Peer Reviewed Paper openaccess Special Issue on Chemometrics in Hyperspectral Imaging

Unsupervised classification of individual foodborne bacteria from a mixture of bacteria cultures within a hyperspectral microscope image

Matthew Eady[a] and Bosoon Park[b,*]

[a]The US National Poultry Research Center, US Department of Agriculture – Agricultural Research Services, Athens, GA 30606, USA. ORCID: https://orcid.org/0000-0002-3617-6636

[b]The US National Poultry Research Center, US Department of Agriculture – Agricultural Research Services, Athens, GA 30606, USA. E-mail: Bosoon.park@ars.usda.gov. ORCID: https://orcid.org/0000-0001-8721-9117

Salmonella is a leading cause of foodborne illness. Traditional detection methods require lengthy incubation periods or expensive reagent kits. Hyperspectral microscope images (HMIs) have been previously investigated as a method for early and rapid detection of bacteria by using a spectral signature that is unique to the organism. Previous HMI use with bacteria has consisted of supervised classification with hypercubes collected for single culture images isolated from highly selective growth media. In order to move forward with HMI as a detection tool in the food industry, unsupervised classification of bacteria cells in mixed culture HMIs was investigated. Four foodborne bacteria cultures, *S.* Typhimurium (ST) *E. coli* (Ec), *S. aureus* (Sa) and *L. innocua* (Li) were combined in seven different culture combinations with HMIs collected between 450 nm and 800 nm. A *k*-means divisive cluster analysis (CA) was implemented and mixed culture image sets were found to contain between two and four clusters. CA cluster accuracy was obtained by assigning a dummy variable of the proposed CA classification, then carrying out a discriminant analysis. From the mixed culture HMIs, 700 bacteria cells were classified and accuracies were between 91.92% and 100%, with six of the seven HMI sets resulting in >97% accuracies. A distance measure between clusters was applied to identify unknown clusters based on single culture reference samples of the four bacteria used. Results showed that the CA has potential for unsupervised classification of bacteria cells, but the distance metric was not an adequate method for identifying the unknown cluster based on reference spectra, potentially due to the collinearity amongst bacteria spectra.

Keywords: foodborne bacteria, hyperspectral microscope images, cluster analysis, discriminant analysis, distance classifier

Introduction

Worldwide, it is estimated that 600 million people become ill due to foodborne illness and 230,000 deaths occur annually.[1] *Salmonella* is a pathogenic bacterium that has been associated with foodborne

Disclaimer: Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture.

Correspondence
Bosoon Park (Bosoon.park@ars.usda.gov)

Received: 12 October 2017
Revised: 20 December 2017
Accepted: 31 January 2018
Publication: 6 March 2018
doi: 10.1255/jsi.2018.a6
ISSN: 2040-4565

Citation
M. Eady and B. Park, "Unsupervised classification of individual foodborne bacteria from a mixture of bacteria cultures within a hyperspectral microscope image", *J. Spectral Imaging* **7,** a6 (2018). https://doi.org/10.1255/jsi.2018.a6

illness around the world. The organism is ubiquitous in the environment and can find its way into the food system through cross-contamination at stages of production, processing or distribution. Significant outbreaks have been noted in eggs, peanut butter and sprouts.[2]

Early and rapid detection of harmful microbial organisms is critical for several purposes. In the case of an outbreak, traditional methods such as the use of nutrient-enriched growth media, DNA-based methods or immunological detection may take up to 10 days for confirmation.[3] During this time the contaminated product may still be in the market causing illness. Early and rapid detection methods can offer food chain suppliers with the information necessary to prevent contaminated products from initially entering the market. Also, early detection of microbes in hospital settings could lead to a quicker response with targeted antibiotics for a specific infection-causing organism.

The use of hyperspectral imaging (HSI) and hyperspectral microscope imaging for food safety has seen some advancement in recent years.[4] HSI has been investigated as a method of *Enterobacteriaceae* counts on a chicken using PLSR.[5] Determination of shiga toxin-producing *E. coli* (STEC) serogroup using HSI to identify colonies growing on agar plates has been explored by two groups.[6,7] At the microscopic level, hyperspectral microscope imaging has been used to characterise deformed red and white blood cells.[8]

Our previous hyperspectral microscope imaging work has focused on classification of HMIs containing pure cultures of one target organism under a microscopic slide.[9] In moving forward with a food industry application of the HMI method for early and rapid detection, it is important to be able to identify target species of bacteria from HMIs containing more than one species of bacteria. Cluster analysis (CA) has seen many uses, including applications by biologists to classify various taxonomical species of animals.[10] CA is an unsupervised classification method that works by clustering together similar samples, with separate clusters forming for dissimilar samples. Here, the objective is to determine if CA can be used to differentiate individual cells in HMIs containing multiple bacteria cultures.

Materials and methods

Sample preparation

All bacteria cultures were obtained from the Poultry Microbiological Safety and Processing Research Unit located at the US National Poultry Research Center in Athens, GA, USA. Cultures were isolated from chicken broiler carcass rinses, purified and held at −80°C until needed. Short-term tryptic soy agar (TSA) slants were then prepared by inoculating the frozen cultures of *E. coli* (Ec), *Listeria innocua* (Li), *Staphylococcus aureus* (Sa) and *Salmonella* Typhimurium (ST) into TSA slants and incubating overnight at 37 ± 2°C prior to sampling. Samples were removed after overnight incubation and stored at 4°C for up to two weeks for hyperspectral microscope imaging data collections. The short-term agar slants were streaked onto TSA plates and incubated for 18–24 h at 37 ± 2°C.

Park *et al.*[11] describe the preparation of microscope sample slides for collecting HMIs of live cells. In brief, a colony from the TSA plate was picked with an inoculating loop and suspended in 100 µL of UV-sterilised water.[11] The suspension was then vortexed briefly. Three µL of the suspension was spread onto the centre of a glass slide and allowed to air dry for 15 min under a biosafety cabinet (Baker, Sanford, ME, USA). A cover slip was then applied with 0.8 µL of water, pressing firmly to remove any potential air bubbles. One inherent advantage of this detection method is sampling live-viable cells without the use of added stains or dyes. The common heat fixation method was avoided because it would damage the cells. The reference set of images was collected in this manner with one culture per hyperspectral microscope imaging slide. For the mixtures of two, three and four species, the same method was used, but 3 µL of each bacteria suspension were mixed together and 3 µL of that mixture was placed on a glass slide. The mixed culture HMIs were combinations of ST with the other three bacteria. HMI sets A–C were combinations of two bacteria sets A) = ST and Li, B) = ST and Sa, and C) = ST and Ec. Hyperspectral microscope imaging sets D–F contain three cultures D) = ST, Ec and Sa, E) = ST, Li and Sa, F) = ST, Ec and Li. Finally, all four species were combined into the G set of HMIs. For the mixed culture HMI sets, these bacteria were grown from culture in three repetitions, with three to six HMIs collected at each repetition for each A–G.

Hyperspectral microscope imaging system and image acquisition

The hyperspectral microscope imaging system consisted of a standard upright digital microscope (Eclipse, e80i, Nikon, Lewisville, TX, USA). To this was attached an acousto optical-tunable filter (AOTF) (Gooch and Housego, Ilminster, UK), with a 16-bit electron multiplying charge coupled device (EMCCD) (iXon Andor Technology, Belfast, UK) and a 24-W tungsten halogen (TH) light source (Ushio America, Cypress, CA, USA). The components of the hyperspectral microscope imaging system can be seen in Figure 1.

System spectral calibration was performed by collecting HMIs with a multi-ion discharge lamp (MIDL) for wavelength calibration (LightForm Inc., Ashville, NA, USA). A linear regression analysis was carried out using five distinguishable peaks from the wavelength calibration lamp with the exposure time set to 250 ms and with a gain of 1.6%, which maintained high signal-to-noise ratio for acquiring quality images from the bacterial cells. The five peaks from the calibration lamp were found at 486, 542, 586, 610 and 706 nm. The linear regression yielded an R^2 value = 1.00, with an equation of:

$$\lambda_i = 446 + 4x_i \quad (1)$$

where the λ_i = wavelength (nm), x_i = band number, R^2 = coefficient of determination. Radiometric calibration was carried out with a standard yield lamp for photonic calibration (SYLPH) (Lightform Inc., Ashville, NC, USA). Spectra were consistent with no abnormal peaks or patterns observed. The hypercube images for calibration and sampling were collected between 450 nm and 800 nm, every 4 nm, resulting in images captured at 89 spectral bands. Hypercubes consisted of (1002 × 1002 × 89) pixels. Using a micrometer, it was determined that the pixel size equals 0.0847 μm^2, while single bacterial cells have approximately between 100 pixels and 7500 pixels. *Salmonella* usually have 300–500 pixels, *Staphylococcus* have as few as 120, while *Listeria* have up to 700.

HMI processing

HMIs were converted in the ENVI software program (Harris Geospatial, Boulder, CO, USA). Regions of interest (ROIs) depicting single cells were extracted in a two-step method. First, a spatial ROI was drawn around the cell. Then, image intensity thresholds were determined by an intensity histogram at the 650 nm peak, which had highest scattering intensity from the HMIs. Thresholds

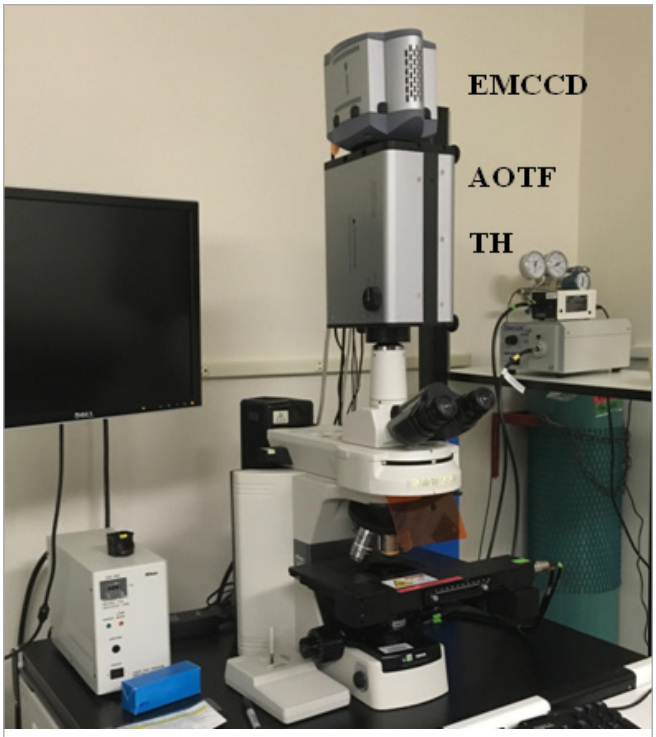

Figure 1. Hyperspectral microscope imaging system.

were then applied to each cell selected in an ROI to effectively extract pixels representing the backscatter signal from a single cell, while preventing under and oversaturated pixels from being extracted. From here, a mean spectrum was calculated for each bacteria cell. Using the mean spectra from each cell, the Mahalanobis distance (MD) from each sample to the class mean was calculated by:

$$\mathrm{MD} = d(x_i) = \left[(x_i - \bar{x})^T \; C^{-1} (x_i - \bar{x})\right]^{0.5} \text{ for } i = 1, \ldots, n \quad (2)$$

where x_i is an object vector and $\bar{x}$ is the centre of the cluster. The MD takes into account the distribution of objects in the variable space, independent of variable scaling.[12] Outliers were identified as MD values greater than ± 2 standard deviations from the mean MD value.[13] An important assumption of the MD is a normal distribution.[14] Here, the MD values representing distances from sample to class centroid were found to be normally distributed. From a sample size of 1220 cells, only 0.82% (10 cells) were identified as outliers and removed from the data set. Previous work using HMIs to classify single bacteria cells showed that cells identified as spectral outliers could be linked to damaged cells extracted from the HMIs.[15] Upon review of the HMIs these cells were

identified as severely damaged, appearing malformed, cleaved or with severely damaged cell walls. In our experiment, a total of 10 cells were identified as outliers. Here, we can see that the cell's backscatter pattern is not consistent. The ST cell appears divided, while the Sa cell is deformed and not a cocci shape, and the unknown cell obtained from a mixed culture image is also damaged. These inconsistencies most likely contribute to the outlier classification (see Supplementary Figure S1).

Data preprocessing

As previously mentioned, prior use of hyperspectral microscope imaging for pathogenic bacteria identification was conducted with sample slides prepared with one bacteria strain per slide. A two-step spectral pretreatment method was applied consisting of a normalisation step where the single cell mean spectra was normalised to the light source spectra at each collected wavelength. This was then followed by application of multiplicative scatter correction (MSC).[16] A slight difference in cell size was noticed in the reference sets of Sa, most likely due to the differences in minor cell growth size between experiment repetitions. Given that these HMIs of Sa were collected from the same strain, but grown at different times, a slight difference in size is not surprising. In first analysing the reference HMIs (single culture), a global piecewise multiplicative scatter correction (GPMSC), previously described by Burger and Geladi, was applied.[13] This was a two-step approach applying MSC by class first then globally. Here, the GPMSC was effective in removing the influence of cell size on the spectra, eliminating the repetition effect in intensity for Sa. If the goal of using HMIs is confirmation of a microbial isolate where one is fairly certain that there is only one organism per HMI, then the GPMSC showed promise for removing or reducing the influence of size from an organism's mean spectra. However, if the objective is to use hyperspectral microscope imaging for differentiation of two or more species in one HMI, then the MSC or GPMSC are not effective. Because MSC/GPMSC uses a class mean as a reference point to scale sample spectra, it will blend the two or more classes together, convoluting the resulting classification boundaries when *a priori* knowledge of cell cultures are not known. In order to reduce the spatial influence on the single cell mean spectra, standard normal variate (SNV) was used in place of MSC, and calculated by:

$$\tilde{x}_i = \frac{x_i - m_i}{\sigma_i} \tag{3}$$

where $\tilde{x}_i$ is the SNV adjusted spectra, m_i is the sample's mean, σ_i is the sample's standard deviation and x_i is the sample's spectra. These preprocessing steps are similar, but offer a subtle difference that is necessary to account for the unsupervised classification of mixed cultures for HMIs. The SNV does not require a class reference spectrum as the MSC does, because SNV subtracts the sample mean, and divides by the standard deviation for each sample.[17] Tungsten halogen light normalised spectra show slight rep effects; however, after applying SNV, more sample homogeneity from different experiments is observed, and reducing the rep effect on the spectra (see Supplementary Figure S2). This step can be done when more than one species of bacteria are present in an image, without blurring the proceeding cluster boundaries.

Multivariate data analysis

Analysis of spectral data was carried out in the Unscrambler software V.10.1 (Oslo, Norway). Here, a divisive CA is proposed with the objective of identifying individual cells at the species level, based on the cell's mean spectra. First, the CA approach was tested with a supervised reference data set taken from our spectral image library of common foodborne bacteria. Reference HMIs contained one culture per HMI for Ec, Li, Sa and ST. The four bacteria were grown from culture for either two or three repetitions with three total HMIs of each species used for the reference data set. Each level of the CA was carried out with a two-cluster *k*-means clustering method, with 50 iterations for cluster assignment, using the squared Euclidean as a measure of distance and cluster assignment. Intracluster similarity was used as a stopping mechanism. For the mixed cluster HMIs, a two-cluster *k*-means CA was performed. From here, each of the two clusters formed had another round of *k*-means cluster analysis carried out. If the results came back as all or most of the samples belonging to the same cluster, then the CA was stopped. Next, the CA method developed from the spectral library was applied to the HMIs containing a mixture of ST with one, two or three other bacteria species. Mean spectra from the unknown clusters were then compared to the reference data set, which contained single culture HMIs for all four bacteria. The "cos α" of the angle between two spectra were calculated by:

$$cos\ \alpha = \frac{x_A^T x_B}{\sqrt{\left(x_A^T\ x_A\right)\left(x_B^T\ x_B\right)}} = \frac{x_A^T x_B}{x_A \times x_B} \tag{4}$$

where $\cos\alpha$ is the angle α between two vectors and x_A and x_B are the two vectors. Here, x_A would be the mean spectra spectral library (e.g. either Ec, Li, Sa or ST in this study) and x_B is the mean spectra of the unknown samples. This is a similarity measure with the aim of identifying the unknown samples as one of the four species, based on the largest $\cos\alpha$ value between the unknown sample's mean spectra and the reference spectra.[16]

In an effort to further assess the efficiency of the CA to correctly identify single bacteria cells from the mixed culture HMIs, a MD-based discriminant analysis (DA) was applied. Spectral signatures obtained from bacteria cells are inherently multi-collinear. For this reason, the MD measure was chosen due to its ability to standardise by scaling in terms of standard deviation, and summing the pooled within-group variance–covariance, adjusting for correlations among variables.[10] DA is a supervised classification method. A categorical dummy variable was inserted into the matrix, containing the proposed CA classification result for each cell (i.e., 0, 1, 2 or 3). The DA was carried out using the spectra as predictors, and the proposed CA classification as the response. DA was performed with the first five principal components, which explained ≥ 95% of the variance between classes.

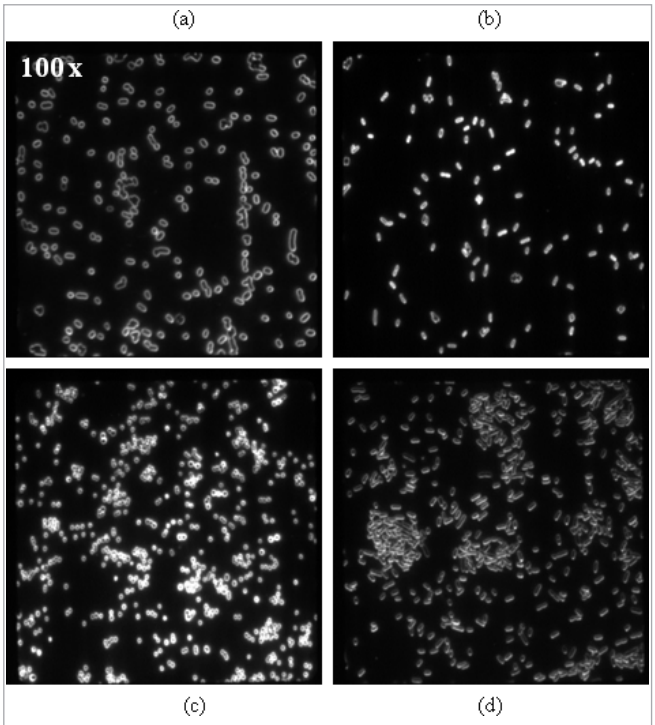

Figure 2. Hyperspectral microscope image bands collected at 650 nm for pure culture images (a) *E. coli*, (b) *S.* Typhimurium, (c): *S. aureus* and (d) *L. innocua*.

Results and discussion

Single culture samples

One experimental repetition of the single culture reference HMIs is shown in Figure 2. We can see that Sa in Figure 2(c) has higher scattering intensities than the other three images of rod-shaped bacteria. Cocci also have a tendency to cluster together, which can be seen in Figure 2(c). In Figure 2(d) the Li cells are more clumped together than the other rod images along with the characteristic slightly bent rod shape visible in some cells. The Ec (Figure 2a) and ST (Figure 2b) are the most similar, with no real discernible visual difference between the two. The mean spectra of cells from the reference images can be seen in Figure 3. The plot on the left (Figure 3a) shows the raw spectra of the four species. Spectra for Ec and ST are similar, while the Li and Sa mean spectra are not as similar to Ec and ST, which are in the same family of bacteria (*Enterobacteriaceae*) and have a similar short rod shape. The plot on the right (Figure 3b) shows the preprocessed spectra. We can see that preprocessing has brought the mean spectra of all four species closer together. This removes intensity variability in the spectra, while pattern deviations are present. The Sa spectra appear to be the most different from the others with changes in both pattern and intensity noticed. Overall, as the spectra approach the red colour bands between 680 nm and 800 nm, we see more variance in the spectral patterns.

Given these cells are imaged with a 100× oil objective in the visible light range, there are some benefits and limitations to what can be discerned from the HMIs. The backscatter signal collected from the cells with a 100× objective gives us some detail of an individual cell, but not the sensitivity to observe how selected pixels may represent specific cellular characteristics, such as flagellum or the actin polymerising tail of *Listeria*. This does give us an overview of the cell. Considering that these HMIs in the visible light range describe a generalisation of molecular and physical interactions with the tungsten halogen (TH) light occurring within a cell, but not necessarily specific biochemical processes.

A divisive dendrogram of the reference data set is shown in Figure 4. The three levels of clustering can be seen. Cellular shape is a deciding factor in the first level of the CA, resulting in 100% accuracy. The three rod-

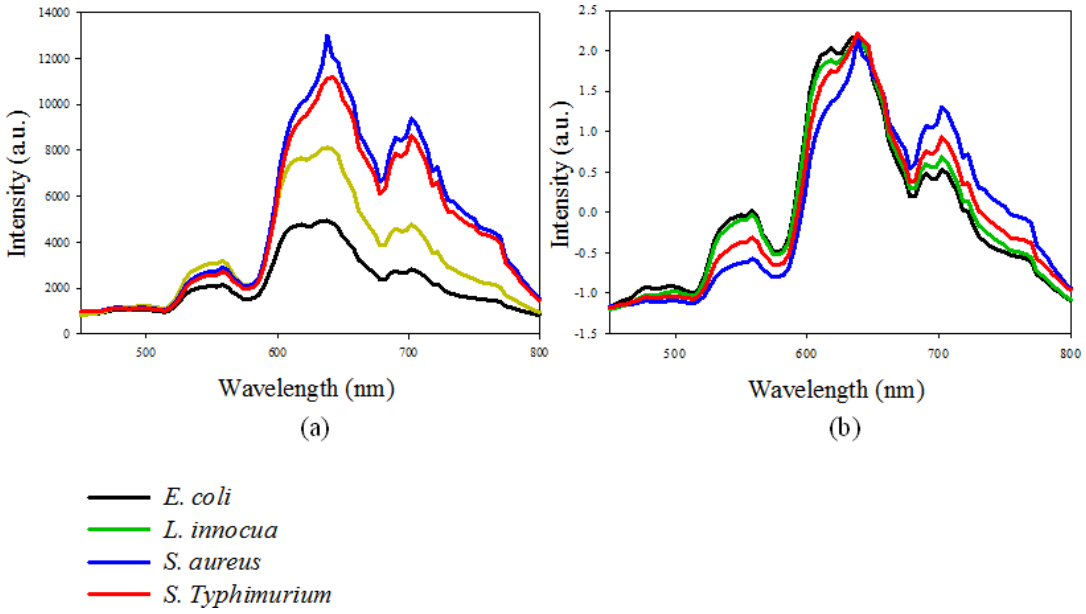

Figure 3. Mean spectra for four species obtained from single culture hyperspectral microscope images. (a) Raw spectra, (b) normalisation and standard normal variant preprocessing steps applied.

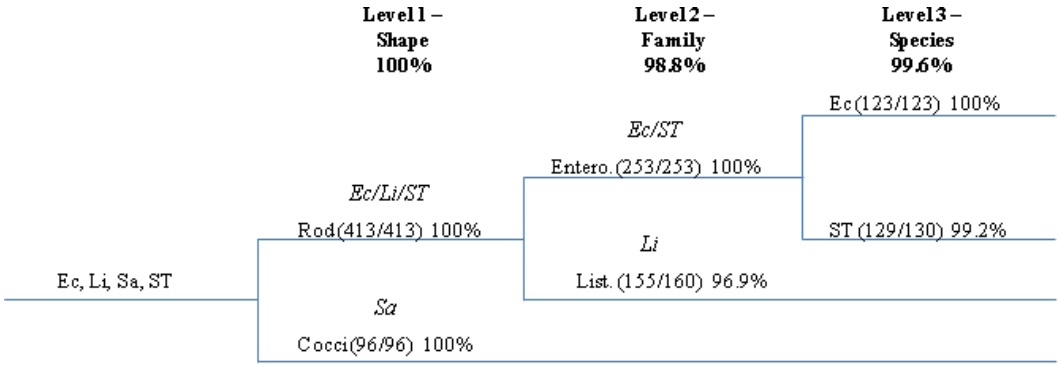

Figure 4. Divisive dendrogram showing results of a cluster analysis carried out on pure culture images of single culture hyperspectral microscope images.

shaped bacteria (Ec, Li and ST) ($N = 413$) represent one cluster, while the cocci (Sa) formed a second cluster alone ($N = 96$). Figure 5 shows the reference spectra for Ec, Li and Sa adjusted to ST, by subtracting the ST spectra for each. Ec and Li have similar patterns, but it is important to note the distinct difference in Sa pattern, showing this was the clearest CA cluster separation. From here, the second level of the CA differentiated between the family level of taxonomy with *Enterobacteriaceae* and *Listeriaceae*. This step had a total accuracy of 98.8%, with *Enterobacteriaceae* (Ec and ST) achieving 100% classification ($N = 253$), and *Listeriaceae* at 96.9% accuracy with 5 cells out of 160 misclassified. Finally, in the third CA step, the two *Enterobacteriaceae* species were differentiated. From biological and physiological standpoints these two species were most alike out of the four used. Both are members of the same family, are non-spore formers, have G– cell walls (do not retain crystal violet stain and have additional cell wall components not found in G+ bacteria), and are facultative anaerobes that use flagellum for motility.[18] These similarities in traits are responsible for requiring an additional level of CA. From the reference data, attempting to cluster Ec, Li and ST into three clusters after removing Sa failed. Li formed one cluster, while Ec and Li formed a second. The third level had a combined accuracy of 99.6%, Ec classifying at 100% with 123 cells and ST = 99.2% with 129 out of 130 correctly identified. Combining the three levels resulted in a total of 1175 classifications and 99.5% accuracy. The CA showed promise when tested under supervised

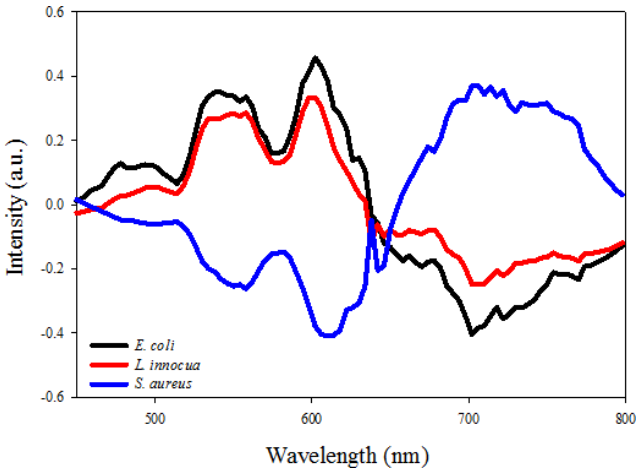

Figure 5. Mean species spectra for *E. coli*, *L. innocua* and *S. aureus* adjusted by the *S.* Typhimurium mean spectra.

terms, with levels of the CA being tied to shape and taxonomical differences between the four species.

Unsupervised classification of mixed cultures

Salmonella is of particular interest in this work due to the relevance to the poultry industry, and was prepared with a mixture of one, two and three other bacteria species. *Salmonella* was prepared in a mixture with each possible combination of the other three bacteria species for seven different sets of images (A–G). Table 1 lists the four bacteria with some physiological traits listed for comparison. We can see many similarities in Ec and ST. Mixed cultures were repeated in triplicate for each set, and ROIs of single cells were extracted for analysis as previously mentioned. Figure 6 shows two image slices from the hypercube, both at the 650 nm band. These images contain a mixture of species. Visually, it was usually possible to differentiate between some rod and cocci shapes (Figure 6a), but are by no means reliable through visual analysis alone. Typically, cocci are smaller than rods, at least in the case of these organisms. Smaller spherical shapes have higher intensity backscatter patterns.[12] It became increasingly more difficult to visually differentiate between the three rod bacteria as seen in Figure 6(b), likely containing Ec, Li, Sa and ST.

After performing the CA on the mixed culture HMIs, the $\cos\alpha$ of the angle between each of the reference mean spectra and the unknown cluster's mean spectra were calculated to attempt to identify the unknown cluster based on the reference data. These results can be seen in Table 2. Each image set (A–G) is listed along with the number of clusters from each set. The $\cos\alpha$ was calculated comparing the unknown cluster to Ec, Li, Sa and ST. The highest value by row represents the highest measure of similarity to the reference data, and is marked with an asterisk. The results in Table 2 were unreliable. The highest values of similarity appear to misrepresent the cluster in some image sets. Because the mean cluster spectra were compared to each of the four reference bacteria, even though certain species may not have been present in the image, misclassifications are obvious. It appears that most $\cos\alpha$ values appear to classify clusters as ST, indicating that $\cos\alpha$ is unreliable in terms of a validation tool for the CA. The spectra are highly collinear in nature making a spectral angle distance value difficult to determine from the $\cos\alpha$ values alone. Results from the DA can be seen in Table 3. Because the CA is somewhat

Table 1. Bacteria information on physical attributes and taxonomical levels.

		E. coli	*L. innocua*	*S. aureus*	*S.* Typhimurium
Taxonomy	Species[c]	*E. coli*	*L. innocua*	*S. aureus*	*S. enteric*
	Genus	*Escherichia*	*Listeria*	*Staphylococcus*	*Salmonella*
	Family[b]	*Enterobacteriaceae*	*Listeriaceae*	*Staphylococcaceae*	*Enterobacteriaceae*
	Order	*Enterobacteriales*	*Bacillales*	*Bacillales*	*Enterobacteriales*
Physical attributes	Shape[a]	Rod	Rod	Cocci	Rod
	Cell Wall	G–	G+	G+	G–
	Motility	Yes—flagella	Yes—flagella*	No	Yes—flagella
	Metabolism	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic

**L. innocua* motility consists of flagella, but at 37°C it polymerises actin from a host cell to form a "comet" tail. [a]First level of cluster analysis, [b]second level of cluster analysis, [c]third level of cluster analysis.

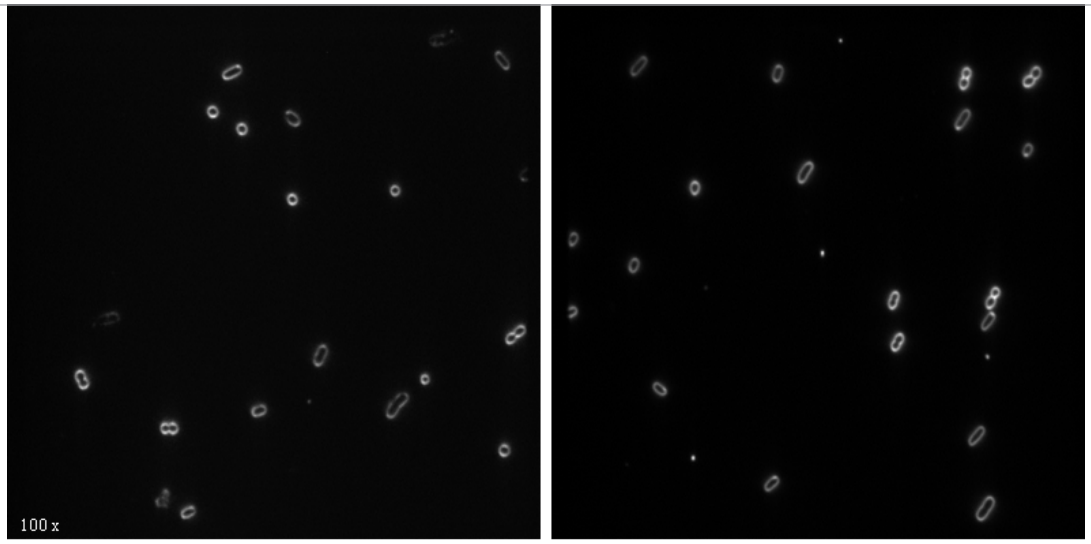

Figure 6. 650 nm band from hyperspectral microscope images of mixed bacteria cultures. (a) *S. aureus* and *S.* Typhimurium, (b) *E. coli*, *L. innocua*, *S. aureus* and *S*. Typhimurium.

Table 2. The angle the between reference set mean spectra (pure culture images) and the unknown cluster's mean spectra (mixed images) calculated as the cos α.

Set	Bacteria	Cluster	Ec	Li	Sa	ST
A	ST, Li	0	0.989	0.989	0.985	0.996*
		1	0.987	0.988	0.996*	0.987
B	ST, Sa	0	0.974	0.972	0.982	0.990*
		1	0.981	0.977	0.979	0.992*
C	ST, Ec	0	0.992*	0.980	0.976	0.992*
		1	0.973	0.977	0.992	0.994*
D	ST, Ec, Sa	0	0.980	0.984	0.994*	0.991
		1	0.994*	0.984	0.992	0.981
		2	0.985	0.987	0.989	0.996*
E	ST, Li, Sa	0	0.980	0.978	0.980	0.993*
		1	0.980	0.981	0.994*	0.988
		2	0.981	0.982	0.988	0.994*
F	ST, Ec, Li	0	0.974	0.981	0.985	0.989*
		1	0.989	0.978	0.991*	0.988
		2	0.982	0.978	0.988	0.993*
G	ST, Ec, Li, ST	0	0.975	0.962	0.954	0.981*
		1	0.977	0.965	0.958	0.983*
		2	0.967	0.963	0.951	0.981*
		3	0.967	0.954	0.952	0.977*

*Indicates the closest similarity between unknown cluster mean spectra and reference spectra.

Table 3. Cluster analysis results (unsupervised) for mixtures of bacteria cells, with confirmation through a discriminant analysis to confirm accuracy percentages.

Set	Bacteria	Cluster	Correctly classified	Accuracy (%)	Total accuracy (%)
A	ST, Li	0	49/50	98	98.91
		1	42/42	100	
B	ST, Sa	0	72/72	100	100
		1	19/19	100	
C	ST, Ec	0	33/34	97.06	97.06
		1	66/68	97.06	
D	ST, Ec, Sa	0	30/31	95.24	91.92
		1	19/25	76	
		2	42/43	97.67	
E	ST, Li, Sa	0	53/55	96.36	97.56
		1	39/39	100	
		2	28/29	96.55	
F	ST, Ec, Li	0	30/30	100	98.89
		1	53/54	98.15	
		2	6/6	100	
G	ST, Ec, Li, ST	0	21/21	100	98.06
		1	32/33	96.97	
		2	23/24	95.83	
		3	25/25	100	

[a]Ec = *E. coli*, Li = *L. innocua*, Sa = *S. aureus* and ST = *S.* Typhimurium.

of an exploratory analysis, the DA was calculated to assess the proposed clustering of the CA. Here, unknown samples were assigned to clusters 0, 1, 2 or 3 (cluster ID), depending on how many clusters were formed per HMI. Using the proposed cluster number as the response variable DA classification accuracies ranged between 91.9% and 100% for mixed culture HMIs. Image set D (Ec, Sa and ST) had one cluster formed through the CA that had six cells misclassified through the DA, resulting in the lowest image accuracy of 91.9%. The other six image sets had classification accuracies > 97%. Because the cluster's identification was unreliable with the $\cos\alpha$ it is not clear as to the cause of the six misclassified cells. G was the final image set and it contained all four bacteria with a DA accuracy of CA cluster assignment of 98.1%.

Cellular walls contain large varieties of proteins unique to the organism and its survival mechanisms.[19] The Frölich theory states that cellular structures exhibit coherent longitudinal vibrations of electrically polar structures.[20,21] Most proteins are electrically polar structures.[22] Given the constant TH light source used for HMI collection, the backscatter signal we are observing is potentially from the steady-state, non-linear vibration of those electrically polar protein structures found in the cell walls, unique to the organism being imaged. Park *et al.* showed that Gram-positive cell walls of *S. aureus* differed from the Gram-negative cell walls of *Salmonella*.[23] Here, CA portioned the individual samples into clusters and DA verified the accuracy of the clustering. The unsupervised classification of CA may be possible due in part to the spectral differences influenced by the diversity of backscatter signal caused by the unique profiles of polar protein structures found in the cell walls.

Figure 7 shows several mixed culture HMIs at the 650 nm band, with cell classification noted on the images. The top row is the HMI band from 650 nm, and the bottom row is the same image with the cell assignment based on unsupervised method marked. Column A is a mixture of Sa and ST. These cells were the most different in terms of shape, and are clearly separated by the spectral information. Sa cells can be seen in the image as being smaller, and showing higher levels of pixel inten-

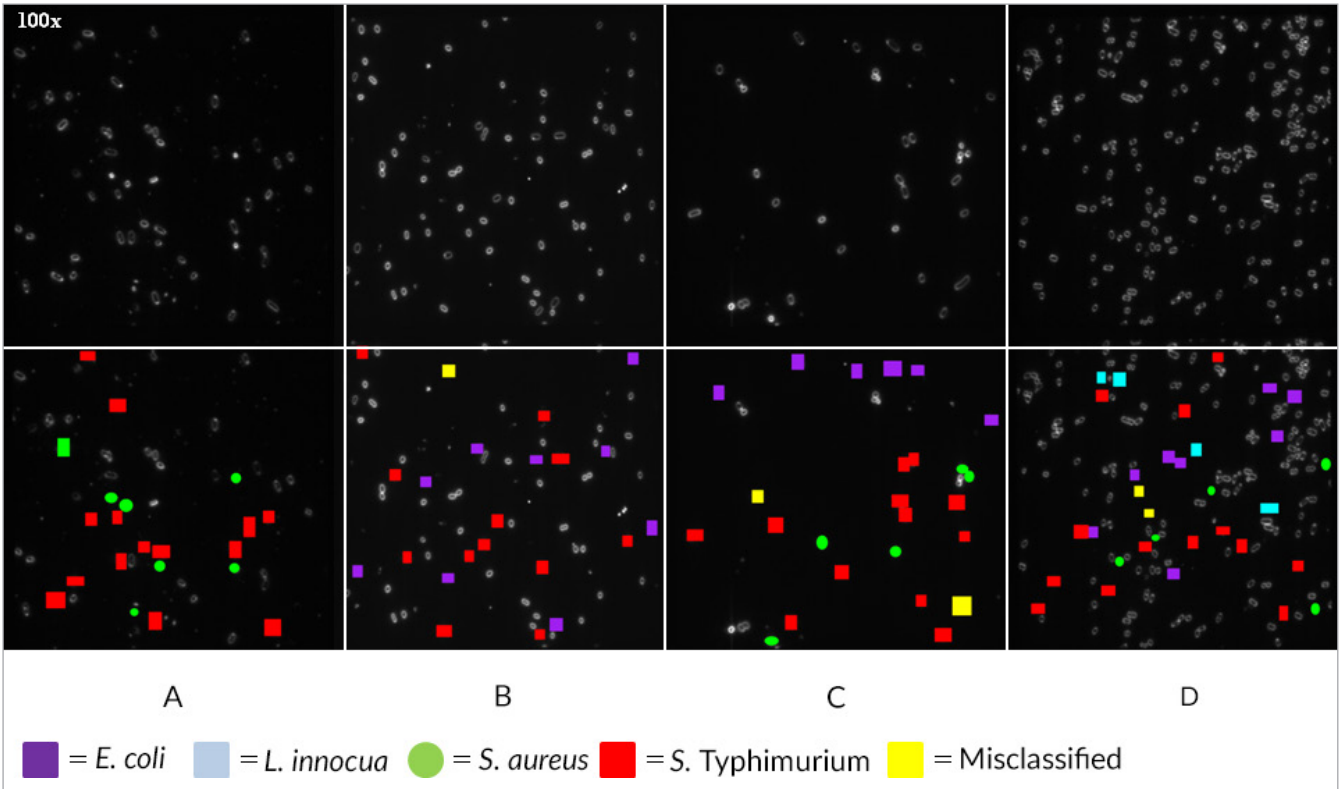

Figure 7. Hyperspectral microscope images collected at 650 nm of mixed bacteria cultures. Top row: 650 nm image, bottom row: the same image with cells marked as cluster analysis results. Columns (a) *S. aureus* and *S.* Typhimurium, (b) *E. coli* and *S.* Typhimurium, (c) *E. coli*, *S. aureus* and *S.* Typhimurium, (d) *E. coli*, *L. innocua*, *S. aureus* and *S.* Typhimurium.

sity. Column B shows the Ec and ST mixture image. It is difficult to see any difference between the two species, given that they are both members of the same bacteria family. Spectral information was similar as well, requiring the removal of Li and Sa from the CA, before these two species could be differentiated. In column C we have a mixture of Ec, Sa and ST cells. A few Sa cells are present. This image was taken from a different Sa growth repetition and the cells are slightly larger than the Sa cells seen in column A of Figure 7. This reduces the cell's pixel intensity slightly, but the spectral information was clearly different from that of the three rod-shaped bacteria. Finally, in column D we have a mixture of all four species. Only a few Li cells were found in this image. Combining all reps with four images had an accuracy of 98.1% when verified by the DA.

Conclusions

Moving towards the use of hyperspectral microscope imaging as an applied early and rapid detection tool in the food industry, we investigated the use of a *k*-means CA to differentiate unknown cells in HMIs containing mixed bacteria cultures. The unsupervised CA classification was followed with a discriminant analysis, in which a dummy variable containing the CA class results per cell was inserted and used as the response variable. The cos-α of the angle from unknown clusters to reference data was unsuccessful in clearly identifying each cluster in the mixed culture HMIs due to the high collinearly of the data and small alpha (α) space between vectors. Individual cells from the mixture of two, three or four cultures saw classification accuracies > 91.9%. This showed promise in the ability of a CA to perform unsupervised classification of bacteria cells. However, the use of the cos-α as a validation tool for comparing unknown cluster mean spectra to a reference data set failed, potentially due to the high collinearity of the spectra. Perhaps the use of scale-invariant shape metrics and statistical moments can be used to assign a cluster species classification. In a real-world application of this technology, target organisms such as *Salmonella* will need to be classified against an indefinite number of other microbial organisms. This study is meant to be a step towards early and rapid classification of *Salmonella* from a food

matrix by investigating the ability to classify *Salmonella* from several other common foodborne bacteria in mixed culture HMIs. Future efforts will be aimed at enhancing HMI detection sensitivity with bacteria isolated from food matrices.

Acknowledgements

The authors would like to acknowledge Dr Nasreen Bano of the US National Poultry Research Center, located in Athens, GA, USA, for her contributions to this project.

References

1. WHO, *Food Safety*. Vol. 2017, World Health Organization (2017) [Accessed 1 September 2017]. http://www.who.int/foodsafety/areas_work/foodborne-diseases/en/
2. N.A. Cox, J.G. Frye, W. McMahon, C.R. Jackson, J. Richardson, D.E. Cosby, G. Mead and M.P. Doyle, "Salmonella", in *Compendium of Methods for the Microbiological Examination of Foods*, Ed by M.P. Doyle and R.L. Buchanan. ASM Press, Hendon, VA, USA, p. 225 (2013).
3. O. Lazcka, F.J. Del Campo and F.X. Munoz, "Pathogen detection: a perspective of traditional methods and biosensors", *Biosens. Bioelectr.* **22(7),** 1205–1217 (2007). doi: https://doi.org/10.1016/j.bios.2006.06.036
4. A.A. Gowen, Y. Feng, E. Gaston and V. Valdramidis, "Recent applications of hyperspectral imaging in microbiology", *Talanta* **137,** 43–54 (2015). doi: https://doi.org/10.1016/j.talanta.2015.01.012
5. Y.Z. Feng, G. ElMasry, D.-W. Sun, A.G.M. Scannell, D. Walsh and N. Morcy, "Near-infrared hyperspectral imaging and partial least squares regression for rapid and reagentless determination of *Enterobacteriaceae* on chicken fillets", *Food Chem.* **138(2–3),** 1829–1836 (2013). doi: https://doi.org/10.1016/j.foodchem.2012.11.040
6. W.R. Windham, S.C. Yoon, S.R. Ladely, J.A. Haley, J.W. Heitschmidt, K.C. Lawrence, B. Park, N. Narrang and W.C. Cray, "Detection by hyperspectral imaging of shiga toxin–producing *Escherichia coli* serogroups 026, 045, 0103, 0111, 0121, and 0145 on rainbow agar", *J. Food Prot.* **76(7),** 1129–1136 (2013). doi: https://doi.org/10.4315/0362-028X.JFP-12-497
7. Y. Tang, H. Kim, A.K. Singh, A. Aroonnual, E. Bae, B. Rajwa, P.M. Fratamico and A.K. Bhunia, "Light scattering sensor for direct identification of colonies of *Escherichia coli* serogroups 026, 045, 0103, 0111, 0121, 0145 and 0157", *PLoS One* **9,** e105272 (2014). doi: https://doi.org/10.1371/journal.pone.0105272
8. G.S. Verebes, M. Melchiorre, A. Garcia-Leis, C. Ferreri, C. Marzetti and A. Torreggiani, "Hyperspectral enhanced dark field microscopy for imaging blood cells", *J. Biophot.* **6(11–12),** 960–967 (2013). doi: https://doi.org/10.1002/jbio.201300067
9. B. Park, "Future trends in hyperspectral imaging", *NIR news* **27(1),** 35–38 (2016). doi: https://doi.org/10.1255/nirn.1583
10. J. Hair, W. Black, B. Babin and R. Anderson, *Multivariate Data Analysis*. Prentice Hall, Upper Sadle River, NJ, USA (2010).
11. B. Park, S. Yoon, S. Lee, J. Sundaram, W. Windham, A. Hinton Jr and K. Lawrence, "Acousto-optic tunable filter hyperspectral microscope imaging method for characterizing spectra from foodborne pathogens", *Trans. ASABE* **55(5),** 1997–2006 (2012). doi: https://doi.org/10.13031/2013.42345
12. K. Varmuza and P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*. CRC Press, Boca Raton, FL, USA (2009). doi: https://doi.org/10.1201/9781420059496
13. J. Burger and P. Geladi, "Spectral pre-treatments of hyperspectral near infrared images: analysis of diffuse reflectance scattering", *J. Near Infrared Spectrosc.* **15(1),** 29–37 (2007). doi: https://doi.org/10.1255/jnirs.717
14. T. Fearn, "Limitations of Mahalanobis and H distances", *NIR news* **22(8),** 16–17 (2011). doi: https://doi.org/10.1255/nirn.1279
15. M. Eady, B. Park, S.C. Yoon, M. Haidekker and K. Lawrence, "Methods for hyperspectral microscope calibration and spectra normalization from images of bacteria cells", *Trans. ASABE* **61(2),** 1–12 (2018). doi: https://doi.org/10.13031/trans.12222
16. B. Park and M. Eady, "New application of hyperspectral imaging for bacterial cell classification", *NIR news* **27(8),** 4–6 (2016). doi: https://doi.org/10.1255/nirn.1647
17. S.C. Yoon and B. Park, "Hyperspectral image processing methods", in *Hyperspectral Imaging*

Technology in Food and Agriculture, Ed by B. Park and R. Lu. Springer, New York, NY, USA, p. 91 (2015). doi: https://doi.org/10.1007/978-1-4939-2836-1_4

18. J. Jay, M. Loessner, and D. Golden, *Modern Food Microbiology*. Springer Science & Business Media, New York, NY, USA (2008).

19. T.J. Silhavy, D. Kahne and S. Walker, "The bacterial cell envelope", *CSH Perspect. Biol.* **2,** a000414 (2010). doi: https://doi.org/10.1101/cshperspect.a000414

20. H. Fröhlich, "Long-range coherence and energy storage in biological systems", *Int. J. Quant. Chem.* **2(5),** 641–649 (1968). doi: https://doi.org/10.1002/qua.560020505

21. H. Fröhlich, "Bose condensation of strongly excited longitudinal electric modes", *Phys. Lett. A* **26(9),** 402–403 (1968). doi: https://doi.org/10.1016/0375-9601(68)90242-9

22. M. Cifra, J.Z. Fields and A. Farhadi, "Electromagnetic cellular interactions", *Prog. Biophy. Molec. Biol.* **105(3),** 223–246 (2011). doi: https://doi.org/10.1016/j.pbiomolbio.2010.07.003

23. B. Park, Y. Seo, S.C. Yoon, A. Hinton Jr, W.R. Windham and K.C. Lawrence, "Hyperspectral microscope imaging methods to classify gram-positive and gram-negative foodborne pathogenic bacteria", *Trans. ASABE* **58(1),** 5–16 (2015). doi: https://doi.org/10.13031/trans.58.10832

Supplementary figures

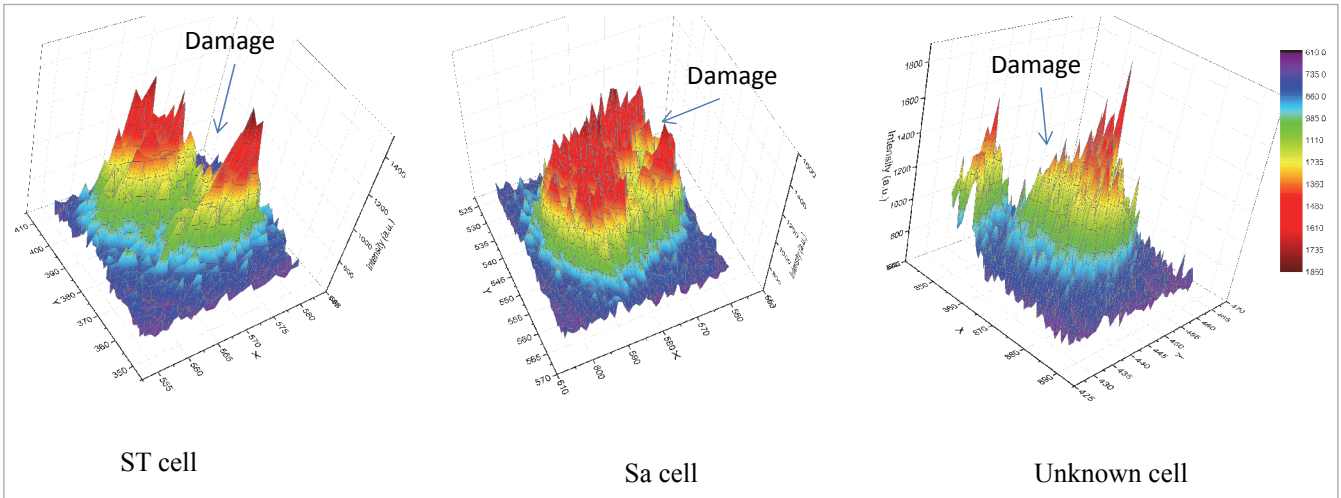

Figure S1. 3D surface plots of three damaged cells determined by Mahalanobis distances to be statistical outliers. The unknown cell was obtained from an image containing both ST and Sa. ST = *S.* Typhimurium and Sa = *S. aureus*.

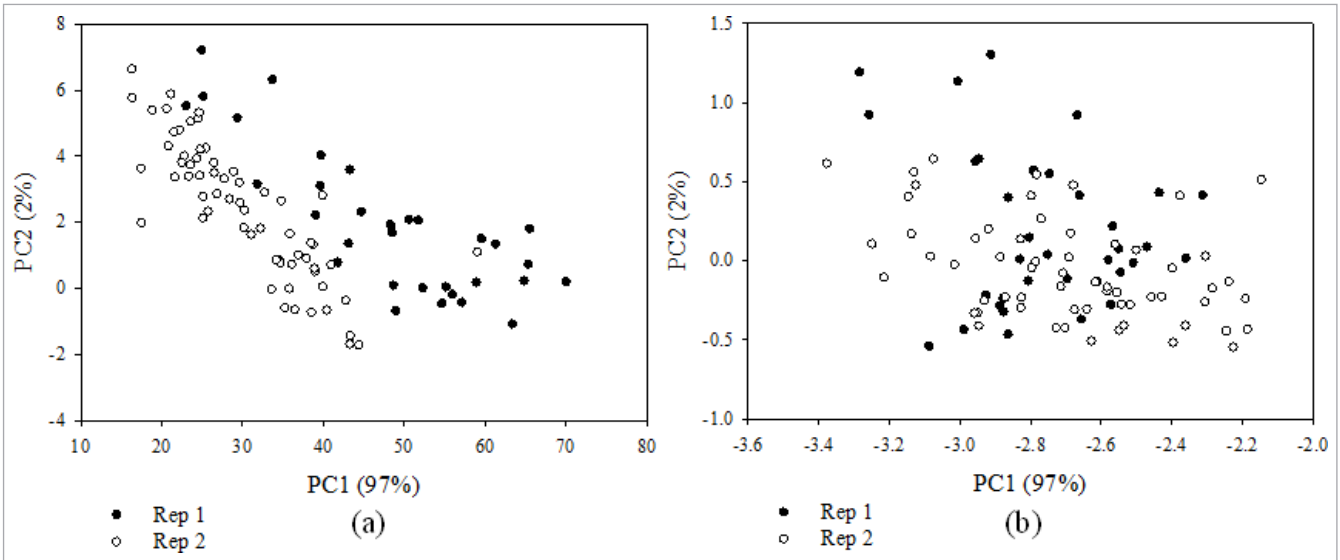

Figure S2. PCA score plots for *S. aureus.* (a) Tungsten halogen normalised spectra, (b) additional standard normal variant treated spectra.