluate the sensitivity and specificity of this approach. Multiwavelength transmission spectra were generated from a set of clinical isolates including *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus*. Spectra of an initial training set of these target organisms were used to create identification models representing the spectral variability of each species using multivariate statistical techniques. Next, the spectra of the blinded isolates of targeted species were identified using the model achieving >94% sensitivity and >98% specificity, with 100% accuracy for *P. aeruginosa* and *S. aureus*. The results from this on-going clinical study indicate this approach is a powerful and exciting technique for identification of pathogens. The menu of models is being expanded to include other bacterial genera and species of clinical significance. 
[DOI: 10.1117/1.JBO.17.10.107002]

Keywords: biomedical optics; transmission; spectrophotometry.

Paper 12368 received Jun. 12, 2012; revised manuscript received Aug. 3, 2012; accepted for publication Sep. 5, 2012; published online Oct. 1, 2012.

## 1 Introduction

The current health care crisis is forcing the modernization of the existing state of medical delivery. The rapid, cost-effective identification of infectious bacteria is an area of significant medical need for hospitals, clinics, and other laboratory settings, especially in remote locations or those serving a large indigent population in low-resource settings. Diagnostic systems in use today are often unable to identify life-threatening infections in time to effectively treat them in many patients. Bacterial infections leading to septicemia are especially deadly. Sepsis is the tenth-leading cause of death in the United States, and treating these infections adds $17 billion annually to the total health care expenditures for the country.[1] Early diagnosis and treatment of bacterial infections is key to better patient outcomes.[2] Providing the clinical laboratory with the appropriate diagnostic instrumentation to meet those needs is critical world-wide.

### 1.1 *Established and New Approaches*

Technologies that are used in the clinical laboratory to identify bacterial pathogens are primarily based upon either biochemical assays or molecular methodologies. Biochemical assay-based instruments, such as the Phoenix (BD) or Vitek 2 (bioMerieux), rely upon a series of chemical reactions conducted automatically on a card. Molecular methods primarily used include real-time polymerase chain reaction (PCR), peptide nucleic acid fluorescence in situ hybridization (PNA FISH) and can also be applied to a variety of sample types (i.e., blood, wounds, nasal swabs, etc.). Matrix-assisted laser desorption/ionization time of flight mass spectroscopy (MALDI-TOF MS), a recent advancement for the identification of blood culture pathogens, has gained approval in Europe for use in clinical applications.[3,4] Unlike some techniques, MALDI-TOF MS does not require any initial assessment, such as gram staining, choice of PCR primers, or usage of selective growth media; however, specific protocols involving chemical reagents are required to break cell walls to expose intracellular proteins for analysis. Despite these more recent approaches, sample processing and preparation remains laborious, complex, and expensive, and samples are susceptible to contamination, especially for DNA extraction. Further, these techniques are not necessarily within reach of smaller health-care centers with limited financial resources.

Optical biosensors are increasingly being considered as the next-generation tool for detection and identification of microbial pathogens in bodily fluids and tissue, as well as food and environmental samples.[5,6] The interaction of light with matter fundamentally changes the properties of the input light, and the measurement of such changes provides insights into the nature of a sample. Optical techniques such as fluorescence spectroscopy,[7] flow cytometry,[8] and immunological tests[9] among others[5,10,11] detect specific enzymes, antibodies, nucleic acids, labels, or pathogenic markers and require specialized reagents. There is also a renewed interest, in optical techniques that directly use the structural[5,12–17] or chemical properties of cell constituents[18–22] without the use of specialized reagents. From the structural perspective, the light scattering of cells has been used to identify bacteria based upon metrics like size and shape.[16,17,23] Intrinsic fluorophores such as tryptophan and

Address all correspondence to: Jennifer M. Smith, Claro Scientific, LLC, 10100 Dr. Martin Luther King Jr. Street North, St. Petersburg, Florida 33710. Tel: 727 543 2171; Fax: 727 543-1177; E-mail: jennifer.smith@clarosci.com

0091-3286/2012/$25.00 © 2012 SPIE

NADH have been used in fluorescence spectroscopy to discriminate among bacteria.[24–26] Vibrational spectroscopic techniques, specifically Raman spectroscopy and infrared spectroscopy, have been used to create "fingerprints" based upon the chemical composition of bacteria,[18,20,22,27] and systems such as the SpectraCell RA (River Diagnostics) are now commercially available. While the aforementioned techniques typically rely on the resolution of either structural or chemical information about the cells[5] to identify bacteria, it is also worthwhile to investigate the use of other types of optical measurements, like multiwavelength transmission spectroscopy, that include more than one type of light-matter interactions in a single measurement, thus increasing the diversity of variables available to discriminate among bacteria.

### 1.2 *Pathogen Identification from Transmission Measurements*

Both the absorption and scattering properties of a sample are measured in UV-Visible-NIR transmission spectra.[28] Such spectra of microorganisms and cells contain rich information pertaining to their physical structure and chemical composition.[29–35] The spectral features of cells and cellular organisms result from numerical density (concentration), size, shape, and internal composition.[29] A model that utilizes light scattering theory, spectral deconvolution techniques, and the approximation of the wavelength-dependent optical properties of the basic constituents of cells (e.g., DNA, etc.) has been proposed for the interpretation of the spectra of microorganisms.[31] It was shown that the features of the transmission spectra of bacteria (*Escherichia coli*, *Pseudomonas agglomerans*, *Bacillus globigii* vegetative cells and spores, and *Bacillus subtilis* spores) can be quantitatively linked to the internal and external characteristics of the bacteria thus enabling discrimination among them.[31] Although the ability of UV-visible-NIR spectroscopy for the detection, identification, and characterization of cells, pathogens, and disease markers using their spectral signatures has been demonstrated,[30,32,35] a statistically meaningful study of clinical isolates is still needed.

The objective of this study is a clinical demonstration of the utility of multiwavelength UV-vis-NIR transmission measurements for the identification of bacterial pathogens. The study utilizes the theoretically confirmed findings that multiwavelength transmission spectroscopy uniquely characterize bacteria and accounts for the natural variability in bacteria spectral signatures using multivariate statistical techniques. The approach is demonstrated to be an excellent discriminatory tool for clinically relevant pathogens such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Staphylococcus aureus* which accounted for more than 50% of the isolates received for this study.

## 2 Methods

### 2.1 *Targeted Organisms*

The bacteria *E. coli, K. pneumoniae*, *P. aeruginosa,* and *S. aureus* were selected as targets because of their high collective prevalence and severity of infection. A total of 555 bacterial isolates from positive blood cultures were received during the course of the study (Table 1) from the BayCare Health System, Florida. The identities of the isolates were blinded and arrived in sealed envelopes. The identity of the initial 188 isolates received were opened and used to construct and test identification models ($n = 116$ isolates of target organisms). Due to the substantial size differences in the number of isolates received for *K. pneumoniae* and *P. aeruginosa* in the initial set of 188 isolates (11 and 6, respectively *c.f.* 38 and 61 for *E. coli* and *S. aureus*, respectively), this set was further supplemented with an additional 45 clinical isolates of *K. pneumoniae* and *P. aeruginosa* (23 *K. pneumoniae* and 22 for *P. aeruginosa*) from a second hospital (Tampa General Hospital, Florida) to ensure adequate sample sizes for each target organism. The remaining 367 isolates were analyzed as blinded samples ($n = 204$ isolates for target organisms).

### 2.2 *Generation of Bacterial Spectra*

Participating hospitals subcultured positive blood cultures onto either blood agar or MacConkey agar plates, which were sent via courier for analysis. Upon arrival all plates were inspected for sample integrity and colonial morphology. Samples were processed as follows: a single well-isolated colony was selected and inoculated using sterile technique into a 250 mL flask containing 50 mL of sterile tryptic soy broth. The flask was incubated at 37°C with agitation for $17 \pm 2$ *h*. After incubation, a 1 mL aliquot was collected from each flask and transferred to a 1.5 mL sterile Eppendorf tube. Samples were immediately placed into a ThermoFisher Accuspin Microfuge 12 and spun for 3 min at 12,000 RPMs. Upon completion of the centrifugation cycle, the tubes were removed from the Microfuge and the supernatant was slowly drawn off and discarded. The remaining pellets were re-suspended using de-ionized water and vortexed for a few seconds. This washing process was repeated three times to remove all traces of the growth media. After the last wash, the pellet of clean cells was re-suspended well in de-ionized water. Aliquots of the final cell suspension (30 to 60 $\mu$L) were diluted into 3 mL of deionized water in a 1 cm path length quartz cuvette. For the transmission measurements, the following diode array spectrometers were tested and found to be equivalent for pathogen identification: Agilent 8453, Santa Clara, California; Ocean Optics HR4000, Dunedin, Florida; and Ocean Optics HR2000, Dunedin, Florida. The data reported herein are from the Agilent 8453. Other commercial spectrophotometers can be used provided that the spectrophotometers have a minimum of 1 nm wavelength resolution, a high signal-to noise ratio (>99%) and an acceptance angle smaller than 2 deg. The degree of dilution of each sample was selected to yield maximal optical density values between 0.4 and 0.8 absorbance units in the wavelength range of 190 to 220 nm. In the case of spectra that exceed maximum allowable optical density, samples were diluted to achieve the requisite optical density range. Data were collected with an integration time of 0.5 s. Collected spectra were inspected and removed from the data set if they exhibited significant scattering signal in the visible-IR wavelength region due to aggregation, diffraction artifacts, or exceeded allowable optical densities. All spectra were pretreated by normalizing by the area under the curve between 210 to 900 nm.

### 2.3 *Model Development and Validation*

The generation of an accurate identification model for an organism must be based on a set of spectra that represents as many of the spectral variations as possible that can be exhibited by an organism. Natural spectral variability can be related to bacterial growth due to changes in the number of cells, their shape,

**Table 1** Species prevalence of bacteria from positive blood cultures from BayCare Hospital System.

| Genus | Species | Known | % | Blind | % |
|---|---|---|---|---|---|
| *Achromobacter* | *denitrificans* | 0 | — | 2 | 0.5 |
| *Achromobacter* | *xylosoxidans* | 1 | 0.01 | 3 | 0.8 |
| *Acinetobacter* | *lwoffi* | 0 | — | 1 | 0.3 |
| *Aeromonas* | *hydrophilia/caviae* | 0 | — | 1 | 0.3 |
| *Chryseobacterium* | *meningosepticum* | 1 | 0.01 | 0 | — |
| *Citrobacter* | *braakii* | 0 | — | 1 | 0.3 |
| *Citrobacter* | *freundii* | 0 | — | 3 | 0.8 |
| *Enterobacter* | *aerogenes* | 1 | 0.01 | 1 | 0.3 |
| *Enterobacter* | *cloacae* | 1 | 0.01 | 14 | 3.8 |
| *Enterococcus* | *avium* | 1 | 0.01 | 0 | — |
| *Enterococcus* | *durans* | 1 | 0.01 | 1 | 0.3 |
| *Enterococcus* | *faecalis* | 8 | 0.04 | 18 | 4.9 |
| *Enterococcus* | *faecium* | 2 | 0.01 | 3 | 0.8 |
| *Escherichia* | *coli* | 38 | 0.2 | 71 | 19.3 |
| *Klebsiella* | *oxytoca* | 0 | — | 1 | 0.3 |
| *Klebsiella* | *pneumoniae* | 11 | 0.06 | 16 | 4.4 |
| *Pasteurella* | *multocida* | 0 | — | 2 | 0.5 |
| *Pasteurella* | *sp.* | 0 | — | 1 | 0.3 |
| *Proteus* | *mirabilis* | 0 | — | 7 | 1.9 |
| *Pseudomonas* | *aeruginosa* | 6 | 0.03 | 16 | 4.4 |
| *Pseudomonas* | *oryzihabitans* | 0 | — | 1 | 0.3 |
| *Salmonella* | *sp.* | 2 | 0.01 | 1 | 0.3 |
| *Serratia* | *marcescens* | 3 | 0.02 | 4 | 1.1 |
| *Staphylococcus* | *aureus* | 61 | 0.3 | 101 | 27.5 |
| *Staphylococcus* | *auricularis* | 2 | 0.01 | 3 | 0.8 |
| *Staphylococcus* | *capitis* | 2 | 0.01 | 8 | 2.2 |
| *Staphylococcus* | *cohnii* | 0 | — | 1 | 0.3 |
| *Staphylococcus* | *epidermidis* | 28 | 0.2 | 46 | 12.5 |
| *Staphylococcus* | *haemolyticus* | 1 | 0.01 | 4 | 1.1 |
| *Staphylococcus* | *hominis* | 9 | 0.05 | 17 | 4.6 |
| *Staphylococcus* | *lugdunensis* | 1 | 0.01 | 0 | — |
| *Staphylococcus* | *saprophyticus* | 2 | 0.01 | 0 | — |
| *Staphylococcus* | *simulans* | 0 | — | 3 | 0.8 |
| *Staphylococcus* | *warneri* | 4 | 0.02 | 5 | 1.4 |

**Table 1** (*Continued*).

| Genus | Species | Known | % | Blind | % |
|---|---|---|---|---|---|
| *Staphylococcus* | *xylosus* | 0 | — | 1 | 0.3 |
| *Stenotrophomonas* | *maltophilia* | 1 | 0.01 | 0 | — |
| *Streptococcus* | *acidominimus* | 0 | — | 1 | 0.3 |
| *Streptococcus* | *agalactiae* | 1 | 0.01 | 6 | 1.6 |
| *Streptococcus* | *anginosus* | 0 | — | 1 | 0.3 |
| *Streptococcus* | *infantarius* | 0 | — | 1 | 0.3 |
| *Streptococcus* | *salivarus* | 0 | — | 1 | 0.3 |
| **Total** | | **188** | 100%* | **367** | 100%* |

*Totals may not equal 100% due to rounding.

chemical composition, and internal structure.[29] For example, a spectral data set representing the complete growth cycle of an organism must include spectra of the organism at the lag, exponential, stationary, and death phases since they have distinct spectral signatures.[29] Moreover, a set of spectra describing the stationary phase alone should contain any spectral variation that can occur within this phase as well as any contribution due to sample processing. In order to account for such spectral variations, the set of spectra used in the training set was composed of multiple spectra of each isolate collected while in stationary phase.

Following this, 10 spectra (Fig. 1) were selected from the full set of spectra available for each species from their respective training sets to create an optimal identification model (i.e., a set of eigenvectors) solving the generalized eigenvalue problem [Eq. (1)].

$$Ax = \lambda x \tag{1}$$

$$A = \frac{XX'}{n}, \tag{2}$$

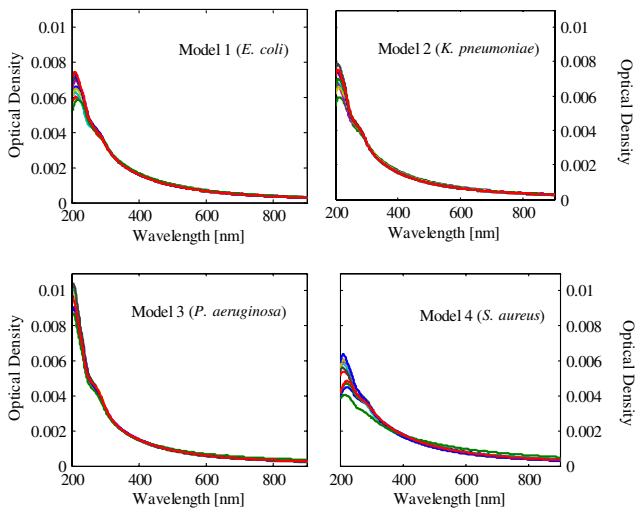


**Fig. 1** Transmission spectra used to construct each identification model. Spectra have been normalized by the area under the curve.

where $\lambda$ are the eigenvalues, $x$ are the corresponding eigenvectors, $A$ is the covariance matrix of the spectral data, $X$, and $n$ is the number of spectra.[36] The eigenvectors represent the underlying spectral features of the corresponding microorganisms and are defined as the "model" for each organism. Therefore, the spectral features contained within the set of eigenvectors derived from the training set are present in the transmission spectra of any sample of a given microorganism. The spectra used to generate each model were chosen to span the range of spectral variability observed for each species. The largest breadth of this variability is apparent in the UV region of the measured spectra (Fig. 1). While the selection of 10 spectra to represent each organism is somewhat arbitrary, given the small sample sizes of *K. pneumoniae* and *P. aeruginosa*, it is a result of trying strike a balance between including enough spectra to capture the variability and having enough spectra left over to validate the model.

Under the aforementioned assumptions, a regression step is required to calculate the sample residuals and compare among models for classification. The entire set of generated eigenvectors ($n = 10$ for each model) were regressed in an ordinary least squares sense[37] against the remaining spectra from the training set and the entire set of blinded spectra. This is described generally in Eq. (3).

$$B = mx, \tag{3}$$

where $B$ is a the spectral data, $m$ is a set of weights describing the relationship between $B$ and the eigenvectors generated from the training set, $x$. Specifically, for $i$ spectra and $j$ models,

$$m_i = \frac{B_i}{\lambda_j} \tag{4}$$

$$\hat{B}_i = m_i \lambda_j. \tag{5}$$

Typically only the eigenvectors with the highest eigenvalues are used in this type of regression; however, we found that using the entire set yielded the best sensitivity and specificity in the training set.[38] A residual sum of squares (RSSQ) was generated for the regression of each spectrum against each model [Eq. (6)]

**Table 2** Results of the training set for the sensitivity and specificity for each of the target organisms. These results confirm the validity of the model for each organism.

| Species | # Spectra (Isolates) | Sensitivity | Specificity | Accuracy* |
|---|---|---|---|---|
| *E. coli* | 164 (38) | 98.2 ± 0.2% | 98.9 ± 0.1% | 98.6 ± 0.1% |
| *K. pneumoniae* | 46 (28) | 93.5 ± 1.1% | 99.2 ± 0.4% | 98.6 ± 0.5% |
| *P. aeruginosa* | 30 (22) | 100% | 100% | 100.0% |
| *S. aureus* | 202 (61) | 100% | 100% | 100.0% |

*Accuracy = (TP + TN)/(TP + TN + FP + FN); Confidence intervals from Harper 1999.[39]

$$\mathrm{RSSQ}_{ij} = \sum \frac{(\hat{B}_{ij} - B_i)^2}{nwav}, \tag{6}$$

where *nwav* is equal to the number of wavelengths used for each spectrum. The model regression that generated the smallest RSSQ value was judged to best represent the spectrum, and the spectrum was correspondingly assigned the identification of the species represented by that model. Replicate spectra of many of the blinded isolates were generated from independently grown cultures and analyzed with the identification models to establish the reproducibility of the results. Identification results were compared to those obtained by the hospital using bioMerieux Vitek-2.

The spectra from the training set not selected for model creation were used to validate the model. Table 2 summarizes the identification results obtained for the spectra for each training set, excluding those used to create the models. Sensitivity and specificity values of the training sets were greater than 93.5% for all four target organisms.

## 3 Results and Discussion

### 3.1 *Targeted Species*

Collected spectra of 204 blind isolates were analyzed with four identification models, and the results are presented in Table 3. Strong sensitivity and specificity values were achieved for the target species. The overall accuracy for the identification of the targeted species was 98%. Replicate spectra generated from independently grown cultures and representing 169 isolates were also tested to confirm the robustness of the approach.

The RSSQ values obtained from the regression of the eigenvector sets against the blinded spectra of target organisms were the lowest and were within the same order of magnitude as those for the training sets (Table 4). However, the RSSQ values for training sets were still smaller than those obtained from the blinded spectra of the target species. Given the limited size of the training sets, as is evidenced by the larger spread of the residuals in the blinded data, this would indicate that the identification models underestimate the full scope of spectral variability for each species and could be enhanced with additional data in the training sets (i.e., a larger number of spectra from additional isolates from which to generate the eigenvectors). *P. aeruginosa* had the largest difference in RSSQ values between the training set and blinded isolates (*c.f.* $4.84e^{-12}$ versus $8.22e^{-12}$). This difference speaks to the small sample size used for the construction of its model ($n = 30$ spectra). Notwithstanding this difference, the spectral features of *P. aeruginosa* are exceptionally distinct from the other targeted organisms so much that there was no impact on the identification results (100% accurate). Another example of how spectral differences among species can be effectively exploited for correct identification was with *S. aureus*, which showed 100% sensitivity and specificity in both training and blind data sets.

In both the training set and the blind samples, there was one cross-identification between *E. coli* and *K. pneumoniae* and one *E. coli* isolate was incorrectly identified as *K. pneumoniae*.

**Table 3** Sensitivity and specificity results for the blind samples. Analysis of independently grown replicate cultures confirms the same overall accuracy rate.

| | Reported data | | | | Replicate data | | | |
|---|---|---|---|---|---|---|---|---|
| | # Isolates | Sensitivity | Specificity | Accuracy | # Spectra (Isolates) | Sensitivity | Specificity | Accuracy |
| *E. coli* | 71 | 95.8 ± 0.6% | 99.2 ± 0.2% | 98.0 ± 0.4% | 38 (25) | 97.4 ± 0.4% | 98.5 ± 0.3% | 98.2 ± 0.4% |
| *K. pneumoniae* | 16 | 93.8 ± 3% | 98.4 ± 1.5% | 98.0 ± 1.7% | 10 (6) | 90.0% | 98.7% | 98.2% |
| *P. aeruginosa* | 16 | 100% | 100% | 100.0% | 5 (4) | 80.0% | 100.0% | 99.4% |
| *S. aureus* | 101 | 100% | 100% | 100.0% | 116 (77) | 99.1 ± 0.2% | 100.0% | 99.4 ± 0.1% |
| Total | 204 | | | | 169 (112) | | | |

*Confidence intervals are calculated from Harper and Reeves, 1999 on samples with appropriate sample sizes.[39]

**Table 4** A comparison of residuals among data sets.

| Organism/model | *N* | Mean RSSQ | St. dev of RSSQ |
|---|---|---|---|
| **Training set** | Spectra/isolates | | |
| *E. coli* | 164/38 | 2.56*e* – 12 | 2.20*e* – 12 |
| *K. pneumoniae* | 46/28 | 1.93*e* – 12 | 1.58*e* – 12 |
| *P. aeruginosa* | 30/22 | 4.84*e* – 12 | 4.58*e* – 12 |
| *S. aureus* | 202/61 | 3.70*e* – 12 | 2.82*e* – 12 |
| **Blinded targeted species** | *N* (isolates) | | |
| *E. coli* | 68 | 3.17*e* – 12 | 2.42*e* – 12 |
| *K. pneumoniae* | 15 | 2.62*e* – 12 | 1.42*e* – 12 |
| *P. aeruginosa* | 16 | 8.22*e* – 12 | 4.16*e* – 12 |
| *S. aureus* | 101 | 4.77*e* – 12 | 2.80*e* – 12 |
| **Blinded nontargeted species(based on model assignment)** | *N* (isolates) | | |
| Model 1 | 31 | 2.03*e* – 11 | 1.89*e* – 11 |
| Model 2 | 21 | 1.23*e* – 11 | 5.52*e* – 12 |
| Model 3 | 13 | 3.82*e* – 11 | 5.29*e* – 11 |
| Model 4 | 102 | 1.46*e* – 11 | 2.25*e* – 11 |

This crossover is due to the fact that the spectra of *E. coli* and *K. pneumoniae* are very similar (Fig. 1). On the other hand the spectra of *P. aeruginosa*, another Gram-negative rod, are clearly distinct. Unlike *E. coli* and *K. pneumoniae*, *P. aeruginosa* is known to produce phenazine compounds, namely pyocyanin, which are thought to confer a competitive growth advantage and are a virulence factor in certain diseases, such as cystic fibrosis.[40] These fluorescent compounds are strong absorbers in the UV-visible portion of the spectrum[41,42] and likely contribute to the distinct spectral features of *P. aeruginosa*.

### 3.2 *Nontargeted Species*

The blind spectra of nontargeted species show distinct differences in spectral features from those of the targeted organisms (not shown). The RSSQ values of the regression of the target organism models against the spectra of the blind isolates of nontarget species are given in Table 4. These RSSQ values were much greater than those for the blind isolates of the target species indicating that the models performed substantially better for their respective organisms. A contrast of more than one order of magnitude in the RSSQ values between the target and nontarget organisms suggests that a RSSQ threshold can be imposed to discriminate the target organisms from nontarget organisms and thus, the utility of using the models for target organisms in the absence of those for others remains relevant. A retrospective analysis that imposed an RSSQ threshold adequate to ensure the same reported sensitivity for each model (Table 3) resulted in specificity values of 89%, 98%, and 98% for *E. coli*, *K. pneumonia,* and *P. aeruginosa*, respectively. The specificity value for *S. aureus* was substantially lower, which is not particularly surprising given that this was the only model representing cocci and captured the widest number of species (see Table 5).

Because all spectra were collected on a blinded basis, the nontarget organisms also received an identification based on the best model fit. We have evaluated the merit of the models when applied to organisms they were not designed to represent. This dataset included 32 species representing 14 genera. A summary of the results is presented in Table 5.

The models did well at selecting organisms that express the same morphological characteristics as their targets. For example, 97% of the *Staphylococcus* isolates were assigned to Model 4 (183/189 including *S. aureus*; 82/88, excluding *S. aureus*). Model 4 also captures half of the *Enterococcus* isolates (11/22) and the majority of the *Streptococcus* (8/10) isolates. *Enterobacter* is described almost equally well by Models 1 and 2—two of the models that describe Gram-negative organisms. Interestingly, one isolate of *Pseudomonas oryzihabitans* was not classified under Model 3. Closer inspection of the spectrum of this isolate revealed substantial differences from the spectra of *P. aeruginosa*. In fact, this may not be surprising considering that *P. oryzihabitans* (previously known as *Flavimonas oryzihabitans*) is in a different rRNA homology group than *P. aeruginosa* and has a different phenotypic expression.[43,44] Given that these models were designed for specific species, it encourages confidence that the inclusion of additional models can provide a higher degree of taxonomic resolution.

The vast majority of the blinded samples were assigned to either Model 1 or 4 (82%). This can be explained in two different manners. The spectral variance of *E. coli* and *S. aureus* isolates is so large that they act like "catch all" models. This large variance can be explained by either the natural variability among

**Table 5** Classification of blinded isolates using four models. Model assignment is indicated by the four right-most columns. Models 1 to 4 represent *E. coli*, *K. pneumoniae*, *P. aeruginosa*, and *S. aureus*, respectively.

| Genus | Species | *n* | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|---|---|
| *Achromobacter* | *dentrificans* | 2 | | 1 | 1 | |
| *Achromobacter* | *xylosoxidans* | 3 | | | 3 | |
| *Acinetobacter* | *lwoffi* | 1 | | 1 | | |
| *Aeromonas* | *hydrophila/caviae* | 1 | | 1 | | |
| *Citrobacter* | *braakii* | 1 | 1 | | | |
| *Citrobacter* | *freundii* | 3 | 3 | | | |
| *Enterobacter* | *aerogenes* | 1 | | | 1 | |
| *Enterobacter* | *cloacae* | 14 | 8 | 6 | | |
| *Enterococcus* | *durans* | 1 | | | | 1 |
| *Enterococcus* | *faecalis* | 18 | 3 | 7 | | 8 |
| *Enterococcus* | *faecium* | 3 | 1 | | | 2 |
| *Escherichia* | *coli* | 71 | 68 | 3 | | |
| *Klebiella* | *oxytoca* | 1 | | 1 | | |
| *Klebiella* | *pneumoniae* | 16 | 1 | 15 | | |
| *Pasteurella* | *multocida* | 2 | 1 | | 1 | |
| *Pasteurella* | *spp.* | 1 | | | 1 | |
| *Proteus* | *mirabilis* | 7 | 4 | | 3 | |
| *Pseudomonas* | *aeruginosa* | 16 | | | 16 | |
| *Pseudomonas* | *orzyhabitans* | 1 | | | | 1 |
| *Salmonella* | *spp.* | 1 | 1 | | | |
| *Serratia* | *marcescens* | 4 | 1 | | 3 | |
| *Staphylococcus* | *aureus* | 101 | | | | 101 |
| *Staphylococcus* | *auricularis* | 3 | | | | 3 |
| *Staphylococcus* | *capitis* | 8 | | | | 8 |
| *Staphylococcus* | *cohnii* | 1 | | | | 1 |
| *Staphylococcus* | *epidermidis* | 46 | 1 | | | 45 |
| *Staphylococcus* | *haemolyticus* | 4 | | | | 4 |
| *Staphylococcus* | *hominis* | 17 | 3 | | | 14 |
| *Staphylococcus* | *simulans* | 3 | 1 | | | 2 |
| *Staphylococcus* | *warneri* | 5 | 1 | | | 4 |
| *Staphylococcus* | *xylosus* | 1 | | | | 1 |
| *Streptococcus* | *acidominimus* | 1 | | | | 1 |
| *Streptococcus* | *agalactiae* | 6 | 1 | 1 | | 4 |
| *Streptococcus* | *anginosus* | 1 | | | | 1 |

**Table 5** (*Continued*).

| Genus | Species | *n* | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|---|---|
| *Streptococcus* | *infantarius* | 1 | | | | 1 |
| *Streptococcus* | *salivarius* | 1 | | | | 1 |
| **Totals** | | **367** | **99** | **36** | **29** | **203** |

isolates or can be attributed to their large sample sizes. Alternatively, it could be postulated that the models representing *K. pneumoniae* and *P. aeruginosa* (Models 2 and 3) are formulated from spectra that are so distinct in character from the other organisms that they have elevated skill in excluding them. Given that the size of the sample set from which the models were built was relatively small ($n = 46$ and 30, respectively) these models may only represent a fraction of the variance expressed by these organisms; however, this is not supported by the results from the blinded samples of *K. pneumoniae* and *P. aeruginosa*. Again, the numbers of blind samples were small relative to the other target organisms, so it is difficult to extrapolate beyond the results at hand. In the absence of equal representation of other species, it is difficult to determine the number of spectra required to describe the full variance of spectral features from stationary phase cultures for each species (some may require more than others). Because of this, numerical comparisons of sampling sizes among different species are challenging. Models will be revisited to further discriminate among cocci and include other organisms with discriminating spectral features.

This study demonstrates that the transmission spectra of clinical pathogens can be utilized for pathogen identification. This method eliminates the need for specialized reagents and can be directly applied to liquid cultures (e.g., in a research laboratory or food safety testing). In a clinical setting, colonies could be sampled directly from a culture plate and suspended in water for analysis. Maximum utility of this approach will be achieved when complemented with an appropriate sample preparation system (see Refs. 45 to 47 for promising developments) and thus this method could be applied directly to the positive blood cultures, eliminating the need for plating or culturing.

## 4 Conclusions

This pilot study demonstrates that the use of multiwavelength transmission measurements is an effective and economical way to identify pathogens without specialized reagents. Sensitivities and specificities >93.5% to 100% were achieved for *E. coli*, *K. pneumoniae*, *P. aeruginosa*, and *S. aureus*. The sampling protocol, hardware capabilities, and models used in this study are being expanded. Subsequent generations of sample handling protocols will include refinement and automation of sample preparation (samples were processed by hand in this study) that will significantly decrease the testing turn-around time. The menu of targeted organisms will be extended to include other prevalent organisms with special focus upon *Staphylococcus* species. To this end, it is expected that there is an upper limit on the utility of transmission measurements alone for bacterial identification. To overcome such a limitation, additional independent, yet complementary optical measurements (i.e., angular scattering) will be necessary to discriminate among some species. Statistical models will be further informed by incorporating such additional discriminatory information and will also be expanded to include theoretical and other multivariate approaches. In summary, our studies have demonstrated that standard UV-vis-NIR transmission measurements can be utilized for reagent-free pathogen identification.

### Acknowledgments

Financial support for this study was provided by Claro Scientific LLC St. Petersburg, Florida. Bacterial isolates were provided by BayCare Health Systems and Tampa General Hospital. The authors would like to acknowledge the insightful comments from the reviewers and facilities support from Florida Blood Services, a division of OneBlood Inc.

### References

1. D. C. Angus et al., "Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care," *Crit. Care Med.* **29**(7), 1303–1310 (2001).
2. P. G. Brindley, N. Zhu, and W. Sligl, "Best evidence in critical care medicine—early antibiotics and survival from septic shock: it's about time," *Can. J. Anaesth.* **53**(11), 1157–1160 (2006).
3. J. O. Lay, "MALDI-TOF mass spectrometry of bacteria," *Mass Spectrom. Rev.* **20**(4), 172–194 (2001).
4. M. Drancourt, "Detection of microorganisms in blood specimens using matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a review," *Clin. Microbiol. Infect.* **16**(11), 1620–1625 (2010).
5. A. K. Bhunia et al., "Light scattering, fiber optic-and cell-based sensors for sensitive detection of foodborne pathogens," *J. Rapid Methods Automat. Micro.* **15**(2), 121–145 (2007).
6. M. Nayak et al., "Detection of microorganisms using biosensors—A smarter way towards detection techniques," *Biosens. Bioelectron.* **25**(4), 661–667 (2009).
7. K. Oliveira et al., "Direct identification of Staphylococcus aureus from positive blood culture bottles," *J. Clin. Microbiol.* **41**(2), 889–891 (2003).
8. H. M. Davey and D. B. Kell, "Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses," *Microbiol. Rev.* **60**(4), 641–696 (1996).
9. P. B. Luppa, L. J. Sokoll, and D. W. Chan, "Immunosensors—principles and applications to clinical chemistry," *Clin. Chim. Acta* **314**(1–2), 1–26 (2001).
10. D. Ivnitski et al., "Biosensors for detection of pathogenic bacteria," *Biosens. Bioelectron.* **14**(7), 599–624 (1999).
11. O. Lazcka, F. J. Del Campo, and F. X. Munoz, "Pathogen detection: a perspective of traditional methods and biosensors," *Biosens. Bioelectron.* **22**(7), 1205–1217 (2007).
12. A. Dunn and R. Richards-Kortum, "Three-dimensional computation of light scattering from cells," *IEEE J. Sel. Top. Quantum Electron.* **2**(4), 898–905 (1996).
13. X.-T. Su et al., "2D light scattering patterns of mitochondria in single cells," *Opt. Express* **15**(17), 10562–10575 (2007).
14. I. Buzalewicz, A. Wieliczko, and H. Podbielska, "Influence of various growth conditions on Fresnel diffraction patterns of bacteria colonies examined in the optical system with converging spherical wave illumination," *Opt. Express* **19**(22), 21768–21785 (2011).
15. M. Venkatapathi et al., "High speed classification of individual bacterial cells using a model-based light scatter system and multivariate statistics," *Appl. Opt.* **47**(5), 678–686 (2008).
16. P. J. Wyatt and D. T. Phillips, "Structure of single bacteria from light scattering," *J. Theor. Biol.* **37**(3), 493–501 (1972).

17. P. J. Wyatt, "Differential light scattering: a physical method for identifying living bacterial cells," *Appl. Opt.* **7**(10), 1879–1896 (1968).
18. K. Maquelin et al., "Identification of medically relevant microorganisms by vibrational spectroscopy," *J. Microbiol. Methods* **51**(3), 255–271 (2002).
19. U. Neugebauer et al., "Towards a detailed understanding of bacterial metabolism—spectroscopic characterization of Staphylococcus epidermidis," *Chem. Phys. Chem.* **8**(1), 124–137 (2007).
20. K. Maquelin et al., "Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures," *J. Clin. Microbiol.* **41**(1), 324–329 (2003).
21. K. Gaus et al., "Classification of lactic acid bacteria with UV-resonance Raman spectroscopy," *Biopolymers* **82**(4), 286–290 (2006).
22. A. Harz, P. Rosch, and J. Popp, "Vibrational spectroscopy—a powerful tool for the rapid identification of microbial cells at the single-cell level," *Cytometry Part A* **75A**(2), 104–113 (2009).
23. A. Katz et al., "Bacteria size determination by elastic light scattering," *IEEE J. Sel. Top. Quantum Electron.* **9**(2), 277–287 (2003).
24. H. E. Giana et al., "Rapid identification of bacterial species by fluorescence spectroscopy and classification through principal components analysis," *J. Fluorescence* **13**(6), 489–493 (2003).
25. L. Leblanc and E. Dufour, "Monitoring the identity of bacteria using their intrinsic fluorescence," *FEMS Microbiol. Lett.* **211**(2), 147–153 (2002).
26. M. S. Ammor, "Recent advances in the use of intrinsic fluorescence for bacterial identification and characterization," *J. Fluorescence* **17**(5), 455–459 (2007).
27. M. Harz et al., "Micro-Raman spectroscopic identification of bacterial cells of the genus Staphylococcus and dependence on their cultivation conditions," *Analyst* **130**(11), 1543–1550 (2005).
28. M. Kerker, *The Scattering of Light, and Other Electromagnetic Radiation*, Academic Press, New York (1969).
29. C. E. Alupoaei and L. H. Garcia-Rubio, "Growth behavior of microorganisms using UV-Vis spectroscopy: Escherichia coli," *Biotechnol. Bioeng.* **86**(2), 163–167 (2004).
30. C. E. Alupoaei and L. H. Garcia-Rubio, "An interpretation model for the UV-VIS spectra of microorganisms," *Chem. Eng. Comm.* **192**(2), 198–218 (2005).
31. C. E. Alupoaei, J. A. Olivares, and L. H. Garcia-Rubio, "Quantitative spectroscopy analysis of prokaryotic cells: vegetative cells and spores," *Biosens. Bioelectron.* **19**(8), 893–903 (2004).
32. Y. Mattley et al., "Light scattering and absorption model for the quantitative interpretation of human blood platelet spectral data," *Photochem. Photobiol.* **71**(5), 610–619 (2000).
33. S. Narayanan et al., "Ultraviolet and visible light spectrophotometric approach to blood typing: objective analysis by agglutination index," *Transfusion* **39**(10), 1051–1059 (1999).
34. M. R. Callahan, J. B. Rose, and L. Garcia-Rubio, "Use of multiwavelength transmission spectroscopy for the characterization of Cryptosporidium parvum oocysts: quantitative interpretation," *Environ. Sci. Technol.* **37**(22), 5254–5261 (2003).
35. Y. M. Serebrennikova, Y. M. Patel, and L. H. García-Rubio, "Interpretation of the UV-visible spectra of malaria parasite Plasmodium falciparum," *Appl. Opt.* **49**(2), 180–188 (2010).
36. E. Kreyszig, *Advanced Engineering Mathematics*, 8th ed., Wiley, New York (1999).
37. E. R. Malinowski, *Factor Analysis in Chemistry*, John Wiley & Sons, Ltd., New York (2002).
38. I. T. Jolliffe, "A note on the use of principal components in regression," *J. Royal Statistical Soc. Series C (Applied Statistics)* **31**(3), 300–303 (1982).
39. R. Harper and B. Reeves, "Reporting of precision of estimates for diagnostic accuracy: a review," *Br. Med. J.* **318**(7194), 1322–1323 (1999).
40. J. M. Turner and A. J. Messenger, "Occurrence, biochemistry and physiology of phenazine pigment production," *Adv. Microb. Physiol.* **27**, 211–275 (1986).
41. C. Propst and L. Lubin, "Light-mediated changes in pigmentation of Pseudomonas aeruginosa cultures," *J. Gen. Microbiol.* **113**(2), 261–266 (1979).
42. D. V. Mavrodi et al., "Functional analysis of genes for biosynthesis of pyocyanin and phenazine-1-carboxamide from Pseudomonas aeruginosa PAO1," *J. Bacteriol.* **183**(21), 6454–6465 (2001).
43. N. J. Palleroni et al., "Nucleic acid homologies in the genus Pseudomonas," *Int. J. Syst. Bacteriol.* **23**(4), 333–339 (1973).
44. Y. Anzai, Y. Kudo, and H. Oyaizu, "The phylogeny of the genera Chryseomonas, Flavimonas, and Pseudomonas supports synonymy of these three genera," *Int. J. Syst. Bacteriol.* **47**(2), 249–251 (1997).
45. F. Petersson et al., "Free flow acoustophoresis: Microfluidic-based mode of particle and cell separation," *Anal. Chem.* **79**(14), 5117–5123 (2007).
46. H. W. Hou et al., "Microfluidic devices for blood fractionation," *Micromachines* **2**(3), 319–343 (2011).
47. A. A. S. Bhagat et al., "Microfluidics for cell separation," *Med. Biol. Eng. Comput.* **48**(10), 999–1014 (2010).