

Peer Reviewed Paper **openaccess**

# Effect of colony age on near infrared hyperspectral images of foodborne bacteria

Paul J. Williams,<sup>a\*</sup> Terri-Lee Kammies,<sup>b</sup> Pieter A. Gouws<sup>c</sup> and Marena Manley<sup>d</sup>Department of Food Science, Stellenbosch University, Private Bag X1, Matieland (Stellenbosch) 7602, South Africa. E-mail: [pauljw@sun.ac.za](mailto:pauljw@sun.ac.za)<sup>a</sup> <https://orcid.org/0000-0002-6014-2049>, <sup>b</sup> <https://orcid.org/0000-0001-5382-8395><sup>c</sup> <https://orcid.org/0000-0002-7061-2403>, <sup>d</sup> <https://orcid.org/0000-0001-7581-7208>

Near infrared hyperspectral imaging (NIR-HSI) and multivariate image analysis were used to distinguish between foodborne pathogenic bacteria, *Bacillus cereus*, *Escherichia coli*, *Salmonella* Enteritidis, *Staphylococcus aureus* and a non-pathogenic bacterium, *Staphylococcus epidermidis*. Hyperspectral images of bacteria, streaked out on Luria–Bertani agar, were acquired after 20h, 40h and 60h growth at 37°C using a SisuCHEMA hyperspectral pushbroom imaging system with a spectral range of 920–2514nm. Three different pre-processing methods: standard normal variate (SNV), Savitzky–Golay (1<sup>st</sup> derivative, 2<sup>nd</sup> order polynomial, 15-point smoothing) and Savitzky–Golay (2<sup>nd</sup> derivative, 3<sup>rd</sup> order polynomial, 15-point smoothing) were evaluated. SNV provided the most distinct clustering in the principal component score plots and was thus used as the sole pre-processing method. Partial least squares discriminant analysis (PLS-DA) models were developed for each growth period and was tested on a second set of plates, to determine the effect the age of the colony has on classification accuracies. The highest overall prediction accuracies where test plates required the least amount of growth time, was found with models built after 60h growth and tested on plates after 20h growth. Predictions for bacteria differentiation within these models ranged from 83.1% to 98.8% correctly predicted pixels.

**Keywords:** colony age, foodborne bacteria, growth media, PLS-DA, NIR hyperspectral imaging

## Introduction

The widespread occurrence of foodborne diseases in developed and developing countries indicate underlying food safety flaws in adequate pathogen detection.<sup>1</sup> Conventional methods for pathogen detection and identification are labour intensive and time consuming—often taking two to three days for initial results (enrichment and growth on specialised media) and up to a week for species identification. Pure culture identifications are typically required for regulatory and legal purposes.<sup>2</sup> Thus, rapid methods for pathogen detection and improvements on existing methods are constantly being investigated.

Recently, near infrared (NIR) hyperspectral imaging (HSI) has been explored as a tool to differentiate pathogenic bacteria from each other as well as from non-pathogenic bacteria.<sup>3–6</sup> Barbin *et al.*<sup>7</sup> determined total viable count (TVC) and psychrotrophic plate count (PPC) on porcine meat over 21 days. Chemical changes in the meat, due to bacterial growth, were used to predict whether samples were commercially sound (those that had a bacterial count lower than  $\log 6$  CFU g<sup>-1</sup>). Linear discriminant analysis (LDA) models showed over 95% accuracy for spoilage detection,

### Correspondence

Paul J. Williams ([pauljw@sun.ac.za](mailto:pauljw@sun.ac.za))**Received:** 1 November 2018**Revised:** 14 December 2018**Accepted:** 4 January 2019**Publication:** 30 January 2019**doi:** 10.1255/jsi.2019.a5**ISSN:** 2040-4565

### Citation

P.J. Williams, T.-L. Kammies, P.A. Gouws and M. Manley, "Effect of colony age on near infrared hyperspectral images of foodborne bacteria", *J. Spectral Imaging* 8, a5 (2019). <https://doi.org/10.1255/jsi.2019.a5>

© 2019 The Authors

This licence permits you to use, share, copy and redistribute the paper in any medium or any format provided that a full citation to the original paper in this journal is given.



while partial least squares (PLS) achieved coefficients of determination of 0.82 and 0.85 for TVC and PPC, respectively. In a study by Yoon *et al.*,<sup>8</sup> NIR hyperspectral imaging was used to facilitate the detection and identification of *Campylobacter* and non-*Campylobacter* species, using the 400–900 nm wavelength range. They incubated colonies for 48 h on three different agars (Campy–Cefex, Campy–Line and blood agar) and developed a two-step classification algorithm to differentiate *Campylobacter* colonies from non-*Campylobacter* colonies (*Sphingomonas paucimobilis*, *Acinetobacter baumannii*, *Brevundimonas diminuta*, *Ochrobacterium sp.* and *Flavobacterium odoratum*). Cultures grown on Cefex had the highest classification accuracies (up to 99%). In an attempt to increase the rapidity of this method, the authors conducted a follow-up study.<sup>4</sup> They incubated the same organisms for a shorter period (24 h) before imaging, and compared the results to their previous study. Three different classification methods (band ratio algorithm, spectral feature fitting and single band thresholding algorithm) were applied. With the single band thresholding algorithm, the overall classification accuracy of the 48 h Cefex cultures (97%) was higher than the overall 24 h Cefex cultures (86%). Blood agar cultures incubated for 24 h showed up to 99% accuracy with the band ratio algorithm. These results demonstrated that it was possible to detect *Campylobacter* species using NIR-HSI after 24 h of growth.

When an old or damaged bacterial culture is introduced onto fresh media, the culture will undergo the normal growth phases (lag, exponential, stationary and death) in a closed system.<sup>9,10</sup> Various genera of bacteria have different generation times and will thus not display growth phases at the same time.<sup>9,10</sup> Furthermore, it is well known that as bacterial colonies proliferate and age on growth medium, various biochemical and physical changes occur<sup>11,12</sup> and these, too, depend on the organism. Thus, colony age could prove to be an important factor in certain detection methods, as various metabolites could hinder detection or inhibit reagents. Coutinho *et al.*<sup>13</sup> showed that colony age had a negative effect on polymerase chain reaction fingerprinting. The authors found that banding patterns were more consistent and clearer in younger (two days) than in older cultures (five days). In another study, Arnold *et al.*<sup>14</sup> noted profound changes in the mass spectra of *Escherichia coli* strains with time. The variation was attributed to nutrient depletion, metabolite and waste accumulation and cell

death. In both cases, the effect of colony age was self-explanatory, i.e., using young cultures produce different results than older colonies, and would influence identification. To date, no studies investigated the effect of colony age on NIR hyperspectral images of bacteria. Thus, this study aims to evaluate the effect of colony age on classification of foodborne bacteria.

## Materials and methods

### Sample preparation

Bacterial isolates of *Bacillus cereus* (ATCC 13061), *Escherichia coli* (ATCC 25922), *Salmonella* Enteritidis (ATCC 13076), *Staphylococcus aureus* (ATCC 25923) and *S. epidermidis* (ATCC 12228) were acquired from the culture collection of the Department of Food Science, Stellenbosch University. Two sets of streak plates were prepared for each bacterium using Luria–Bertani (LB) agar in 100 mm glass Petri dishes and incubated for 60 h at 37 °C. Streak plates were preferred as not only do they produce single colonies, but also large areas of bacterial growth where the initial streak was made. This permits the collection of more spectral data from the bacteria for analyses. LB agar was chosen because it is a common, general growth medium which contains only tryptone, yeast extract, sodium chloride and bacteriological agar powder.<sup>15</sup> To minimise the spectral response from the growth media, only LB agar was used throughout this study. Furthermore, we decided to refrain from using selective media as these could result in the production of substances that would interfere with NIR hyperspectral imaging measurements. Plates were removed from the incubator every 20 h, digitally imaged and then imaged with the NIR hyperspectral system. All inoculations were performed under aseptic conditions and bacterial cultures were screened for purity after NIR hyperspectral images were collected. To ensure images were not affected by the high incubation temperature, plates were left at 21 °C for 20 min before imaging.

### NIR hyperspectral imaging system and image acquisition

A SisuCHEMA short wave infrared camera (Specim, Spectral Imaging Ltd, Oulu, Finland) was used to acquire NIR hyperspectral images. The sensor consisted of a 2D array mercury-cadmium-telluride (HgCdTe) detector, with a light source of quartz halogen lamps as described

in Reference 5. Once the imaging system was calibrated, images of the entire Petri dish, with the lid on to avoid contamination, were collected every 20h for a period of 60h. These time points were chosen arbitrarily, to explore the effects of growth time on hyperspectral images, to determine if changes in culture age can be visualised and assess its effect on classification accuracies.

In total, 30 images were collected (five different bacteria, imaged at three time-intervals, in duplicate). Using streak plates allowed many single colonies to be imaged, incorporating more variation within a plate due to biochemical differences between colonies. In total (duplicate plates included), 10 images were collected. To compensate for effects of the growth media, images of Petri dishes with only agar were also acquired.

## Data analysis

Individual images were imported into Evince v.2.7.0 (Prediktera AB, Umeå, Sweden) hyperspectral image analysis software package. Principal component analysis (PCA) was calculated with three components before image segmentation. This involved removing the background, agar, bad pixels and reflection from the Petri dish with the brushing technique.<sup>16</sup> Noisy wavelengths from 920nm to 1097nm and 2477nm to 2514nm were removed and mosaics were then constructed with the cleaned images, before PCA was recalculated. To facilitate data analysis, individual images were grouped to form three mosaics, each containing a different combination of bacteria (Table 1). A mosaic was constructed for each of the growth periods, with a total of nine mosaics. This made it possible to study the bacteria and visualise differences in one image (mosaic). Similar to our previous study,<sup>5</sup> the bacteria were selected and grouped to include a variety of characteristics, i.e.

- Gram-positive and Gram-negative pathogens
- Bacteria which appeared similar in colour on the growth media, as well as those which appeared different (to prove that colour does not affect the method)

**Table 1.** Grouping of bacteria for mosaic formation, used for multivariate data analysis.

| Group 1               | Group 2               | Group 3               |
|-----------------------|-----------------------|-----------------------|
| <i>B. cereus</i>      | <i>B. cereus</i>      | <i>S. aureus</i>      |
| <i>E. coli</i>        | <i>S. aureus</i>      | <i>S. epidermidis</i> |
| <i>S. Enteritidis</i> | <i>S. epidermidis</i> |                       |

■ *S. epidermidis* was included to determine whether it was possible to distinguish between pathogens and non-pathogens.

However, in this study we aim to determine whether colony age would influence the ability to identify these characteristics and thus the classification accuracies. To achieve defined clustering and optimum separation in PCA score plots, thereby improving subsequent classification results, three pre-processing treatments were evaluated. We investigated SNV and Savitzky–Golay first derivative (2<sup>nd</sup>-order polynomial; 15-point smoothing) and second derivative (3<sup>rd</sup>-order polynomial; 15-point smoothing). Baseline correction was excluded, as this was done in combination with smoothing when Savitzky–Golay polynomial derivatives are calculated. Score plots and score images were inspected for clustering (score plots) and colour differences based on score values (score images). In the score images, warm colours (yellow to red) are indicative of high score values whereas cold colours (blue to cyan) show lower score values. A pre-processing method/combination was considered suitable when a distinct clustering of pixels for each bacterium, with minimal overlap, was observed.

The data was processed further in Matlab (The Mathworks Inc., Natick, MA, USA) and PLS\_Toolbox with MIA\_Toolbox 8.6.2 (Eigenvector Research, Inc., Manson, WA, USA). To determine the effect colony age has on classification, PLS-DA models were developed for each group, at each time interval and were validated independently on the duplicate plates of each time interval. The optimum number of latent variables for each model was determined using venetian blinds cross-validation with 10 splits and 20 samples per split. As too few samples were available, we decided to perform all data analyses with the pixel-wise approach. Had the object-wise approach (one image per Petri dish) been adopted, only 30 objects would be available for analysis. This was considered too few samples for classification. The data analysis process is shown in Figure 1.

## Results and discussion

Digital images of each Petri dish were collected after each growth period, before hyperspectral image acquisition, using a Huawei IDEOS S7 camera (3.15 MP, 2048 × 1536 pixels). An example of the digital images is shown in Figure 2.

## Pre-processing

After background removal and image correction, each image comprised at least 15,000 pixels per bacteria. All calculations were done on the duplicate plates individually, yielding similar results and thus duplicate plates were deemed sufficient. When studying the PCA score plots (Figures 3a–d), it was found that SNV-pretreated data produced the best clustering for all groups over the three growth periods. Bacteria were imaged in glass Petri dishes with the lid on, when images were acquired, irregularities in the glass and reflection of the agar could have caused some scattering of light. Moreover, the elevation of all colony morphologies was raised, this slight curvature could have contributed to light scatter as well. As there were a large number of sources for possible scattering, removing it with SNV was an essential step.

Score plots and score images of groups 1 and 2 after 40h growth were used as examples to illustrate the effect of pre-processing method (Figures 3–6). For group one (Figure 3), the Gram-positive *B. cereus* was represented by the cluster with a red centre (highest density of pixels)

in each score plot (except Figure 3c), whereas Gram-negative *E. coli* and *S. Enteritidis* were represented by clusters with a yellow/cyan centre. Score plots of raw data, as well as first and second derivative pre-processed data (Figure 3a, c and d) showed either one or two clusters. Only in the score plot of SNV-transformed data, were three clusters clearly visible (Figure 3b). *B. cereus* was separated from *E. coli* and *S. Enteritidis* along PC 1 [30% sum of squares (SS)] and *E. coli* was separated from *S. Enteritidis* along PC2 (11.8% SS).

In the score image of the raw data (Figure 4a) each bacterium appears different, with varying levels of score values (positive and negative), when compared to each other. This is possibly due to scattering effects caused by varying morphology of the bacterial colonies on the growth media. In the pre-processed score images (Figure 4b, c and d), *E. coli* and *S. Enteritidis* seem similar, whereas *B. cereus* is different, signifying the presence of notable chemical variances between Gram-positive *B. cereus* and the two Gram-negative bacteria—*E. coli* and *S. Enteritidis*. Differences between Gram-positive and Gram-negative

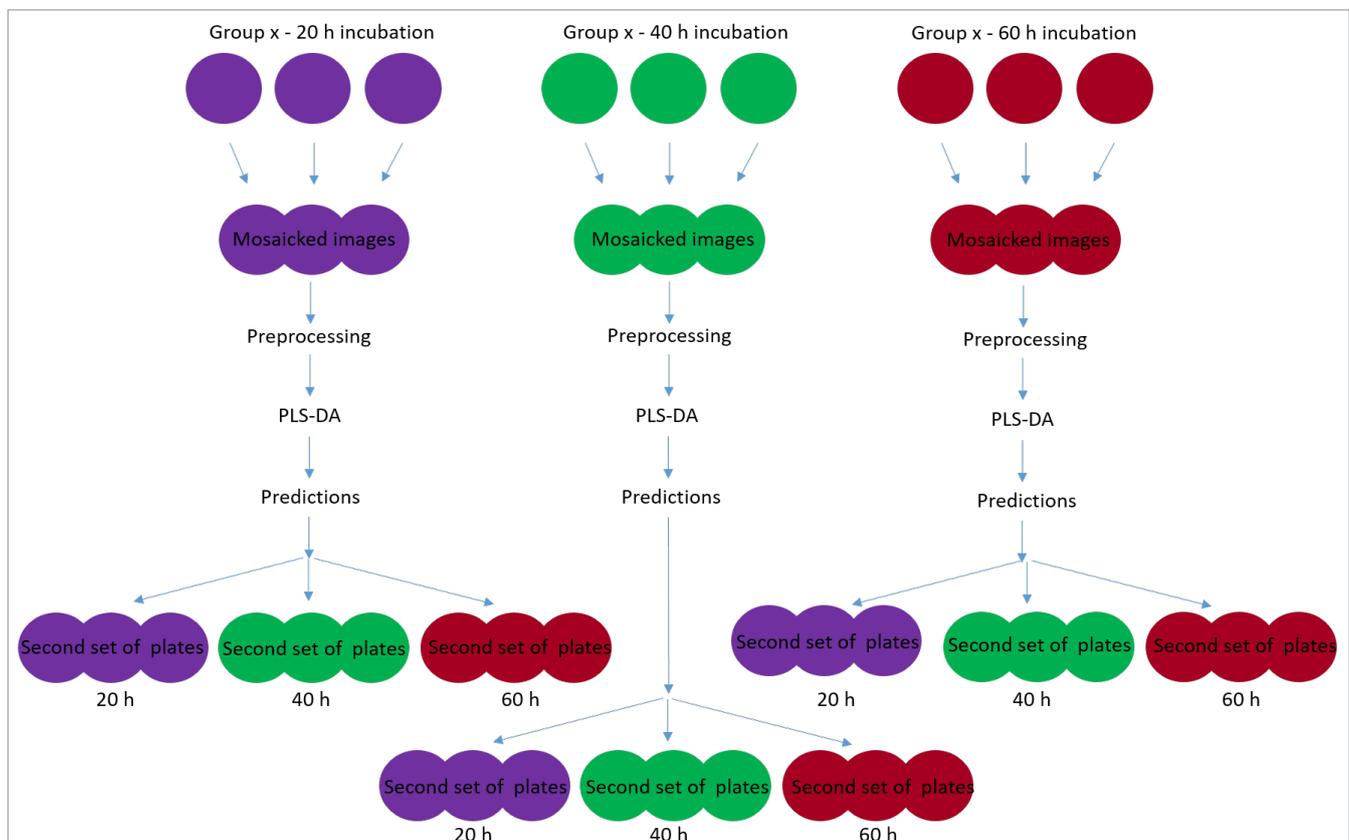


Figure 1. Schematic representation of data analysis process. PLS-DA models built for each time increment (20h, 40h and 60h growth), tested on a second set of plates of bacteria imaged after 20, 40 and 60h growth.

bacteria are due to distinct variation in cell wall composition.<sup>17-19</sup> In an earlier study, Kammies *et al.*<sup>5</sup> reported that differences in protein and carbohydrate structures in Gram-positive and Gram-negative bacteria, combined with the absence of teichoic acid in the cell walls of Gram-negative bacteria, were the main factors contributing to the variation observed.

For group 2, SNV-transformed data again exhibited the best clustering, where the two pathogens (*B. cereus* and *S. aureus*) were separated from non-pathogenic *S. epidermidis* (Figure 5b). The score image of this pre-processing method (Figure 6b) also showed the most variation in score values between the three Gram-positive bacteria. Both pathogens appeared to

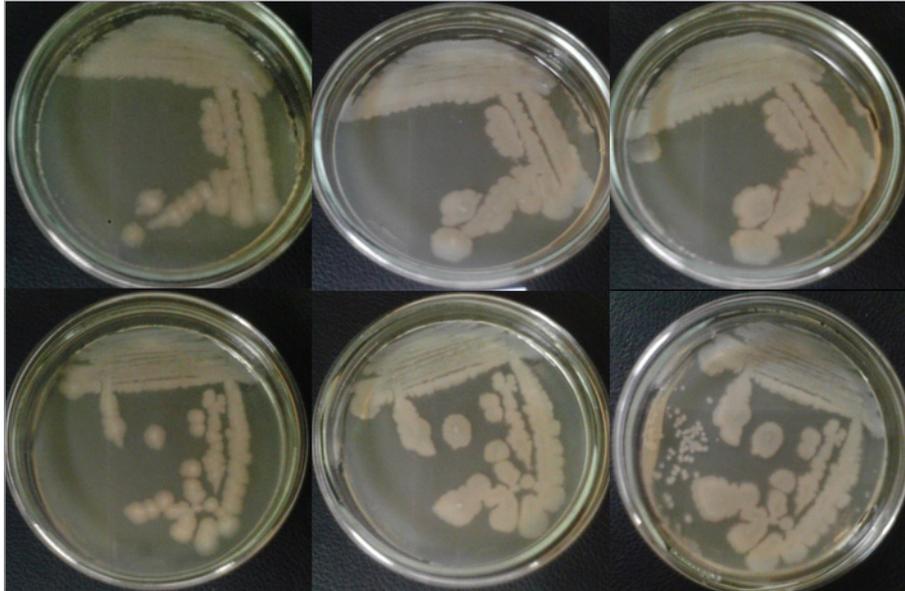


Figure 2. Digital images of *B. cereus* after 20h, 40h and 60h (left to right) growth at 37°C with the second set of plates in the bottom row.

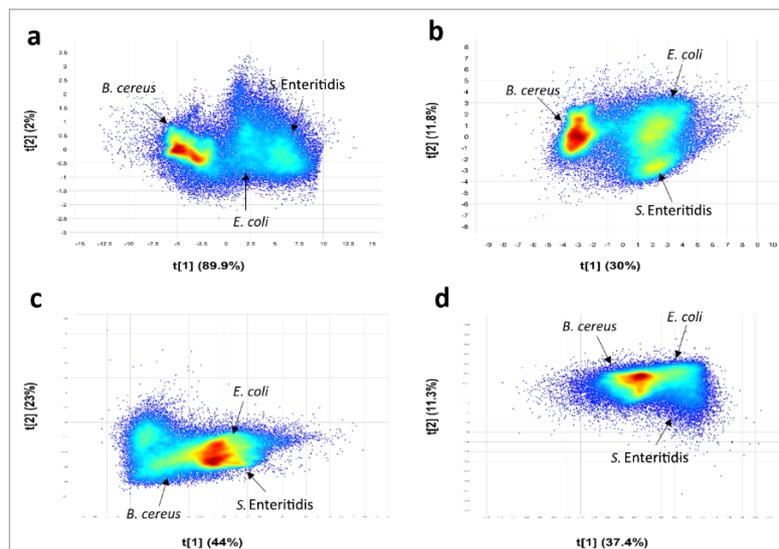


Figure 3. Score plots of PC1 vs PC2 for the mosaic of group 1 (*B. cereus*, *E. coli* and *S. Enteritidis*) with different pre-processing methods applied. a Raw data, b SNV, c Savitzky-Golay (1<sup>st</sup> derivative, 2<sup>nd</sup> order polynomial, 15-point smoothing) and d Savitzky-Golay (2<sup>nd</sup> derivative, 3<sup>rd</sup> order polynomial, 15-point smoothing).

have low score values, however, *B. cereus* was dark blue (high negative score values) while *S. aureus* was cyan (less negative score values). In contrast, *S. epidermidis* predominantly had positive score values. In the Savitzky–Golay pre-processed score images (Figure 6c

and d), the two *Staphylococcus* species appeared very similar. These results suggest that scattering effects caused by bacteria morphology, and the other sources mentioned previously, play a major role in bacteria differentiation.

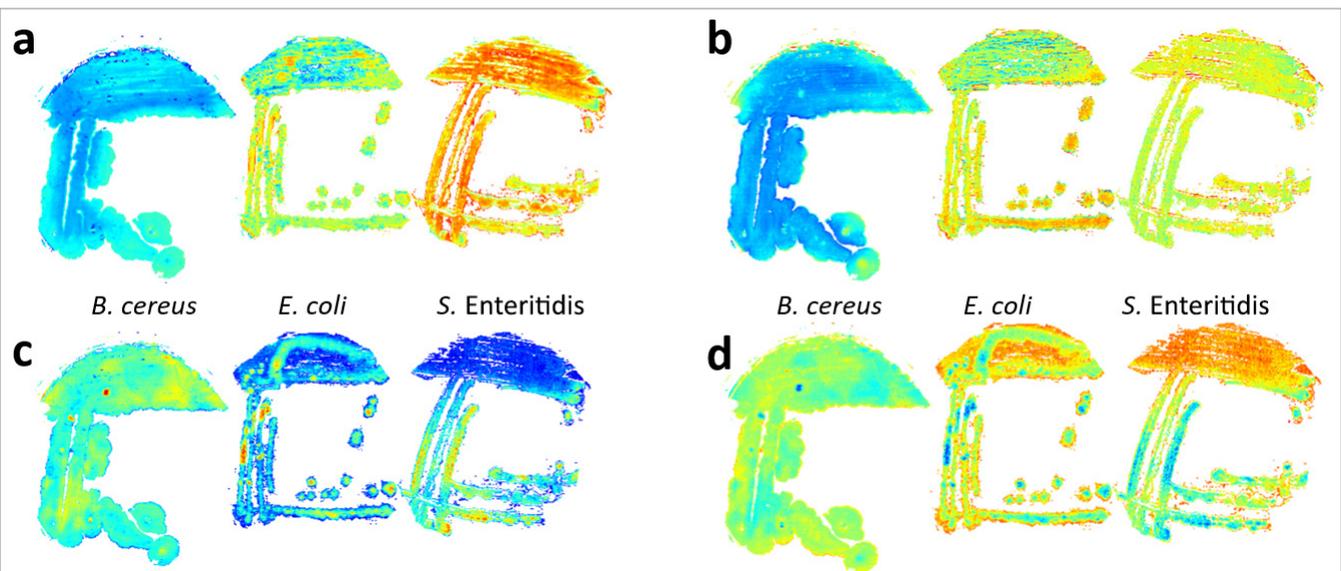


Figure 4. Score image of PC1 for the mosaic of group 1 (*B. cereus*, *E. coli* and *S. Enteritidis*) with different pre-processing methods applied. a Raw data, b SNV, c Savitzky–Golay (1<sup>st</sup> derivative, 2<sup>nd</sup> order polynomial, 15-point smoothing) and d Savitzky–Golay (2<sup>nd</sup> derivative, 3<sup>rd</sup> order polynomial, 15-point smoothing). Warm colours (yellow/red) are indicative of high score values and cold colours (blue/cyan) show lower score values.

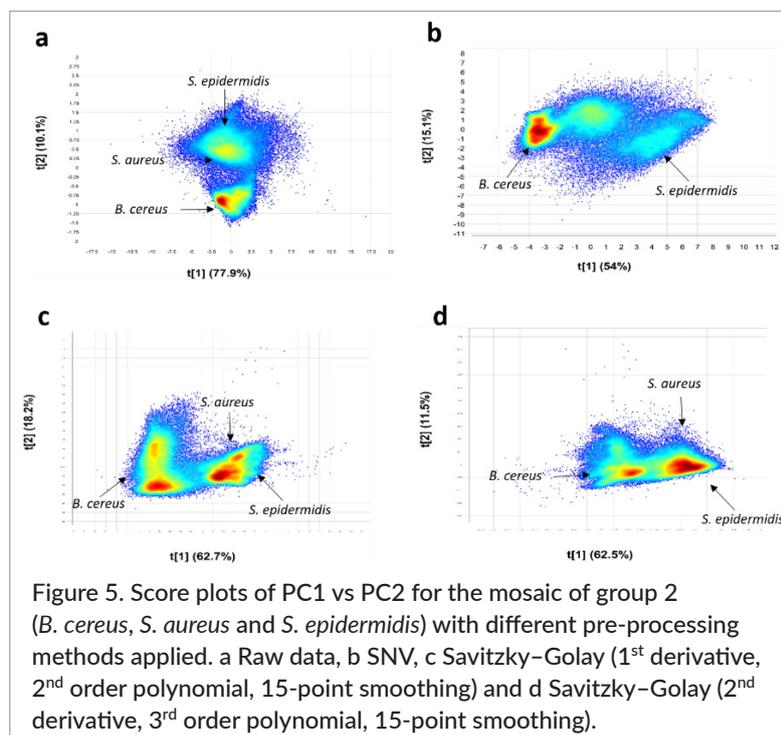


Figure 5. Score plots of PC1 vs PC2 for the mosaic of group 2 (*B. cereus*, *S. aureus* and *S. epidermidis*) with different pre-processing methods applied. a Raw data, b SNV, c Savitzky–Golay (1<sup>st</sup> derivative, 2<sup>nd</sup> order polynomial, 15-point smoothing) and d Savitzky–Golay (2<sup>nd</sup> derivative, 3<sup>rd</sup> order polynomial, 15-point smoothing).

## Partial least squares discriminant analysis

Results from our previous study<sup>5</sup> indicated that it was possible to distinguish between different bacteria after 20h growth, however, some classification results were not satisfactory. In order to improve classification accuracies while keeping rapidity in mind, we used older cultures (from advanced time points) to calibrate PLS-DA models and validated them on younger colonies (20h). For example, models were calibrated on images after 60h incubation and used to predict agar plates after 20h growth.

The performance of the 20h model for group 1 (Table 2) was good, except for the classification of *B. cereus* and *E. coli* tested on the second set of 20h growth plates, where only 69% and 71% of pixels were correctly classified. As streak plates were used in this study and pixel-wise classifications were done, each pixel was regarded as a colony. For this reason, these results were deemed unsatisfactory. In practice, if 30% of pathogenic colonies on a plate were not identified as such, it implies that the detection method is inadequate since a high percentage of possibly harmful contamination goes undetected. Therefore, in this study any classification accuracy under 90% correctly predicted pixels was considered low. Although a non-selective media was used that would not form part of a confirmative test, using a selective media would introduce variation that would hinder classification. Furthermore, as the proposed method aims to reduce time and costs, using a general-purpose media is appropriate.

The prediction accuracy of the model built from training data generated after 20h for group 1 was nevertheless 16% higher for *B. cereus* than the 40h model. However, for *E. coli*, this increased by 12.4%. Classification results from the 60h model showed an overall decrease in accuracy, except for *B. cereus* and *E. coli* after 20h growth, where it increased to 97% and 83%, respectively. The  $R^2$  for the 20h, 40h and 60h models were 0.45, 0.57 and 0.54, respectively. The lowest prediction accuracies for group 1 was for *B. cereus* with the 20h and 40h models, tested with the 20h growth plates (69% and 53% correctly predicted pixels, respectively). The SNV-corrected mean spectra, of each bacterium on the respective days, were studied to obtain information about chemical compounds within each bacterium to aid in the explanation of the classification results. At 1937nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H<sub>2</sub>O),<sup>20</sup> the spectrum of *B. cereus* after 20h growth (Figure 7) was closer to that of *E. coli* after 40h growth. Whereas, after 40h and 60h growth the peak intensities were nearly identical across the entire wavelength range. This explains the low prediction accuracy of the 20h *B. cereus* plate for the 40h model (model built from training data generated after 40h), where it might have been misclassified as *E. coli*. These differences are also likely due to changes in protein conformation or concentration,<sup>14</sup> since bacterial cultures change biochemically during aging.<sup>11,13</sup> Arnold *et al.*<sup>14</sup> reported changes in the peak intensities of matrix-assisted laser desorption/ionisation mass spectra of *E. coli*. They ascribed the

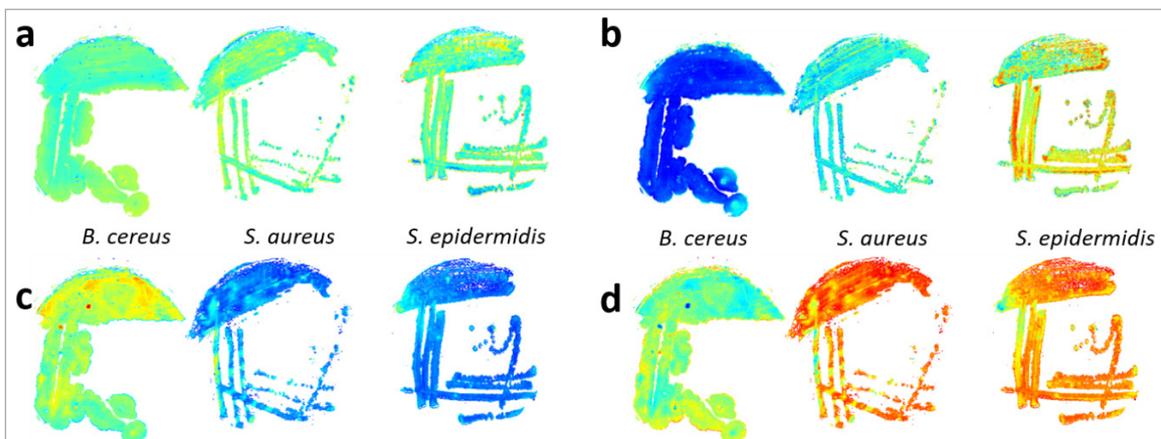


Figure 6. Score image of PC1 for the mosaic of group 2 (*B. cereus*, *S. aureus* and *S. epidermidis*) with different pre-processing methods applied. a Raw data, b SNV, c Savitzky-Golay (1<sup>st</sup> derivative, 2<sup>nd</sup> order polynomial, 15-point smoothing) and d Savitzky-Golay (2<sup>nd</sup> derivative, 3<sup>rd</sup> order polynomial, 15-point smoothing). Warm colours (yellow/red) are indicative of high score values and cold colours (blue/cyan) show lower score values.

Table 2. PLS-DA results of group 1 showing classification results of 20h, 40h and 60 h models as percentage pixels correctly predicted.

|  | <i>B. cereus</i> | <i>E. coli</i> | <i>S. Enteritidis</i> |
|--|------------------|----------------|-----------------------|
| % Correctly predicted pixels of validation data—model built from training data generated after 20h growth ( $R^2=0.45$ ) |                  |                |                       |
| 20h growth   | 69.1             | 70.6           | 90.3                  |
| 40h growth   | 99.8             | 85.9           | 96.0                  |
| 60h growth   | 96.0             | 82.4           | 87.9                  |
| % Correctly predicted pixels of validation data—model built from training data generated after 40h growth ( $R^2=0.57$ ) |                  |                |                       |
| 20h growth   | 53.1             | 83.0           | 81.6                  |
| 40h growth   | 95.8             | 86.5           | 95.2                  |
| 60h growth   | 92.3             | 93.1           | 89.8                  |
| % Correctly predicted pixels of validation data—model built from training data generated after 60h growth ( $R^2=0.54$ ) |                  |                |                       |
| 20h growth   | 96.9             | 83.1           | 86.4                  |
| 40h growth   | 98.8             | 75.3           | 95.5                  |
| 60h growth   | 93.9             | 78.2           | 84.4                  |

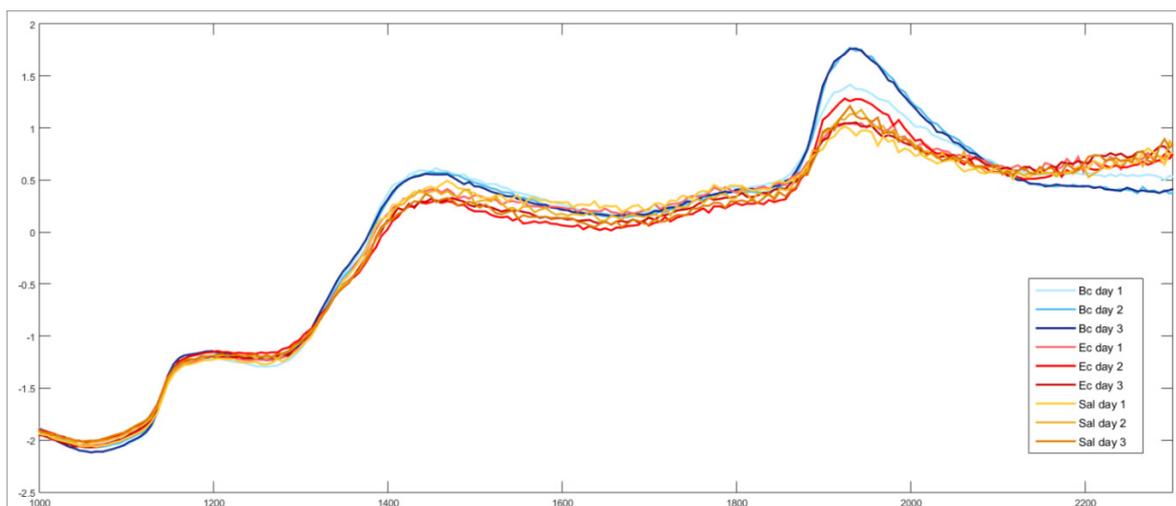


Figure 7. SNV-corrected mean spectra of group 1 [*B. cereus* (Bc), *E. coli* (Ec) and *S. Enteritidis* (Sal)] after 20h, 40h and 60h growth, with the major peak at 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H<sub>2</sub>O).

peaks to ribosomal protein and noted that they changed with increasing growth time, i.e. the peaks representing ribosomal protein decreased in intensity with increasing growth time. Although they could not quantify these changes, they concluded that differences in protein were responsible for the changes in peak intensity. In the early stages of growth, bacteria proliferate rapidly and produce ribosomes for protein synthesis. With time, this requirement is reduced as growth slows. Given that the organisms in this study all exhibit different growth rates and that the same bacteria would vary on the duplicate plate, it was expected that these changes would not be consistent. Thus, the haphazard results. The approximate generation time for bacteria used in the current study is as follows: *E. coli* and *S. Enteritidis*—20 min,<sup>21</sup> *S. aureus* and *B. cereus*—30 min<sup>22,23</sup> and *S. epidermidis*—45 min.<sup>24</sup>

There was an increase in prediction accuracy for *E. coli* with the 40h and 60h models (models built from training data generated after 40h and 60h, respectively), tested on the plates after 20h growth. However, for the 60h model, predictions of plates after 40h and 60h growth decreased. With the 20h model as the baseline, the only increase in correctly predicted pixels for *S. Enteritidis* was in the 40h model tested on the growth plate after 60h—an increase of 1.9% (87.9–89.8%) was observed. The best results for *B. cereus* and *S. Enteritidis* were achieved with the 20h model tested on growth plates after 40h, with 99.8% and 96.0% correctly predicted pixels, respectively. However, for *E. coli* the 60h model

tested on 60h growth plate was the best, with 93.1% correctly predicted pixels.

For group 2 (Table 3), the  $R^2$  for 20h, 40h and 60h models were 0.80, 0.75 and 0.69, respectively. An overall decrease in prediction accuracy was observed, except for *S. aureus*, where classification of pixels increased in consecutive models. The highest prediction accuracy for *B. cereus* was 99.9% with the 20h model, predicting the 40h growth plate. For both *S. aureus* and *S. epidermidis*, the 40h model showed the highest prediction accuracies tested on the 20h growth plates, with 97.3% and 99.3% correctly predicted pixels, respectively. In the 20h and 40h models, *S. aureus* had the lowest prediction accuracies, with only 24.6% for the 20h model tested on the 60h growth plate and 61.4% for the 40h model, predicting the 60h growth plate as well. These poor results are attributed to the slow initial growth of *S. aureus*.<sup>25</sup> This is seen in the SNV-corrected mean spectra (Figure 8), where the intensity of the peak at 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H<sub>2</sub>O)<sup>20</sup> for *S. aureus* is much lower after 20h growth, than 40h and 60h growth. This peak is likely related to protein or moisture, which are key components in all bacterial cell walls, as cells increase in population over time, protein and moisture content in colonies increase too, leading to a higher peak intensity. After 60h growth, the spectral profile of *S. aureus* is more similar to that of *B. cereus* after 40h and 60h growth, which could have led to misclassifications for this growth period.

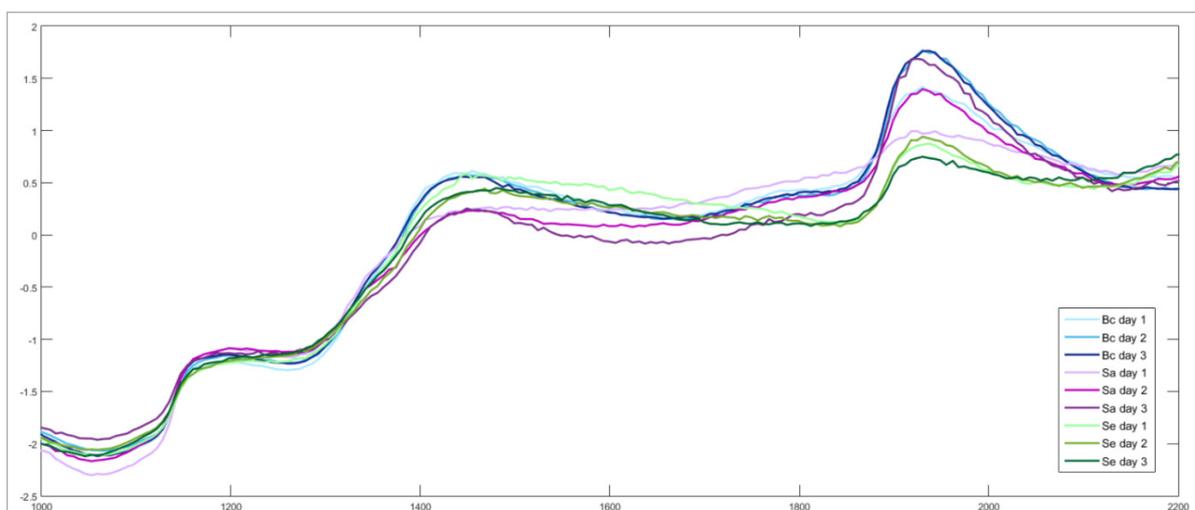


Figure 8. SNV-corrected mean spectra of group 2 [*B. cereus* (Bc), *S. aureus* (Sa) and *S. epidermidis* (Se)] after 20h, 40h and 60h growth, with the major peak at 1937 nm (C=O, second overtone, CONH/O-H stretch + O-H deformation, H<sub>2</sub>O).

**Table 3. PLS-DA results of group 2 showing classification results of 20h, 40h and 60h models as percentage pixels correctly predicted.**

|   | <i>B. cereus</i> | <i>S. aureus</i> | <i>S. epidermidis</i> |
|---|------------------|------------------|-----------------------|
| % Correctly predicted pixels of validation data—models built from training data generated after 20h growth ( $R^2 = 0.80$ ) |                  |                  |                       |
| 20h growth  | 77.6             | 96.7             | 98.2                  |
| 40h growth  | 99.9             | 71.8             | 99.2                  |
| 60h growth  | 97.0             | 24.6             | 98.4                  |
| % Correctly predicted pixels of validation data—models built from training data generated after 40h growth ( $R^2 = 0.75$ ) |                  |                  |                       |
| 20h growth  | 47.4             | 97.3             | 99.3                  |
| 40h growth  | 98.4             | 95.0             | 95.2                  |
| 60h growth  | 89.6             | 61.4             | 95.1                  |
| % Correctly predicted pixels of validation data—models built from training data generated after 60h growth ( $R^2 = 0.69$ ) |                  |                  |                       |
| 20h growth  | 96.9             | 89.1             | 97.6                  |
| 40h growth  | 97.5             | 89.1             | 97.4                  |
| 60h growth  | 93.9             | 89.1             | 97.3                  |

For group 3 (Table 4), the  $R^2$  for the 20h, 40h and 60h models were 0.80, 0.83 and 0.83, respectively. Model performances for this group were the best, with no prediction accuracy under 90%. The highest prediction accuracies for *S. aureus* was seen in the 20h model, predicting the 20h growth plate, with 99.8% correctly predicted pixels. For *S. epidermidis*, the best results were achieved with the 60h model, tested on the 20h growth plate, where 98.8% of pixels were correctly predicted. These two species are closely related,<sup>26</sup> thus these results are encouraging for applications of this technique for differentiation of pathogenic and non-pathogenic species.

## Conclusions

As an exploratory study, this work illustrates that the age of the colonies used to build models for pathogen differentiation proved to play a role in prediction accuracies. Of the three pre-processing methods investigated, SNV produced the best separation and clustering, and was, therefore, used as the only pre-processing method

throughout data classification. Classification accuracies for most models were encouraging, regardless of the day of prediction (age of the colony). However, deviations due to different generation times could have caused misclassification of bacteria.

Granted that the most rapid classifications were desired, it was found that models built from training data generated after 60h provided the overall highest prediction accuracies in each group. For this reason, we propose this period of growth (60h) for model calibration should this method be used for differentiation of these bacteria. However, additional research is required. Future research should include different classification algorithms, such as support vector machines, however, this method requires a large number of samples (thousands) and was not included in this study. Additionally, object-wise classification of bacterial colonies using different plating methods (such as spot plates) should also be investigated as this could also improve results. In practice, spread plates would be the selected plating method used in this detection technique, since this would provide single colonies on the surface of the solid agar. Object-wise classifica-

**Table 4. PLS-DA results of group 3 showing classification results of 20h, 40h and 60h models as percentage pixels correctly predicted.**

|   | <i>S. aureus</i> | <i>S. epidermidis</i> |
|---|------------------|-----------------------|
| % Correctly predicted pixels of validation data—models built from training data generated after 20h growth ( $R^2=0.80$ ) |                  |                       |
| 20h growth  | 99.8             | 97.4                  |
| 40h growth  | 99.6             | 98.2                  |
| 60h growth  | 98.3             | 90.6                  |
| % Correctly predicted pixels of validation data—models built training from data generated after 40h growth ( $R^2=0.83$ ) |                  |                       |
| 20h growth  | 99.3             | 97.0                  |
| 40h growth  | 99.2             | 97.4                  |
| 60h growth  | 96.0             | 97.3                  |
| % Correctly predicted pixels of validation data—models built from training data generated after 60h growth ( $R^2=0.83$ ) |                  |                       |
| 20h growth  | 98.4             | 98.8                  |
| 40h growth  | 98.4             | 94.7                  |
| 60h growth  | 97.8             | 92.8                  |

tion models would allow entire colonies to be recognised as one bacterium and not mixtures. For example, if more than 50% of pixels in a colony were recognised as *B. cereus*, the entire colony would be classed as such, allowing for a more definitive result. This could greatly increase the accuracy of this method. Even though we employed the pixel-wise approach, our results concur with that found in the literature, i.e. culture age influences detection.

## Acknowledgements

This work is based on the research supported in part by the National Research Foundation of South Africa for the grant, Unique Grant No. 94031. The authors wish to thank Professor Alvaro Viljoen, Dr Ilze Vermaak

and Carmen Leonard, Tshwane University of Technology, Pretoria for use of the NIR hyperspectral imaging system and microbiology laboratory.

## References

1. X. Zhao, C.-W. Lin, J. Wang and D.H. Oh, "Advances in rapid detection methods for foodborne pathogens", *J. Microbiol. Biotechnol.* **24**, 297–312 (2014). <https://doi.org/10.4014/jmb.1310.10013>
2. J. Qin, K. Chao, M.S. Kim, R. Lu and T.F. Burks, "Hyperspectral and multispectral imaging for evaluating food safety and quality", *J. Food Eng.* **118**, 157–171 (2013). <https://doi.org/10.1016/j.jfoodeng.2013.04.001>
3. J. Dubois, E.N. Lewis, F.S. Fry and E.M. Calvey, "Bacterial identification by near-infrared chemical imaging of food-specific cards", *Food Microbiol.* **22**, 577–583 (2005). <https://doi.org/10.1016/j.fm.2005.01.001>
4. S.C. Yoon, K.C. Lawrence, J.E. Line, G.R. Siragusa, P.W. Feldner, B. Park and W.R. Windham, "Detection of *Campylobacter* colonies using hyperspectral imaging" *Sens. Instrum. Food Qual. Saf.* **4**, 35–49 (2010). <https://doi.org/10.1007/s11694-010-9094-0>
5. T.-L. Kammies, M. Manley, P.A. Gouws and P.J. Williams, "Differentiation of foodborne bacteria using NIR hyperspectral imaging and multivariate data analysis", *Appl. Microbiol. Biotechnol.* **100**, 9305–9320 (2016). <https://doi.org/10.1007/s00253-016-7801-4>
6. S.-C. Yoon, W.R. Windham, S.R. Ladely, J.W. Heitschmidt, K.C. Lawrence, B. Park, N. Narang and W.C. Cray, "Hyperspectral imaging for differentiating colonies of non-O157 Shiga-toxin producing *Escherichia coli* (STEC) serogroups on spread plates of pure cultures", *J. Near Infrared Spectrosc.* **21**, 81–95 (2013). <https://doi.org/10.1255/jnirs.1043>
7. D.F. Barbin, G. ElMasry, D.W. Sun, P. Allen and N. Morsy, "Non-destructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging", *Innov. Food Sci. Emerg. Technol.* **17**, 180–191 (2013). <https://doi.org/10.1016/j.ifset.2012.11.001>
8. S. Yoon, K. Lawrence, G. Siragusa, J. Line, B. Park and P. Feldner, "Hyperspectral reflectance imaging for detecting a foodborne pathogen: *Campylobacter*",

- Trans ASABE* **52**, 651–662 (2009). <https://doi.org/10.13031/2013.26814>
9. S. Srivastava and P.S. Srivastava, "Bacteria and life processes - I Growth and multiplication", in *Understanding Bacteria*, 1<sup>st</sup> Edn, Ed by S. Srivastava and P.S. Srivastava. Springer, Netherlands, pp. 97–150 (2013). [https://doi.org/10.1007/978-94-017-0129-7\\_5](https://doi.org/10.1007/978-94-017-0129-7_5)
  10. R.E. Buchanan, "Life phases in a bacterial culture", in *Microbial Growth*, Ed by P.S.S. Dawson. Dowden, Hutchinson & Ross Inc., Stroudsburg, PA, pp. 25–41 (1974).
  11. S.J. Pirt, "A kinetic study of the mode of growth of surface colonies of bacteria and fungi", *J. Gen. Microbiol.* **47**, 181–197 (1967). <https://doi.org/10.1099/00221287-47-2-181>
  12. J.W.T. Wimpenny, "The growth and form of bacterial colonies", *Microbiology* **114**, 483–486 (1979).
  13. H.L.C. Coutinho, B.A. Handley, H.E. Kay, L. Stevenson and J.E. Beringer, "The effect of colony age on PCR fingerprinting", *Lett. Appl. Microbiol.* **17**, 282–284 (1993). <https://doi.org/10.1111/j.1472-765X.1993.tb01467.x>
  14. R.J. Arnold, J.A. Karty, A.D. Ellington and J.P. Reilly, "Monitoring the growth of a bacteria culture by MALDI-MS of whole cells", *Anal. Chem.* **71**, 1990–1996 (1999). <https://doi.org/10.1021/ac981196c>
  15. G. Bertani, "Studies on lysogenesis I. The mode of phage liberation by lysogenic *Escherichia coli*", *J. Bacteriol.* **62**, 293 (1951).
  16. K. Esbensen and P. Geladi, "Strategy of multivariate image analysis (MIA)", *Chemometr. Intell. Lab. Syst.* **7**, 67–86 (1989). [https://doi.org/10.1016/0169-7439\(89\)80112-1](https://doi.org/10.1016/0169-7439(89)80112-1)
  17. M. Salton, "Studies of the bacterial cell wall: IV. The composition of the cell walls of some Gram-positive and Gram-negative bacteria", *Biochim. Biophys. Acta* **10**, 512–523 (1953). [https://doi.org/10.1016/0006-3002\(53\)90296-0](https://doi.org/10.1016/0006-3002(53)90296-0)
  18. W.W. Navarre and O. Schneewind, "Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope", *Microbiol. Mol. Biol. Rev.* **63**, 174–229 (1999).
  19. T.J. Beveridge, "Structures of gram-negative cell walls and their derived membrane vesicles", *J. Bacteriol.* **181**, 4725–4733 (1999).
  20. B.G. Osborne, T. Fearn and P.H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific and Technical (1993).
  21. K. Fehlhaber and G. Krüger, "The study of *Salmonella* Enteritidis growth kinetics using rapid automated bacterial impedance technique", *J. Appl. Microbiol.* **84**, 945–949 (1998). <https://doi.org/10.1046/j.1365-2672.1998.00410.x>
  22. I. Keren, N. Kaldalu, A. Spoering, Y. Wang and K. Lewis, "Persister cells and tolerance to antimicrobials", *FEMS Microbiol. Lett.* **230**, 13–18 (2004). [https://doi.org/10.1016/S0378-1097\(03\)00856-5](https://doi.org/10.1016/S0378-1097(03)00856-5)
  23. J. Collins and M. Richmond, "Rate of growth of *Bacillus cereus* between divisions", *Microbiology* **28**, 15–33 (1962).
  24. B. Gottenbos, D.W. Grijpma, H.C. van der Mei, J. Feijen and H.J. Busscher, "Antimicrobial effects of positively charged surfaces on adhering Gram-positive and Gram-negative bacteria", *J. Antimicrob. Chemother.* **48**, 7–13 (2001). <https://doi.org/10.1093/jac/48.1.7>
  25. J. Sutherland, A. Bayliss and T. Roberts, "Predictive modelling of growth of *Staphylococcus aureus*: the effects of temperature, pH and sodium chloride", *Int. J. Food Microbiol.* **21**, 217–236 (1994). [https://doi.org/10.1016/0168-1605\(94\)90029-9](https://doi.org/10.1016/0168-1605(94)90029-9)
  26. A.F. Gillaspay and J.J. Landolo, "Staphylococcus", in *Encyclopedia of Microbiology*, Ed by M. Schaechter. Elsevier Science, San Diego, USA, pp. 293–303 (2009). <https://doi.org/10.1016/B978-012373944-5.00237-6>