Applied and Environmental Microbiology, May 2007, p. 2956–2962 Vol. 73, No. 9
0099-2240/07/$08.00+0 doi:10.1128/AEM.02954-06

Automated Image Analysis for Quantitative Fluorescence In Situ Hybridization with Environmental Samples▽†

Zhi Zhou,[1] Marie Noëlle Pons,[2] Lutgarde Raskin,[1]‡ and Julie L. Zilles[1]*

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Avenue, Urbana, Illinois 61801,[1] and Laboratoire des Sciences du Génie Chimique, CNRS-ENSIC-INPL, 1 rue Grandville, BP 20451, F-54001 Nancy, France[2]

Received 20 December 2006/Accepted 26 February 2007

When fluorescence in situ hybridization (FISH) analyses are performed with complex environmental samples, difficulties related to the presence of microbial cell aggregates and nonuniform background fluorescence are often encountered. The objective of this study was to develop a robust and automated quantitative FISH method for complex environmental samples, such as manure and soil. The method and duration of sample dispersion were optimized to reduce the interference of cell aggregates. An automated image analysis program that detects cells from 4′,6′-diamidino-2-phenylindole (DAPI) micrographs and extracts the maximum and mean fluorescence intensities for each cell from corresponding FISH images was developed with the software Visilog. Intensity thresholds were not consistent even for duplicate analyses, so alternative ways of classifying signals were investigated. In the resulting method, the intensity data were divided into clusters using fuzzy *c*-means clustering, and the resulting clusters were classified as target (positive) or nontarget (negative). A manual quality control confirmed this classification. With this method, 50.4, 72.1, and 64.9% of the cells in two swine manure samples and one soil sample, respectively, were positive as determined with a 16S rRNA-targeted bacterial probe (S-D-Bact-0338-a-A-18). Manual counting resulted in corresponding values of 52.3, 70.6, and 61.5%, respectively. In two swine manure samples and one soil sample 21.6, 12.3, and 2.5% of the cells were positive with an archaeal probe (S-D-Arch-0915-a-A-20), respectively. Manual counting resulted in corresponding values of 22.4, 14.0, and 2.9%, respectively. This automated method should facilitate quantitative analysis of FISH images for a variety of complex environmental samples.

Fluorescence in situ hybridization (FISH) is a method that is used to detect specific RNA or DNA sequences in situ with fluorescently labeled oligonucleotide probes (4, 10). This technique is used widely in environmental microbiology and clinical diagnostics (6). Although FISH has many advantages, automated analysis of FISH images remains challenging. The intensities of positive signals may be different in different experiments, even for the same sample. The differences in intensity are due to a number of factors, including the metabolic state of the cells, the hybridization conditions, and the image acquisition parameters. Many types of environmental samples have additional complications due to the presence of cell aggregates and nonuniform background fluorescence.

The basic task in analysis of FISH images is to classify cells into two groups: target (positive) cells and nontarget (negative) cells. This classification is typically based on a threshold; i.e., all cells with fluorescence intensity higher than a certain threshold are considered target cells, and other cells are considered nontarget cells. The simplest approach for setting a threshold is to choose a fixed value above the background level (26). As noted above, fluorescence intensity varies between experiments, so the threshold is often set manually for each experiment. Another common approach is to set a fixed signal-to-noise ratio (14, 23, 25). Pernthaler et al. used a fixed signal-to-noise ratio and defined the threshold as the mean background gray value of a FISH gray image multiplied by a signal-to-noise factor (25). The factor was empirically determined based on manual counting and varied from 110 to 200%, suggesting that it may require adjustment for each experiment. Using an alternative approach, Langendijk et al., set the threshold at the 95th percentile of the fluorescence intensity distribution of the negative control (21). However, as these authors noted, such a threshold is also not optimal, since the overlap of two fluorescence distributions in combination with a high fraction of potential target cells resulted in underestimation of the hybridization percentage. Furthermore, this method relies on comparable absolute intensities from two hybridizations (negative control and experiment), but as noted above, there is often substantial variability in intensities even in duplicate experiments. None of these methods is well suited for analysis of samples containing mixtures of cell aggregates and individual cells or samples with variable backgrounds.

Because the intensities of background, target cell, and nontarget cell signals vary within an image and among experiments, threshold-based classification methods appeared to have intrinsic problems. We turned to the idea of cluster analysis, i.e., distinguishing groups within a data set, as a way of handling this variability. To classify complex biological signals, several cluster analysis methods have been developed, including model-based clustering (13), neural network classification

* Corresponding author. Mailing address: 3204 Newmark Civil Engineering Laboratory, MC250, 205 North Mathews Avenue, Urbana, IL 61801. Phone: (217) 244-2925. Fax: (217) 333-6968. E-mail: jzilles@uiuc.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

‡ Present address: Department of Civil and Environmental Engineering, University of Michigan, 1351 Beal Ave., Ann Arbor, MI 48109.

▽ Published ahead of print on 9 March 2007.

(5), and *c*-means clustering (12, 15). Mixed model-based clustering with a Bayesian classifier has been used successfully for detecting chromosomal abnormalities in nuclei or for detecting breast cancer (13, 22); however, it has not been used with microorganisms or with environmental samples. Neural network classification has been used to determine bacterial abundance and morphology in natural aquatic communities (5). Both of these techniques require extensive experience in computer science, and for the neural network classification approach a specific data set is required for training. *c*-means clustering (also known as *k*-means clustering) is a method for arranging data points into a specified number of groups, or clusters, such that the data points within a group are more similar than the data points for different groups. Fuzzy *c*-means clustering (FCM) is a modification of *c*-means clustering in which a data point can belong to more than one cluster (for example, a data point can be classified as 60% cluster 1 and 40% cluster 2), and this additional flexibility is helpful when complex environmental samples are examined. FCM clustering has been used to analyze DNA microarray data (12, 30) and to recognize coregulation of gene expression (15), but prior to this study it had not been applied to analysis of FISH images.

In this study, an automated method for analyzing FISH images with FCM was developed using swine manure and soil as representative complex environmental samples. The automated, quantitative image analysis method was validated by comparing results obtained with this method with manual counts.

MATERIALS AND METHODS

Sample collection, fixation, and dispersion. The swine manure samples used in this experiment were obtained from commercial production facilities in Illinois (18). Two-hundred-gram manure samples from a natural farm (designated NF) were collected from the floor of a building that housed pigs, which was covered with straw and soil, and homogenized. At a conventional farm (designated CF), 200 g of manure was collected from pits underneath the slotted floor of a building that housed pigs and was homogenized. Samples from these pits contained a mixture of dilution water, swine waste, straw, and dust. Five kilograms of topsoil was collected from the surface of a cornfield at an Illinois research farm. Stones were removed, and soil samples were homogenized with a blender.

Samples were fixed using paraformaldehyde and ethanol separately as described previously (11). Briefly, samples were incubated in 4% paraformaldehyde or 100% ethanol (AAPer Alcohol & Chemical, Shelbyville, KY) for 2 h on ice. The paraformaldehyde-fixed samples were washed three times in phosphate-buffered saline (PBS) (130 mM NaCl, 10 mM sodium phosphate; pH 7.2). Following fixation, the samples were resuspended in PBS-ethanol (1:1, vol/vol) and stored at −20°C. Subsequently, the fixed swine manure samples were diluted (1:10) in 1× PBS, and the fixed soil sample was diluted (1:100) in 0.1% sodium pyrophosphate (NaPPi) buffer (20). Then 10 μl of a diluted sample was sonicated in 2 ml of 1× PBS (swine manure) or 0.1% NaPPi buffer (soil) with a sonic dismembrator (5-s pulse; output, 250 W; variable duration; model 500; Fisher Scientific, Pittsburgh, PA) and filtered through a black 0.22-μm-pore-size polycarbonate membrane (diameter, 25 mm; Osmonics, Minnetonka, MN). Cells were transferred from the filters to gelatin-coated slides by manually pressing the filters onto the slides for at least 10 s. The gelatin-coated slides were prepared by using the protocol of Amann et al. (2). The efficiency of transfer from the polycarbonate membrane to the microscope slide was evaluated by comparison of the numbers of cells on membranes with and without transfer. The background contamination was quantified by filtering the same volume of 1× PBS or NaPPi buffer for corresponding samples, and the values obtained were subtracted from the cell counts.

Microscopic analysis. For FISH experiments, two 16S rRNA-targeted oligonucleotide probes were used: the general bacterial probe Bact0338 (S-D-Bact-0338-a-18; 6-carboxyfluorescein [6-FAM] labeled; W. M. Keck Center for Comparative and Functional Genomics, Urbana, IL) (1) and the general archaeal probe Arch0915 (S-D-Arch-0915-a-A-20; Alexa488 labeled; Invitrogen Corporation, Carlsbad, CA) (28). FISH was performed as previously described (11), except that 15 μl hybridization solution was mixed with 1 μl (50 ng/μl) probe, and the hybridization and wash temperatures were 46 and 48°C, respectively. Following FISH, slides were incubated in a 1-μg/ml 4′,6′-diamidino-2-phenylindole (DAPI) (Sigma Chemical, St. Louis, MO) solution for 5 min, rinsed twice with 1 ml distilled deionized water, air dried, and mounted with Citifluor (Marivac Limited, Halifax, NS, Canada). Negative control experiments without probes were performed for all samples. Slides were observed using a magnification ×630 with a Zeiss Axioskop 40 microscope equipped for both light and fluorescence microscopy (Carl Zeiss, Oberkochen, Germany) using a green filter set (excitation wavelength, 480 nm; emission wavelength, 535 nm; model 41001; Chroma Technology Corp., Rockingham, VT) for cells hybridized with FAM- and Alexa488-labeled probes and a near-UV filter set (excitation wavelength, 350 nm; emission wavelength, 460 nm; model 31000; Chroma Technology Corp., Rockingham, VT) for DAPI-stained cells. Images were acquired from 15 random locations using a monochrome camera (AxioCam MRm; Carl Zeiss MicroImaging, Inc. Thornwood, NY) and exposure times of 5, 2, and 0.5 s for FAM-labeled, Alexa488-labeled, and DAPI-labeled images, respectively. The resulting gray images were saved in eight-bit tagged image file (tif) format. In this format the intensity is represented by an integer value between 0 (black) and 255 (white). The image size was 1,388 by 1,040 pixels (1 pixel = 0.1027 μm). All experiments were carried out in duplicate.

Automated image processing. The image-processing protocol was designed to detect individual objects in DAPI-labeled images and to recover two types of data from the corresponding FISH images, data describing the individual objects and data describing the background.

(i) Recognition of cells and background. First, DAPI-labeled images were locally equalized to obtain more uniform distributions and intensities and better contrast. The size of the region to be equalized was optimized based on a comparison with actual cell sizes, which resulted in selection of a 4- by 4-pixel window. Following local equalization, individual cells were detected through binarization, using a threshold set automatically with an entropy algorithm integrated in Visilog (version 6; Noésis, Les Ulis, France) (19). Areas with average intensities above this threshold were set to 1, and all other areas were set to 0. To separate overlapping cells, a fast watershed algorithm was used (29). The resulting images were cleaned with a median filter to remove noise. The median filter replaced every pixel with the average value (0 or 1) of the encompassed 3- by 3-pixel square. The resulting images were used as DAPI masks to determine the discrete locations of cells within the FISH image. Statistical differences between counts obtained with DAPI masks and counts obtained by manual counting were analyzed by a paired *t* test with the R statistical program (version 2.3.1; http://www.r-project.org/). To generate a background mask, DAPI masks were first inverted, so that background areas were analyzed rather than cells. The background area was then eroded (reduced in size) to exclude artificially high background measurements in the immediate vicinity of bright cells. The erosion process replaces every pixel with the minimum value (0 or 1) within the structuring element of 8 pixels. The resulting background masks were used to determine the average intensities of background areas in FISH images.

(ii) Feature extraction. The DAPI mask and the background mask described above were overlaid on the corresponding FISH image to detect the regions used for analysis. For each object detected in the DAPI mask, the area, perimeter, shape, length, width, location, maximum intensity, mean intensity, total intensity, and mean background were retrieved automatically. The entire area of the background mask was used to calculate the mean background intensity for a given image. For all analyses except those utilizing signal-to-noise ratios, the mean background intensity was subtracted from the maximum and mean intensities.

The entire image-processing procedure was programmed with VBA (Visual Basic for Applications) using the image analysis software Visilog. Parameters for each cell were exported to a data file with tab-delimited text format for analysis with standard spreadsheet software. This VBA program is available for download from the corresponding author's website (http://www.cee.uiuc.edu/people/jzilles/). Using this procedure, the average time for analyzing each image was less than 20 s with a 2.4-GHz Pentium 4 personal computer with 512 Mbytes of memory.

Classification through FCM. FCM analysis was performed for the maximum and mean intensities of the cells using FuzME, a Windows program (http://www.usyd.edu.au/su/agric/acpa). The number of clusters (*c*) was set at 10 in the final method, while the degree of fuzziness was always set at the default value of 1.3. During this analysis, principle-component analysis was used to maximize the spread of the intensity data. The resulting clusters were classified as positive or negative by comparison of the cluster centroids with the mean values for maximum intensity obtained from the negative control (no probe). This classification

was verified by manually checking the positive and negative clusters closest to the threshold. The positive percentage (percentage of target cells) was calculated based on the number of cells in positive clusters divided by the total number of cells. FCM and manual results were compared using the paired *t* test with the R statistical software.

RESULTS AND DISCUSSION

Optimization of sample dispersion. The swine manure samples used in this study contained large aggregates of cells, and these aggregates interfered with automated FISH analysis. Based on qualitative microscopic analysis, sonication with a sonic dismembrator was identified as the best method to disperse the aggregates. When this method was used, almost no cell aggregates remained, while the four other methods (vortexing, shearing with a syringe and needle, mechanical homogenization, and sonication in a sonication bath, as described in the supplemental material) produced samples that still contained many large aggregates. The efficiencies of transfer from the polycarbonate membranes to the microscope slides were approximately 77, 85, and 69%, for NF manure, CF manure, and soil, respectively. During this analysis, no pattern was observed for the morphology of cells that were retained on the membrane, but it is possible that some types of microorganisms were preferentially transferred, biasing the results.

The sonication time was then optimized for each sample type to maximize the dispersal of large aggregates, while minimizing cell lysis (see the supplemental material). The sonication times used in the remainder of this work were 15, 30, and 15 s for NF manure, CF manure, and soil, respectively. Using this protocol, a relatively uniform particle size was observed; i.e., more than 96% of the particles were particles that had an area less than 5 μm^2. This homogeneous particle size reduced the interference of aggregates and facilitated image analysis. Fortuitously, following this treatment, a more homogeneous background with relatively similar intensities was also observed, possibly due to the breakup of debris and inorganic particles during the sonication treatment.

Validation of cell recognition. The objectives of this analysis were to automatically detect cells in DAPI images and to determine fluorescence intensities for the cells in the corresponding FISH images. DAPI masks were generated automatically as described above. To validate whether the binarized DAPI masks were accurate, 30 random DAPI images were selected, and automatically binarized DAPI masks were compared with the original DAPI images. The numbers of cells recognized in DAPI images were not significantly different from the manual counts ($P > 0.05$). A qualitative comparison also suggested that the cell shapes in the binarized DAPI images were similar to the cell shapes in the original DAPI images (representative images are shown in Fig. 1). These results demonstrated that the automated procedure accurately detected cells.

Evaluation of threshold methods. The intensity information was converted into a classification of cells as target or nontarget cells using two commonly used methods: using a fixed threshold above the background level (26) and using a fixed signal-to-noise ratio (14, 23, 25). To set thresholds, five FISH images from each sample were analyzed manually, and the percentage of target cells based on the number of total cells was calculated for each image. For each of the five images the threshold was adjusted until the percentage of target cells matched the manual counts for the image, and the resulting five thresholds were averaged and used to analyze the entire set of images in the experiment.

The thresholds showed surprising variability. For images from a single experiment (CF manure), the fixed thresholds ranged from 3.90 to 9.15. Although the differences may appear to be slight, the results were very sensitive to the threshold; applying the lowest and highest thresholds to the entire set of images decreased the percentage of target cells from 75.4 to 35.7%. Counting multiple images and using an average threshold reduced but did not eliminate this problem, at least in part due to variations between experiments. Using the average threshold from a duplicate FISH experiment increased the percentage of target cells by 15.1% for CF manure (Table 1). Similar results were observed for FISH experiments with probe Arch0915 and with thresholds based on signal-to-noise ratios and also for the other two samples with both probes (Table 1).

Although methods in which a fixed threshold above the background level or a fixed signal-to-noise ratio is used are simple and straightforward and can give good results if the sample and background fluorescence intensities are uniform and consistent, the results in Table 1 indicate that these methods were not suitable for swine waste and soil samples. The need to perform manual counting to select the threshold for each individual replicate made this approach impractical.

Classification through FCM. To avoid the problems associated with threshold-based approaches when the intensities for the background, target cell, and nontarget cell signals varied within an image and among experiments, we turned to the idea of cluster analysis, i.e., distinguishing similar groups within a data set, using FCM. The idea behind this method is that the intensities of target cells should be more similar to each other than to the intensities of nontarget cells, whether or not the absolute intensities of target cells are constant in an image or in duplicate experiments. Good separation in cluster analysis depends on appropriate selection of the number of clusters. We investigated a number of approaches to optimize this selection. The initial estimates were based on the 13 major bacterial groups observed by Cotta et al. in their clone library analysis of swine waste samples (9) and six major bacterial groups observed in soil samples by Kobabe et al. (20). Using a more formal approach, we calculated the fuzziness performance index and modified partition entropy with numbers of clusters ranging from 2 to 15 (16). This analysis sometimes suggested a value less than 5 clusters but most often returned values of 6 to 10 clusters. Our final determination of whether the clustering was successful was based on whether the resulting clusters were clearly distinguishable as positive or negative. When less than five clusters were used, the clusters retrieved contained similar numbers of positive and negative cells, preventing proper classification. Burrough et al. also suggested that the optimal number of clusters should be chosen based on the required degree of detail (7). Therefore, we selected a more conservative value, 10 clusters, in the remaining range of 6 to 10 clusters. Considering that this value successfully clustered FISH data obtained with three sample types and two probes, we believe that it is suitable for routine analysis of FISH images from environmental samples.

The results of FCM clustering ($c = 10$) for ethanol-fixed CF

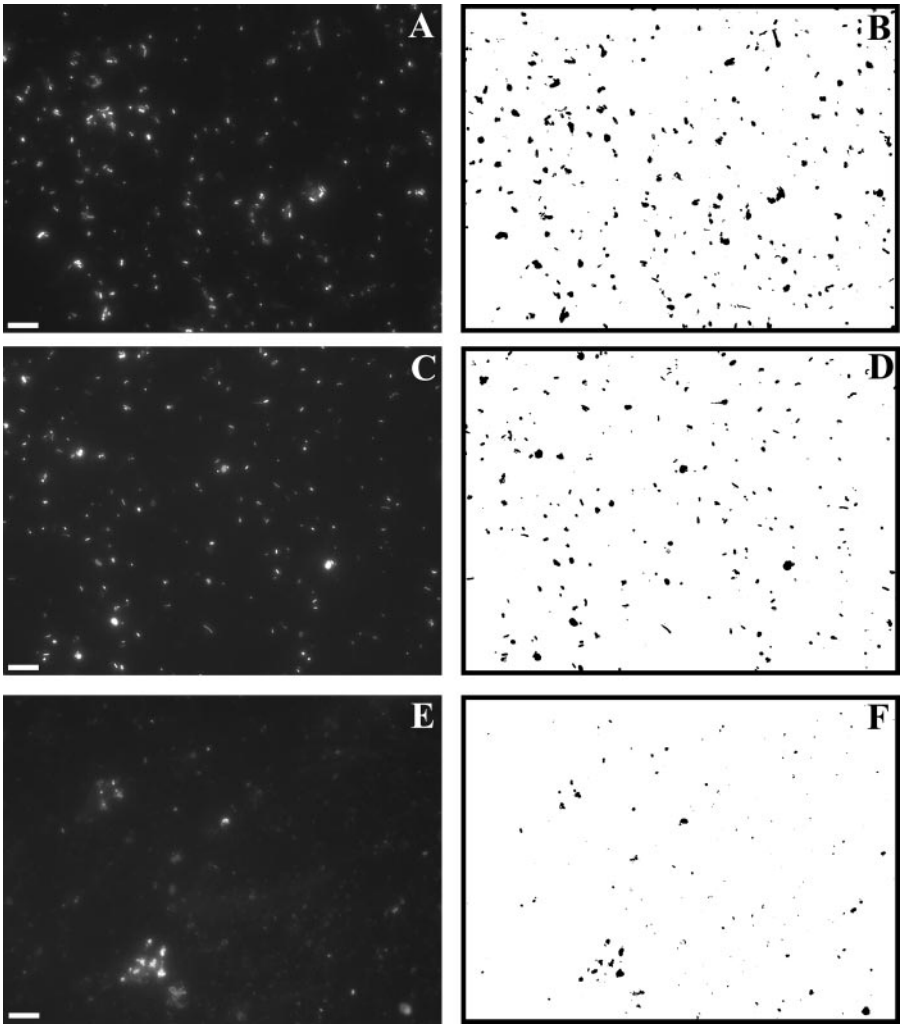

FIG. 1. Effectiveness of automatic cell recognition. (A and B) Original DAPI image (A) and binarized DAPI mask (B) for NF manure; (C and D) original DAPI image (C) and binarized DAPI mask (D) for CF manure; (E and F) original DAPI image (E) and binarized DAPI mask (F) for soil. The black areas in panels B, D, and F represent recognized cells. Scale bar, 10 μm.

manure samples are shown in Tables 2 and 3. The resulting 10 clusters were classified as positive or negative using a fixed threshold approach; in contrast to classification with fixed thresholds for individual objects, subsequent validation demonstrated that the same threshold applied to duplicate experiments and even to different probes with the same label. The threshold used was the average value for the maximum intensity of each cell in the negative control (no probe) for the sample. Hybridizations with an antisense bacterial probe were also performed but could not be used as negative controls because of highly variable fluorescence in both the background and cell areas, as observed previously with soil samples (8). The average values (standard errors) for the maximum intensity for negative controls without probes for the three sample types, NF manure, CF manure, and soil, were 6.24 (0.11), 6.03 (0.46), 5.69 (0.15), respectively, and these values were used as thresholds for classification.

To validate this classification, the clusters on either side of the thresholds were checked manually. Fifty cells from each of these two clusters were randomly selected, interspersed, classified manually, and separated by clusters. The results validated the fixed threshold classification of clusters for all three sample types in duplicate experiments and with both probes. As determined by blind manual counting, cells from all negative clusters were more than 94% negative, while cells from positive clusters were more than 96% positive. These data validate both the number of clusters and the classification method. This manual validation step is an important quality control step for the procedure and was designed to identify any problems arising from the application of 10 clusters and a threshold based on the negative control (no probe) to different samples, probes, and labels.

Validation of complete method. The final results of the new automated image analysis method were validated by comparison to manual counts. Using the automated method, the percentages of *Bacteria* and *Archaea* in the NF manure, CF manure, and soil were not significantly different from those obtained by manual counting ($P > 0.05$) (Fig. 2). These results suggest that classification based on FCM is an effective method for quantification of FISH signals, particularly since this

TABLE 1. Evaluation of existing threshold-based methods: change in percentage of target cells when thresholds from duplicate experiment were used[a]

Samples	Probe	Fixed threshold[b]			Fixed signal-to-noise ratio[c]		
		% Target cells (expt 1 threshold)	% Target cells (expt 2 threshold)	Change (%)	% Target cells (expt 1 threshold)	% Target cells (expt 2 threshold)	Change (%)
NF manure	Bact0338	42.8	76.0	33.2	45.1	81.9	36.8
	Arch0915	15.5	21.1	5.6	14.0	23.0	9.0
CF manure	Bact0338	60.0	75.1	15.1	63.0	66.1	3.1
	Arch0915	14.0	9.8	−4.2	14.0	5.5	8.5
Soil	Bact0338	65.2	79.2	14.0	65.2	73.9	8.7
	Arch0915	17.7	32.6	14.9	17.0	31.7	14.7
Avg				14.5			13.5

[a] Data were obtained for hybridization of ethanol-fixed samples for each probe; experiments 1 and 2 were replicates. The average number of cells per hybridization was 1,772 in 15 images.
[b] Fixed thresholds were set based on manual counts for a subset of images and were applied to the maximum intensity above the mean background level.
[c] Fixed signal-to-noise ratios were set based on manual counts for a subset of images and were the ratio of uncorrected maximum intensity to mean background intensity.

method performed well for two probes targeting populations present at different levels. Moreover, the fact that the same threshold could be used for duplicate experiments, and indeed for different probes with a given label, suggests that the FCM method is more robust than threshold-based classification of individual objects. This is presumably because FCM clustering recovers the intrinsic structure of the fluorescence intensity data, and the subsequent classification is therefore less sensitive to changes in the threshold.

It is also important to evaluate the accuracy of our results.

TABLE 2. Relative cluster sizes and cluster centroids resulting from FCM clustering for ethanol-fixed samples hybridized with Bact0338[a]

Samples	Cluster	Expt 1				Expt 2			
		Relative cluster size (%)	Maximum intensity	Mean intensity	Validated	Relative cluster size	Maximum intensity	Mean intensity	Validated
NF manure	1	**18.83**	**3.82**	**1.81**	−	**17.72**	**2.05**	**0.12**	−
	2	**30.07**	**6.03**	**3.67**	−	**32.50**	**4.53**	**2.20**	−
	3	23.07	8.39	5.32	+	20.51	7.38	3.93	+
	4	14.37	11.22	6.91	+	10.42	11.47	5.82	+
	5	7.30	15.26	9.36	+	6.86	16.91	8.03	+
	6	3.54	20.60	12.10	+	4.63	22.54	10.36	+
	7	1.78	27.95	16.04	+	3.76	30.45	16.22	+
	8	0.64	42.09	19.84	+	2.52	43.11	18.81	+
	9	0.27	67.75	25.88	+	0.54	72.12	30.41	+
	10	0.12	100.82	26.75	+	0.53	87.27	52.78	+
CF manure	1	**31.82**	**2.81**	**0.65**	−	**24.01**	**2.73**	**0.62**	−
	2	28.64	6.52	3.38	+	31.89	6.14	3.32	+
	3	17.53	12.11	6.80	+	18.73	10.51	5.82	+
	4	10.79	19.70	11.22	+	12.30	16.68	8.45	+
	5	6.52	29.37	17.22	+	6.64	25.72	12.49	+
	6	2.11	47.24	24.45	+	3.22	37.81	17.08	+
	7	1.21	71.29	37.19	+	1.34	57.29	24.45	+
	8	0.79	102.42	40.76	+	0.84	87.29	39.82	+
	9	0.36	123.96	67.18	+	0.77	143.02	56.66	+
	10	0.23	198.15	155.97	+	0.27	203.56	90.83	+
Soil	1	**9.87**	**2.31**	**0.30**	−	**10.93**	**1.64**	**−0.20**	−
	2	**25.94**	**4.44**	**2.33**	−	**23.54**	**3.87**	**1.70**	−
	3	28.55	6.44	3.86	+	23.64	5.80	3.32	+
	4	19.27	8.78	5.54	+	19.13	7.99	4.69	+
	5	10.44	11.97	7.89	+	12.89	10.88	6.38	+
	6	3.69	17.83	11.17	+	5.20	15.05	8.29	+
	7	0.76	28.14	13.02	+	2.78	22.00	10.80	+
	8	0.86	36.56	27.52	+	1.15	30.79	16.01	+
	9	0.37	52.88	38.96	+	0.52	48.63	25.42	+
	10	0.26	87.29	68.66	+	0.23	97.56	15.65	+

[a] The values in bold type are below the threshold based on the average maximum intensity of cells in negative controls (no probe).

TABLE 3. Relative cluster size and cluster centroids resulting from FCM clustering for ethanol-fixed samples hybridized with Arch0915[a]

Samples	Cluster	Expt 1				Expt 2			
		Relative cluster size (%)	Maximum intensity	Mean intensity	Validated	Relative cluster size (%)	Maximum intensity	Mean intensity	Validated
NF manure	1	**10.20**	**1.10**	**0.17**	−	**28.41**	**1.94**	**0.75**	−
	2	**21.48**	**2.14**	**0.95**	−	**34.10**	**3.56**	**1.72**	−
	3	**19.46**	**3.12**	**1.58**	−	**16.86**	**5.38**	**2.80**	−
	4	**17.03**	**4.12**	**2.10**	−	10.53	7.95	4.11	+
	5	**9.37**	**5.13**	**2.78**	−	5.60	11.56	5.73	+
	6	11.60	6.53	3.34	+	2.90	17.05	10.22	+
	7	5.80	8.66	4.54	+	0.67	25.51	15.02	+
	8	2.64	12.11	5.40	+	0.61	45.59	23.05	+
	9	2.17	18.16	8.93	+	0.15	164.89	89.46	+
	10	0.24	27.12	24.01	+	0.15	167.32	71.47	+
CF manure	1	**17.32**	**0.81**	**−0.06**	−	**13.93**	**0.81**	**−0.01**	−
	2	**36.46**	**1.95**	**0.69**	−	**33.85**	**1.83**	**0.65**	−
	3	**23.99**	**3.33**	**1.55**	−	**26.92**	**3.07**	**1.44**	−
	4	**11.87**	**5.65**	**3.04**	−	**11.14**	**5.11**	**2.73**	−
	5	6.01	9.44	5.62	+	6.29	7.42	4.25	+
	6	2.79	14.96	8.12	+	4.34	10.21	6.76	+
	7	0.90	23.22	15.29	+	1.61	14.97	8.12	+
	8	0.41	45.99	19.86	+	1.14	21.48	13.49	+
	9	0.17	115.88	57.07	+	0.62	45.15	16.76	+
	10	0.08	235.01	79.63	+	0.16	105.78	50.27	+
Soil	1	**6.33**	**0.79**	**0.00**	−	**6.98**	**0.59**	**−0.30**	−
	2	**26.22**	**1.53**	**0.53**	−	**13.31**	**1.12**	**0.27**	−
	3	**33.73**	**2.37**	**1.03**	−	**16.39**	**1.63**	**0.86**	−
	4	**18.37**	**3.29**	**1.63**	−	**16.24**	**1.75**	**0.38**	−
	5	**7.77**	**4.26**	**2.37**	−	**20.94**	**2.34**	**0.83**	−
	6	**4.12**	**5.59**	**3.32**	−	**16.94**	**2.51**	**1.35**	−
	7	1.98	7.42	4.72	+	**7.77**	**3.47**	**1.78**	−
	8	0.83	11.97	6.29	+	1.17	5.89	3.64	+
	9	0.48	17.16	10.46	+	0.21	13.02	5.93	+
	10	0.19	26.43	17.96	+	0.05	30.10	21.86	+

[a] The values in bold type are below the threshold based on the average maximum intensity of cells in negative controls (no probe).

For example, are the levels of bacterial and archaeal cells in the typical swine manure sample (CF manure) indeed close to 72.1 and 12.3%, respectively? The limited data available for such an analysis suggest that the numbers are reasonable. The most relevant analysis is an analysis of a full-scale anaerobic sequencing batch reactor treating swine waste, and this reactor contained similar distributions of *Bacteria* (65 to 81% of total small-subunit [SSU] rRNA) and *Archaea* (9 to 15% of total SSU rRNA) (3). Cattle manure was found to contain similar concentrations of *Bacteria* (72 to 98% of total SSU rRNA) but lower concentrations of *Archaea* (0.47 to 7% of total SSU rRNA) (17, 24).

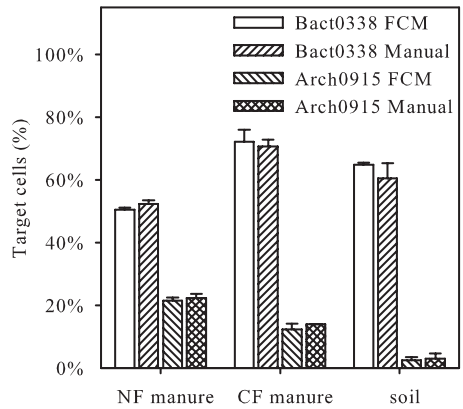

FIG. 2. Comparison of FCM clustering and manual counting results. The error bars indicate the half-range for duplicate experiments.

Concluding remarks. In this study, an optimized dispersion protocol and an automated, quantitative image analysis procedure were developed for analysis of swine waste samples by FISH. The dispersion protocol was a necessary initial step, because the presence of large aggregates of cells prevented automated detection of individual cells and was responsible for large variations in fluorescence intensity. The automated image analysis and classification via FCM are faster than previously described methods that rely on manual calibration and do not require the extensive experience with computational methods or the large training data set that is used in more complicated methods. Together, the methods developed in this work should facilitate quantitative analysis of FISH images in complex environmental samples, such as manure samples and soil samples, particularly when they are used in combination with automated image acquisition systems (25, 27).

ACKNOWLEDGMENTS

We acknowledge funding from the Cooperation Program UIUC/CNRS, France, from the U.S. Department of Agriculture under coop-

erative agreement AG 58-3620-1-179, and from National Pork Board project 05-068.

We thank the anonymous reviewers and Toshio Shimada for helpful comments and discussions. We thank the farmers for their cooperation and Archana Jindal and Matt Robert for collecting the CF manure.

REFERENCES

1. **Amann, R. I., B. J. Binder, R. J. Olson, S. W. Chisholm, R. Devereux, and D. A. Stahl.** 1990. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. **56:**1919–1925.
2. **Amann, R. I., L. Krumholz, and D. A. Stahl.** 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J. Bacteriol. **172:**762–770.
3. **Angenent, L. T., S. Sung, and L. Raskin.** 2002. Methanogenic population dynamics during startup of a full-scale anaerobic sequencing batch reactor treating swine waste. Water Res. **36:**4648–4654.
4. **Bauman, J. G., J. Wiegant, P. Borst, and P. van Duijn.** 1980. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome labelled RNA. Exp. Cell Res. **128:**485–490.
5. **Blackburn, N., A. Hagstrom, J. Wikner, R. Cuadros-Hansson, and P. K. Bjornsen.** 1998. Rapid determination of bacterial abundance, biovolume, morphology, and growth by neural network-based image analysis. Appl. Environ. Microbiol. **64:**3246–3255.
6. **Budny, B., M. Kanik, and A. Latos-Bieleniska.** 2002. Fluorescence in situ hybridization (FISH)—application in research and diagnostics. Folia Histochem. Cytobiol. **40:**107–108.
7. **Burrough, P. A., P. F. M. van Gaans, and R. A. MacMillan.** 2000. High-resolution landform classification using fuzzy k-means. Fuzzy Sets Syst. **113:** 37–52.
8. **Christensen, H., M. Hansen, and J. Sorensen.** 1999. Counting and size classification of active soil bacteria by fluorescence in situ hybridization with an rRNA oligonucleotide probe. Appl. Environ. Microbiol. **65:**1753–1761.
9. **Cotta, M. A., T. R. Whitehead, and R. L. Zeltwanger.** 2003. Isolation, characterization and comparison of bacteria from swine faeces and manure storage pits. Environ. Microbiol. **5:**737–745.
10. **DeLong, E. F., G. S. Wickham, and N. R. Pace.** 1989. Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells. Science **243:**1360–1363.
11. **de los Reyes, F. L., W. Ritter, and L. Raskin.** 1997. Group-specific small-subunit rRNA hybridization probes to characterize filamentous foaming in activated sludge systems. Appl. Environ. Microbiol. **63:**1107–1117.
12. **Dembele, D., and P. Kastner.** 2003. Fuzzy C-means method for clustering microarray data. Bioinformatics **19:**973–980.
13. **Fraley, C., and A. E. Raftery.** 2002. Model-based clustering, discriminant analysis, and density estimation. Technical report no. 380, Department of Statistics. J. Am. Stat. Assoc. **97:**611–631.
14. **Garini, Y., A. Gil, I. Bar-Am, D. Cabib, and N. Katzir.** 1999. Signal to noise analysis of multiple color fluorescence imaging microscopy. Cytometry **35:** 214–226.
15. **Gasch, A. P., and M. B. Eisen.** 2002. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. Genome. Biol. **3:**RESEARCH0059.
16. **Gorsevski, P. V., P. E. Gessler, and P. Jankowski.** 2003. Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. J. Geograph. Syst. **5:**223–251.
17. **Griffin, M. E., K. D. McMahon, R. I. Mackie, and L. Raskin.** 1998. Methanogenic population dynamics during start-up of anaerobic digesters treating municipal solid waste and biosolids. Biotechnol. Bioeng. **57:**342–355.
18. **Jindal, A.** 2002. Antimicrobial resistance in swine waste treatment processes. M.S. thesis. University of Illinois, Urbana.
19. **Kapur, J. N., P. K. Sahoo, and A. K. C. Wong.** 1985. A new method for gray-level picture thresholding using the entropy of the histogram. Comput. Vis. Graphics Image Process. **29:**273–285.
20. **Kobabe, S., D. Wagner, and E.-M. Pfeiffer.** 2004. Characterisation of microbial community composition of a Siberian tundra soil by fluorescence in situ hybridisation. FEMS Microbiol. Ecol. **50:**13–23.
21. **Langendijk, P. S., F. Schut, G. J. Jansen, G. C. Raangs, G. R. Kamphuis, M. H. Wilkinson, and G. W. Welling.** 1995. Quantitative fluorescence in situ hybridization of *Bifidobacterium* spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples. Appl. Environ. Microbiol. **61:** 3069–3075.
22. **Lerner, B.** 2004. Bayesian fluorescence in situ hybridisation signal classification. Artif. Intellig. Med. **30:**301–316.
23. **Li, S., R. N. Spear, and J. H. Andrews.** 1997. Quantitative fluorescence in situ hybridization of *Aureobasidium pullulans* on microscope slides and leaf surfaces. Appl. Environ. Microbiol. **63:**3261–3267.
24. **McMahon, K. D., P. G. Stroot, R. I. Mackie, and L. Raskin.** 2001. Anaerobic codigestion of municipal solid waste and biosolids under various mixing conditions. II. Microbial population dynamics. Water Res. **35:**1817–1827.
25. **Pernthaler, J., A. Pernthaler, and R. Amann.** 2003. Automated enumeration of groups of marine picoplankton after fluorescence in situ hybridization. Appl. Environ. Microbiol. **69:**2631–2637.
26. **Shopov, A., S. C. Williams, and P. G. Verity.** 2000. Improvements in image analysis and fluorescence microscopy to discriminate and enumerate bacteria and viruses in aquatic samples. Aquat. Microb. Ecol. **22:**103–110.
27. **Singleton, S., J. G. Cahill, G. K. Watson, C. Allison, D. Cummins, T. Thurnheer, B. Guggenheim, and R. Gmur.** 2001. A fully automated microscope bacterial enumeration system for studies of oral microbial ecology. J. Immunoass. Immunochem. **22:**253–274.
28. **Stahl, D. A., and R. Amann.** 1991. Development and application of nucleic acid probes, p. 205–248. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons Ltd., Chichester, England.
29. **Vincent, L., and P. Soille.** 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Analysis Machine Intelligence **13:**583–598.
30. **Wang, J., T. H. Bo, I. Jonassen, O. Myklebost, and E. Hovig.** 2003. Tumor classification and marker gene prediction by feature selection and fuzzy c-means clustering using microarray data. BMC Bioinformatics **4:**60.