Automatic image analysis and spot classification for detection of pathogenic *Escherichia coli* on glass slide DNA microarrays

Ron P. Haff\(^a\), Beatriz Quiñones\(^b\), Michelle S. Swimley\(^b\), Natsuko Toyofuku\(^a,\)#

\(^a\) Plant Mycotoxin Research Unit, U.S. Department of Agriculture-Agricultural Research Service, Western Regional Research Center, Albany, CA 94710, USA
\(^b\) Produce Safety and Microbiology Research Unit, U.S. Department of Agriculture-Agricultural Research Service, Western Regional Research Center, Albany, CA 94710, USA

**A R T I C L E   I N F O**

Article history:
Received 26 September 2009
Received in revised form 13 January 2010
Accepted 25 January 2010

Keywords:
DNA microarray
Algorithm
Spot location
*Escherichia coli*
Food safety

**A B S T R A C T**

A computer algorithm was created to analyze and quantify scanned images from DNA microarray slides developed for detecting pathogenic *Escherichia coli* isolates recovered from agricultural food products. The algorithm computed centroid locations for signal and background pixel intensities in RGB space and defined a plane perpendicular to the line connecting the centroids as a decision boundary. The algorithm was tested on 1534 potential spot locations which were visually classified depending on the strength of the signal. Three other standard measures of SNR (SSR, SBR, and SSDR) were also performed for each potential spot location. The number of errors as compared to visual classifications was computed for each of the four measures. SSR and SSDR, which depend on pixel intensity standard deviations, performed poorly with high false positive results, while the current algorithm and SBR, which were independent of standard deviations, performed much better. Overall error rates were 1.4% for the reported algorithm, 2.0% for SBR, 14.2% for SSDR, and 16.8% for SSR.

**Abstract**

Bacterial contamination of agricultural products continues to be a serious health threat, and recent increases in the reported occurrence of outbreaks have led to an increased emphasis on the development of food safety programs in the United States. A leading cause of foodborne illness is considered to be *Escherichia coli* (*E.* coli), which is thought to contribute to more than 73,000 cases of human infection in the United States per year (Rangel et al., 2005). Over the past years, there has been a rise in *E.* coli outbreaks due to the consumption of leafy vegetables, and many of these *E.* coli outbreaks were traced to the Salinas Valley region of California (Centers for Disease Control and Prevention, 2006; Cooley et al., 2007). In September 2006, a multi-state outbreak of *E.* coli infections was linked to baby spinach grown in San Benito County near the Salinas Valley region in California and resulted in 205 confirmed illnesses and 3 deaths (Centers for Disease Control and Prevention, 2006).

The rise in outbreaks linked to the consumption of agricultural food products has heightened the importance of developing better methods to rapidly detect and characterize pathogenic *E.* coli strains. Established culturing methods are very labor-intensive and time-consuming and are limited in the number of samples to be analyzed (Bettelheim and Beutin, 2003). Current methods for automated detection such as sorting based on X-ray, visible light, or near infrared have not been optimized for identifying single bacterial cells. Thus, optimization of procedures for pathogen surveillance is needed with sufficient sensitivity, cost-effectiveness and suitability for routine testing. Recently, methods have been developed using glass slide DNA microarrays for the rapid and economical identification of pathogenic *E.* coli recovered from food products (Quiñones et al., 2009). This information can be vital in guiding subsequent contamination control procedures and preventing contaminated food products from reaching the consumer.

The original glass slide microarrays were produced in 1995 (Schena et al., 1995), and their use as a tool in genomic research has expanded enormously. The development of commercially available printing devices that can precisely situate printing pins over glass slides has accelerated the adoption of this technology. One of the greater challenges has been the extraction, storage, and analysis of the huge amount of generated data (Holloway et al., 2002; Heller, 2002). While printing devices can rapidly produce thousands of spots, extracting the desired data from the microarrays can be time-consuming, slowing research progress. Glass slide microarrays are scanned into image files, and the data analysis then becomes an image processing exercise. Typical software requires selecting and saving pixel value data from multiple regions of interest for each spot on the slide, which can number in the thousands. Consequently, there is a considerable demand for the development of algorithms that can standardize and simplify the extraction of data from the scanned images (Heller, 2002).
There are a number of different glass slide formats used in the field of genomics for which analysis algorithms have been developed (Heller, 2002; Bhandarkar et al., 2004). Most microarray images are generated by using precise robotic controls, resulting in a grid of spots that are scanned and saved into an image file for subsequent analysis. As part of this process, one important goal is to determine some measure of the signal-to-noise ratio (SNR) between each spot in the array and the background. However, there is no consensus on how SNR should be determined, particularly in terms of determining background levels that are measured locally in the neighborhood of each spot or globally at a point outside of the grid. Some arguments have been made that only the local background estimate is adequate (Angulo and Serra, 2002). There has also been disagreement over the formula for computing SNR given the background and signal pixel intensities (Holloway et al., 2002; He and Zhou, 2008). Two common formulas for computing the SNR in the neighborhood of a spot are the signal to standard deviation (σ) ratio (SSR) given by

$$SSR = \frac{S - BG}{\sigma_{BG}},$$

(1)

and the signal to background ratio (SBR) given by

$$SBR = \frac{S}{BG},$$

(2)

where $\bar{S}$ is the mean intensity for signal pixels, $BG$ is the mean intensity for background pixels, and $\sigma_{BG}$ is the standard deviation of the background. Recently a new measure for SNR has been reported that addresses the problem of standard definitions not taking into account the non-uniformity of the intensities of signal pixels (He and Zhou, 2008). The signal to both $\sigma$’s ratio (SSDR) has been defined as

$$SSDR = \frac{S - BG}{\sigma + \sigma_{BG}},$$

(3)

where $\sigma$ represents the $\sigma$ of signal pixel intensities. SNR is not always the method used for determining spot presence. A simple threshold on the pixels in the predominant color channel or a threshold on the grayscale conversion is sometimes used. A threshold on an intensity histogram has also been used to separate background pixels from signal pixels (Steinfath et al., 2001).

Most automated algorithms reported to date attempt to correlate the pixel intensity (or SNR) at the spot location to the concentration of the sample being measured, such as expression levels and DNA copy number in biological samples. Hypothetically, spot intensity can be correlated with the amount of probe at that location of the grid, and some algorithms have been developed to estimate sample concentrations from the arrays (Lopez et al., 2004). In some cases a simple binary decision is the objective of spot analysis; either there is a spot or there is not (Lazo et al., 2005).

Previously reported automated algorithms for analyzing DNA microarray images have been concerned with detecting and analyzing many thousands of spots in a single image. The majority of the developed algorithms follow the same basic blueprint, which involves determining the grid layout and orientation, the spacing between spots and hence the locations of individual spots, measuring signal pixel and background pixel intensities in the neighborhood of each spot, and performing the desired statistics, e.g. the SNR, for each spot (Jain et al., 2002). The grid layout is often determined by simply summing pixels both vertically and horizontally and looking for the peaks in the resulting arrays corresponding to the rows and columns of the grid (Rueda and Vidyadharan, 2006). Proper orientation (or lack thereof) of the grid is also a concern for algorithms and there have been a variety of methods, including Hough transforms (Audic and Zanetti, 1995) to detect and measure the rotation of the grid as compared to the edges of the image.

Recently, researchers have begun to develop techniques using photo-polymerization for rapid and economical identification of toxin producing bacteria. One such study has an objective for the detection of pathogenic E. coli on glass slide DNA microarrays (Quiñones et al., 2009). The system generates bitmap images of the scanned glass slides. For the rapid detection of E. coli, it is necessary to automatically analyze the images. In the present study, the main objective was to develop an automatic computer algorithm for the analysis of images of glass slide DNA microarrays generated by a novel detection system that allows for the rapid and economical detection of pathogenic E. coli (Quiñones et al., 2009). A second objective was to demonstrate a new technique, not related to SNR, for the determination of the presence or absence of spots from scanned microarray images.

2. Materials and methods

2.1. Image parameters

The initial images generated by the scanner were 9 MB bitmap in 24 bit color format with a width of 2048 pixels and a height of 1536 pixels (Fig. 1). The region of interest was a smaller rectangle of 664 × 656 pixels. Fig. 2a shows the layout of the grid, which consisted of an array of 8 columns and 9 rows of potential spot locations, where the presence of a spot indicated a positive result. The 500 μm pins used for printing yielded spots of 400–450 μm (37–42 pixels) in diameter. For this algorithm, the spot diameter was defined as 40 pixels. Center-to-center spacing was 700 μm (66 pixels). The first, fifth, and ninth rows were controls and were expected to always show a positive result. These control spots divided the 24 distinct targets (E. coli strains to be detected) into four groups, with each target spotted in duplicate (Fig. 2a). The location of the grid within the overall image was consistent to within plus or minus ten pixels, and any rotation of the grid as compared to the edges of the image was small, less than about three degrees.

2.2. Image processing

Red, green, and blue pixel intensity values were read into separate arrays of 664 columns by 656 rows covering the region of interest, while the rest of the image was discarded. The array was flipped bottom to top to compensate for the reverse order of data.

Fig. 1. Bitmap image generated by scanning microarray slides to rapidly detect and identify E. coli strains from agricultural products. The initial image generated by the scanner is a 9 MB bitmap in 24 bit color format with a width of 2048 pixels and a height of 1536 pixels. The region of interest is a much smaller rectangle of approximately 664 × 656 pixels.
storage in bmp format images. Finally, the image was inverted by replacing each pixel intensity value \( I \) by \( 255 - I \). This was necessary because the scanning of a glass slide yielded maximum values in the background, and some elements of the algorithm described here depended on low background values.

### 2.3. Grid location

The location of the top left corner of the grid was consistent to within 10 pixels in both the vertical and horizontal directions. To precisely locate the grid corner, a mask was created as an array of integers representing a circle inside a square. The diameter of the circle matched the spot diameter of 40 pixels. Array values inside the circle were initialized to a value of one while those outside were given a value of zero. The mask was convolved over the image as explained below, starting at a location so that the center of the mask circle was ten pixels above and to the left of the expected location of the center of the spot. The value of each pixel in the mask (one or zero) was multiplied by the corresponding pixel value (low for background pixels, high for spot pixels) in the image and the sum of the products computed. The mask was then moved one pixel over and the process repeated until the center of the mask circle was ten pixels to the right and below the expected spot center. For purposes of the convolution, image pixels were converted to grayscale (average of red, green, and blue). Fig. 3 illustrates the convolution for an ideal case where all spot pixels have a maximum value of 255, and all background pixels have a value of zero. When the center of the two circles coincided, the convolution had its maximum value. The computational requirements were minimized since the approximate location of the grid corner was known. This procedure of moving the mask over the image pixels could be used potentially to locate the start of the grid without information about its location by starting at a top corner of the image and continuing until the first maximum occurred. This process assumes that large areas of high intensity, such as the gasket around the circle, are dealt with independently.

### 2.4. Rotation correction

The printing head reliably generated a grid in a square, and spacing between spots was consistent at 66 pixels. However, placement of the slide in the scanner could result in a slight rotation of the grid that required correction. Grayscale pixel intensity values were summed horizontally into a one-dimensional array with peaks corresponding to the rows of spots. Grid rotation was quantified by measuring the width and height of the array peaks at the pixel locations of the three rows of control spots. With no rotation, the peak

Fig. 3. Corner locating convolution for an ideal case where all spot pixels have maximum value (255) and all background pixels are zero. The convolution has its maximum value when the center of the two circles coincides.
was calculated as:

\[ \text{spots was known to be 392 pixels so the magnitude of the rotation} \]

by the direction of the peak shift. The length of the row of control information of the rotation. The direction of the rotation was indicated increase in width by the same amount, allowing a second confirma-

tion. For a rotated grid, each of the three peaks was expected to substrates were smeared outside the normal spot boundary. For a rotated grid, each of the three peaks was expected to occurrence when the grid was generated on the glass slide and the substrate material was smeared outside the normal spot boundary. For a rotated grid, each of the three peaks was expected to increase in width by the same amount, allowing a second confirmation of the rotation. The direction of the rotation was indicated by the direction of the peak shift. The length of the row of control spots was known to be 392 pixels so the magnitude of the rotation was calculated as:

\[ \theta = \sin^{-1} \left( \frac{dx}{392} \right) \]  

(4)

The plot in Fig. 4 was generated first from an image containing the three rows of control spots with no rotation and then after performing a three degree clockwise rotation. Each of the three peaks shows the same widening of approximately 21 pixels, which is the expected amount based on Eq. (4). Once the amount of rotation is determined, a standard rotation transformation (Ballard and Brown, 1982; Fisher et al., 2000) was performed on the pixel arrays so that the grid was properly oriented.

2.5. Mapping the spot coordinates and determining the closest spot to each pixel

Starting with the coordinates of the corner spot, pixel locations of all other grid points were computed based on the known center-to-center spacing and dimensions of the grid. A radius of 30 pixels was assigned to define the area around each grid point over which the signal strength and background levels were measured. A mask was applied to assign a pixel intensity of zero to all pixels outside the circles surrounding each grid point. An option to generate a new image of the grayscale values of the masked array was included to allow a visual verification of the correct grid location and rotation (Fig. 5).

2.6. Determination of the decision plane and pixel classification

For each grid point, the pixel intensities were measured over an area defined as background (outside the defined spot radius) above, below, and on either side of the grid point, and the centroid (average intensity of the selected pixels in RGB space) was calculated. Pixels within one half of the defined spot radius distance from the center of the nearest control spot were similarly measured and their centroid computed. The factor of one half was introduced for small deviations away from locations in the expected spot center, reducing the likelihood of background pixels being included in the calculation of the centroid of signal pixels. This exercise generated two points for each potential spot in the three-dimensional RGB space that represented the average intensities of signal pixels and background pixels in the neighborhood of each spot location. Given these two points and a line between them, the equation of a plane perpendicular to the line was computed as:

\[ A(x - x_0) + B(y - y_0) + C(z - z_0) = 0, \]  

(5)

where \((x, y, z)\) was the current pixel and \((x_0, y_0, z_0)\) was the intersection of the line between the centroids and the plane (Fig. 6). If the resultant of Eq. (5) was greater than zero, then the pixel was on the signal side of the plane and was considered a signal pixel. Otherwise, the pixel was considered a background pixel. In either case, the magnitude of the resultant was a measure of the strength of the signal. The location of the plane along the line can be used as a sensitivity factor. If the plane was moved closer to the signal centroid, then more pixels would be counted as background, reducing the odds for false positive results. For this study, the plane was situated at the halfway point between the two centroids. There are simpler ways to differentiate between signal and background pixels. A simple threshold could be imposed on any of the three color components, or the combination of them (grayscale). Determining the decision boundary in three-dimensional color space resulted in the maximum use of the available information.

For each pixel on the grid that lay within the defined radius of thirty pixels from each potential spot location (Fig. 5), the power
Fig. 6. Known background and signal pixels in the neighborhood of the spot are selected and the centroid for each group computed. A plane perpendicular to the line connecting the two centroids serves as a decision boundary between background and signal pixels.

\[ P_s = \sum_{i=1}^{n} A_i (x_i - x_0) + B_i (y_i - y_0) + C_i (z_i - z_0), \]  
where the summation was over all pixels within the 30 pixel radius classified as signal pixels. \( A_i, B_i, \) and \( C_i \) were the calculated coefficients for the plane at \( (\text{potential}) \) spots. The power of the signal was subsequently used for spot classification.

2.7. Measuring SNR's

In addition to the signal power computations, the three SNR measures (Eqs.(1)–(3)) were computed for each potential spot location. Signal means and \( \sigma_s \) were measured over a radius of 15 pixels from the computed spot center. The full 20 pixel radius was not used in an attempt to avoid mistakenly including background pixels. Background pixel intensity mean and \( \sigma_b \) were also measured over a 10 pixel \( \times \) 50 pixel rectangular area directly below each spot location. This thin rectangular region was used as the distance between the spot edges was only 16 pixels (Fig. 2b).

2.8. Training and testing

Twenty-two images were generated as described above yielding 1584 potential spots. For each spot, SNR values for each of the four classification methods were calculated. Each potential spot location in each image was visually classified as: 1 (full spot), 2 (weak spot), 3 (barely discernible spot), or 0 (no spot) (Fig. 7). Within each classification, half of all potential spots were assigned to a training set and the other half were assigned to a validation set. The training sets were sorted by SNR results for each of the four classification methods. Thresholds were applied based on the criteria that all spots that had been visually classified as 1 or 2 would be correctly classified \( \text{(i.e. on the same side of the threshold).} \) These thresholds were then applied to the validation set. Results were tabulated as the number of false positives, false negatives, and the total number misclassified when the derived thresholds were applied to the validation set. A false positive result occurred when a particular spot was identified by the algorithm but was not identified upon visual inspection of the image. In contrast, a false negative result occurred when a full spot was not correctly detected.

3. Results and discussion

Of 1584 potential spot locations in the 22 images, 50 were visually eliminated because of impurities on the slides at the spot locations. Most of these impurities were caused by debris in the scanning apparatus (Fig. 1). The distribution of the remaining 1534 spot locations between training and validation sets and visual classifications are shown in Table 1. Table 2 shows the ranges of values and selected thresholds for the power, SSB, SSR, and SSDR.
calculations. Although typical threshold values for the SSB (3.0) and SBR (1.5–2.0) methods have been reported for analysis of DNA microarrays (He and Zhou, 2008), in reality these values are dependent on both the system generating the images as well as an appropriate “risk assessment” which determines the relative importance of false positives vs. false negatives. The lower threshold values reported here were due to the requirement that all spots, visually classified as 1 or 2, be correctly identified in the training set. Selecting criteria to minimize false positives would increase the values of the thresholds. Since the system used here attempted to detect pathogenic bacteria in food, a major food safety issue, false negative results were considered to have a higher risk factor than false positive results. Of course, excessive false positive results are also a concern regardless of the accuracy on the false negative side.

Pixel value ranges and means by spot classification are shown in Table 3, along with ranges and means for $\sigma$. The minimum intensity value for background pixels was 42, while the maximum signal intensity was 120. This represents an opportunity for improvement as slightly less than one-third of the dynamic range of the system was employed. However, using software to artificially stretch the dynamic range with a normalization process generates better images for visualization but does not enhance the ability to separate classes with a computer algorithm. The same holds true for background subtraction. As such, this opportunity for improvement exists in the image generating process and not in the spot recognition algorithm. Of more interest is the spread of $\sigma$, both in the background and at the spot locations. Mean background $\sigma$ for background pixels were 1.06 but ranged dramatically from near zero to as high as 20. Fig. 8 is a plot of background residuals for all 1534 spots analyzed with the order randomized along the x-axis. This plot illustrates the pitfalls that can occur when using $\sigma$ as part of the SNR equations, especially in the denominator as is the standard practice. Many of the $\sigma$s are very low, near zero, inflating the SNR value, while others are very high, presumably due to imperfections on the slides. These imperfections could be debris in the scanner, as discussed earlier, or spots in the background where the dye smeared. If the area scanned for a background $\sigma$ reading crosses the edge of one of these regions then the $\sigma$ would be high. This is not generally a problem when the images are being analyzed manually with a software program and humans selecting the background to be measured simply avoid regions that are atypical. For automated algorithms, however, it can lead to errors in classification. Another point of interest is the low $\sigma$ values found in potential spot locations (points on the $8 \times 9$ grid) where no spots occur. The mean $\sigma$ for these points is 0.79, compared with a significantly higher mean of 1.06 for the general background. This is believed to be an artifact of the oligonucleotides spotted on the slides at all grid points. Low $\sigma$ values at points where no spot occurs inflates the SSB and SSDR values, increasing the chances for false positive results.

Misclassification rates are shown in Table 4. Overall error rates were 1.4% for the power algorithm, 16.8% for SBR, 2.0% for SSR, and 14.2% for SSDR. Both SSB and SSDR use $\sigma$ measurements in the denominator and clearly suffered from high false positive rates. In contrast, the power algorithm and SBR, which do not depend on $\sigma$, had much lower misclassification error rates. The power algorithm performed best overall but only slightly better than SBR.

The power algorithm calculation of the plane intersecting the line between the background and signal centroids in the RGB space made use of color information, while the SNR calculations were based on grayscale pixel values. This approach was followed after manually performing the analysis on several images and comparing the results. No advantage was found between using the grayscale values versus measuring SNR by using pixel values of the strongest color channel. This is not a surprising result as the weaker channels are expected to contribute equally to signal and background intensities.

There is a potential to improve the power algorithm by using more sophisticated statistical techniques in determining the separation boundary between signal and background pixels. Rather than computing the plane perpendicular to the line, discriminant analysis could be used to determine the optimal plane for maximizing correct classification of pixel intensities. Other potential techniques would include a nearest neighbor approach or principle component analysis. The adoption of any of these techniques would increase the complexity and computational load of the algorithm. This is an area of ongoing research in this project.
References

Eun Narm and Joseph Gomez for excellent technical assistance.

Acknowledgements

This work was supported by the U.S. Department of Agriculture, Agricultural Research Service CRIS project numbers 5325-42000-045. Special thanks to Erica Dawson and Amber Taylor (InDevR, Inc., Boulder, CO) for providing reagents and instrumentation for generating the microarray images. The authors also wish to thank Robert E. Mandrell and Michael Cooley (USDA/ARS/WRRC, Albany, CA) for providing the bacterial strains used in this study, and Koh-Eun Narm and Joseph Gomez for excellent technical assistance.

4. Conclusions

A “power” algorithm was created to automatically analyze scanned images from DNA microarray slides developed to rapidly detect and identify pathogenic *E. coli*, linked to outbreaks due to contamination of leafy vegetables from agricultural regions in California. The algorithm was tested on a total of 1534 potential spot locations which were visually classified depending on the strength of the signal. Three other standard measures of SNR (SSR, SBR, and SSDR) were also performed for each potential spot location. The number of errors when comparing algorithm results with visual classifications was tabulated for each of the four detection methods. Both SSR and SSDR, which depend on pixel standard deviations, performed poorly with high false positive results. The power algorithm and SBR, which do not include standard deviations in their computations, performed much better. Overall error rates were 1.4% for the power algorithm, 2.0% for SBR, 14.2% for SSDR, and 16.8% for SSR. The conclusions from the present study indicate that SNR measurements with standard deviations appear to work well when the analyses are non-automated. Future work is aimed at optimizing the computational algorithm for circumventing problem areas in the background of the scanned images for automated applications.

Table 4

Error rates for spot detection.

<table>
<thead>
<tr>
<th>Dot intensity</th>
<th># of dots</th>
<th>Number and % misclassified</th>
<th>Power algorithm</th>
<th>%</th>
<th>SSR</th>
<th>%</th>
<th>SBR</th>
<th>%</th>
<th>SSDR</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>489</td>
<td>9</td>
<td>1.8</td>
<td>126</td>
<td>25.8</td>
<td>5</td>
<td>1.0</td>
<td>106</td>
<td>21.7</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>183</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td>0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>36</td>
<td>2</td>
<td>5.6</td>
<td>2</td>
<td>5.6</td>
<td>9</td>
<td>25.0</td>
<td>2</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>769</td>
<td>11</td>
<td>1.4</td>
<td>129</td>
<td>16.8</td>
<td>15</td>
<td>2.0</td>
<td>109</td>
<td>14.2</td>
<td></td>
</tr>
</tbody>
</table>

4. Conclusions

A “power” algorithm was created to automatically analyze scanned images from DNA microarray slides developed to rapidly detect and identify pathogenic *E. coli*, linked to outbreaks due to contamination of leafy vegetables from agricultural regions in California. The algorithm was tested on a total of 1534 potential spot locations which were visually classified depending on the strength of the signal. Three other standard measures of SNR (SSR, SBR, and SSDR) were also performed for each potential spot location. The number of errors when comparing algorithm results with visual classifications was tabulated for each of the four detection methods. Both SSR and SSDR, which depend on pixel standard deviations, performed poorly with high false positive results. The power algorithm and SBR, which do not include standard deviations in their computations, performed much better. Overall error rates were 1.4% for the power algorithm, 2.0% for SBR, 14.2% for SSDR, and 16.8% for SSR. The conclusions from the present study indicate that SNR measurements with standard deviations appear to work well when the analyses are non-automated. Future work is aimed at optimizing the computational algorithm for circumventing problem areas in the background of the scanned images for automated applications.

Acknowledgements

This work was supported by the U.S. Department of Agriculture, Agricultural Research Service CRIS project numbers 5325-42000-045. Special thanks to Erica Dawson and Amber Taylor (InDevR, Inc., Boulder, CO) for providing reagents and instrumentation for generating the microarray images. The authors also wish to thank Robert E. Mandrell and Michael Cooley (USDA/ARS/WRRC, Albany, CA) for providing the bacterial strains used in this study, and Koh-Eun Narm and Joseph Gomez for excellent technical assistance.

References
