

Review

Use of CMEIAS Image Analysis Software to Accurately Compute Attributes of Cell Size, Morphology, Spatial Aggregation and Color Segmentation that Signify *in Situ* Ecophysiological Adaptations in Microbial Biofilm Communities

Frank B. Dazzo * and Brighid C. Niccum

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA; E-Mail: niccumbr@msu.edu

* Author to whom correspondence should be addressed; E-Mail: dazzo@msu.edu; Tel.: +1-517-884-5394.

Academic Editor: Rainer Breitling

Received: 24 November 2014 / Accepted: 25 February 2015 / Published: 9 March 2015

Abstract: In this review, we describe computational features of computer-assisted microscopy that are unique to the Center for Microbial Ecology Image Analysis System (CMEIAS) software, and examples illustrating how they can be used to gain ecophysiological insights into microbial adaptations occurring at micrometer spatial scales directly relevant to individual cells occupying their ecological niches *in situ*. These features include algorithms that accurately measure (1) microbial cell length relevant to avoidance of protozoan bacteriovory; (2) microbial biovolume body mass relevant to allometric scaling and local apportionment of growth-supporting nutrient resources; (3) pattern recognition rules for morphotype classification of diverse microbial communities relevant to their enhanced fitness for success in the particular habitat; (4) spatial patterns of coaggregation that reveal the local intensity of cooperative *vs.* competitive adaptations in colonization behavior relevant to microbial biofilm ecology; and (5) object segmentation of complex color images to differentiate target microbes reporting successful cell-cell communication. These unique computational features contribute to the CMEIAS mission of developing accurate and freely accessible tools of image bioinformatics that strengthen microscopy-based approaches for understanding microbial ecology at single-cell resolution.

Keywords: biofilm; biovolume; cell aggregation; CMEIAS; color segmentation; computer-assisted microscopy; image analysis; morphological diversity; spatial ecology

1. Introduction

Microscopy is an important technique used to examine the microbe's world from its own perspective and spatial scale. Its value is increased when combined with computer-assisted digital image analysis [1,2] to quantify the abundance, ecophysiology, morphotaxa diversity, and spatial distribution of microbes in relation to their interacting neighbors and local environments. These tools can enhance the polyphasic taxonomy approach to analyze complex microbial communities by providing direct visual feedback while minimizing the quantification biases associated with other community analysis methods [1,3,4].

We are interested in developing computing tools that can extract ecologically important information from digital images of microorganisms without laboratory cultivation. Our research team is pursuing this goal by developing a software application suite called CMEIAS (Center for Microbial Ecology Image Analysis System). When fully tested and documented, its copyrighted components are available for free download at its website [5]. The long-term mission of the CMEIAS project is to create accurate computing tools that strengthen quantitative, microscopy-based approaches for understanding microbial ecology. In this review, we describe several of the computational features of computer-assisted microscopy that are unique to CMEIAS digital image analysis software, and applications illustrating how they can be used to gain ecophysiological insights into microbial adaptations occurring with single cells and ecological niches they occupy *in situ*.

2. Descriptions of Unique CMEIAS Algorithms

2.1. Measurement of Microbial Body Size

2.1.1. Accurate Measurement of Cell Length and Width Using Shape-Adaptable Algorithms

Measurements of an object's length should be based on its principal skeleton, but this feature is difficult to calculate and can be very inaccurate for irregularly shaped objects. Alternatively, the length and width of objects are often computed as the major axis and minor axis lengths of the object's best enclosed ellipse. These metrics measure the longest possible straight-line distance between the object's border pixels, and perpendicular to that straight line, respectively. These algorithms suffice for round cocci and elongated morphotypes with a straight medial axis, e.g., regular rods, clubs, and ellipsoids. However, the major axis length will significantly underestimate the true length of curved or branched morphotypes, e.g., spirals, U-shaped rods, curved rods, rudimentary branched rods, and branched filaments. To circumvent this limitation and obtain accurate measurements of object size, CMEIAS uses an alternative "shape-adaptive" algorithm to measure cell length and width automatically [1]. First, CMEIAS classifies each object into one of two groups according to its degree of roundness shape (computed as $4 \times \pi \times \text{Area}/\text{Perimeter}^2$), and then applies the appropriate formula that computes cell

lengths and widths for the appropriate roundness category [1]. The formula used to compute cell length is as follows:

$$\text{Length} = \begin{cases} \frac{2\text{Perimeter} + (\pi - 2)\sqrt{\text{Perimeter}^2 - 4\pi\text{Area}}}{2\pi} & \text{if Roundness} \leq 0.8; \\ \text{Major Axis Length} & \text{otherwise.} \end{cases} \quad (1)$$

This process allows for both accuracy and efficiency in measuring the length of objects with varying shape. The width of an object is approximately computed for these same two roundness shape classes in a similar manner, using the following formula:

$$\text{Width} = \begin{cases} \frac{\text{Perimeter} - \sqrt{\text{Perimeter}^2 - 4\pi\text{Area}}}{\pi} & \text{if Roundness} \leq 0.8; \\ \frac{\text{Area}}{\text{Major Axis Length}} & \text{otherwise.} \end{cases} \quad (2)$$

2.1.2. Accurate Measurement of Biovolume Body Size Using Shape-Adaptable Algorithms

Biovolume body mass is a calculation of a cell's volume in three-dimensional space. Nowadays, the unique characteristics of optisectioned confocal imaging make this type of digital microscopy ideal for measuring the biovolume of 3-D microbial aggregates directly by voxel analysis of image stacks acquired using laser scanning confocal microscopy [2,6–11]. Computation of 3-D biovolume of microbes from 2-D projected images acquired by the more commonly used conventional phase-contrast or epifluorescence light microscopy assumes that the shape of the foreground objects of interest (*i.e.*, individual microbial cells) has approximate axial symmetry [12] and they are sampled with sufficient pixel density so their cell periphery can be accurately defined [13]. Measurements of the 3-D volume of irregularly shaped microbial cells and aggregates that lack axial symmetry can be facilitated by implementing integration algorithms that digitally rotate the object's perimeter to horizontal, measure its two pixel-wide serial cross sectional lengths, calculate the volume of each resultant thin cylinder, and finally sum up the volumes of all cylinders of the whole body mass [14].

Recently, we evaluated the accuracy of 17 different formulas to compute the biovolume of 11 different microbial morphotypes that are accurately classified by CMEIAS, including all of the major and most of the minor morphotypes that develop in natural and managed microbial communities [15]. Object biovolumes computed by 2-D digital image analysis using each formula were compared to ground truth data acquired by volume displacement measurements of 3-D model objects representing each morphotype. The results clearly indicated the strong influence that morphology had on accurate computation of each object's volume. A common feature shared by the two most accurate algorithms (shown below) is their use of computer vision to intelligently adapt the biovolume formula to the classification of the object's shape in computing their cell lengths and widths [15]. When applied to properly segmented images, these two most-accurate biovolume algorithms performed with >98%

accuracy for all morphotype populations and >96% accuracy for communities containing various combinations of these same morphotype populations [15]. This shape-adaptive strategy ultimately improves their accuracy to compute biovolumes for the diverse range of cell morphotypes in actively growing microbial communities that have high morphological diversity, as commonly occurs in nutritionally enriched habitats that support microbial growth.

$$\text{Biovolume Formula 1} = \begin{cases} \frac{\pi \text{Width}^2 \text{Length}}{4} & \text{if Roundness} \leq 0.8; \\ \frac{4\pi}{3} \sqrt{\left(\frac{\text{Area}}{\pi}\right)^3} & \text{otherwise.} \end{cases} \quad (3)$$

$$\text{Biovolume Formula 2} = \frac{\pi \text{Width}^2 \text{Length}}{4} \quad (4)$$

2.2. Rules of Pattern Recognition for Microbial Morphotype Classification

Automatic classification of many microbial morphotypes requires measurement of several shape and size attributes to resolve their distribution of morphological space [1]. The morphotype classification of spheres and regular rods representing the limited morphological diversity of bacterioplankton communities in some oligotrophic marine environments is a rather straightforward process of object shape analysis. However, comprehensive image analysis systems that can automatically classify greater morphological diversity in complex bacterial communities, as commonly exists in most all nutrient-enriched habitats containing actively growing bacteria that are larger in size and typically monomorphic, did not exist prior to development of CMEIAS. Figure 1 shows the diversity of microbial morphotypes classified by CMEIAS, the visual criteria that distinguish each class, and the pseudocolor assigned to each morphotype in the rendered output image that accompanies the CMEIAS morphotype classification analysis.

The CMEIAS Morphotype Classifier is a supervised, hierarchical tree classifier. Some of the classifying elements used are less complex to evaluate than others, as they only require easily-quantifiable structural information to compute. Thus the structural analysis portion of the classifier operates first, thereby reducing the complexity of the program [1]. Next, objects that cannot be classified by structural analysis alone are evaluated by a *k*-Nearest Neighbor classifier [1,16]. Some morphotype categories must be further split into subcategories to facilitate classification: the coccus category is split into spherical, ovoid, and coccobacillus subcategories, and the prosthecate category is split into the long prostheca and short prostheca subcategories; these are later rejoined into single parent categories before the final classification. The pattern recognition rules used by CMEIAS to classify the microbial morphotypes using structural analysis [1] are described next.












Classification Criteria	Morphotype (Label)	Pseudocolor Cell
I. Round cells (spheroid, ovoid, coccobacilli)	Cocci (A)	
II. Elongated Cells		
A. Unbranched (2 poles)		
1. Repeated waveform	Spiral (B)	
2. Single curvature		
a. Vibroid or crescent curvature	Curved Rod (C)	
b. U-shaped curvature	U-rod (D)	
3. Unbranched, no medial axis curvature		
a. Constant cell width		
i. Length/Width < 16:1	Regular Straight Rod (E)	
ii. Length/Width > 16:1	Unbranched Filament (F)	
b. Widest at center, tapered at poles	Ellipsoid (G)	
c. Wider at 1 pole		
i. Tapered at opposite pole	Club (H)	
ii. Thin stalked appendage	Prosthecate (I)	
B. Branched (> 2 poles)		
1. Length/Width < 16:1	Rudimentary Branched Rod (J)	
2. Length/Width > 16:1	Branched Filament (K)	

Figure 1. Visual characteristics of microbial morphotypes classified by CMEIAS (Center for Microbial Ecology Image Analysis Systems).

2.2.1. Rule (r_1) for Category A_1

Subcategory A_1 (spherical coccus) is the only truly spherically shaped subcategory in the CMEIAS morphotype classifier. A comparison of roundness *vs.* compactness allows a clear decision boundary to be drawn between objects of this subcategory and all other objects [1,16].

The formula for classifying an object x by rule r_1 is as follows:

$$r_1 : \begin{cases} \text{if } 24F_2(x) + 25F_3(x) - 43 \leq 0, & x \in \omega_{A_1}; \\ \text{if } 24F_2(x) + 25F_3(x) - 43 \geq 0, & x \in (\Omega - \omega_{A_1}), \end{cases} \quad (5)$$

where $F_2(x)$ and $F_3(x)$ represent roundness and compactness shape metrics, respectively, ω_{A_1} represents the A_1 subcategory, and Ω is the collection of all morphotype categories under consideration. For subsequent rules, it is important to have Ω_1 defined as $(\Omega - \omega_{A_1})$, representing all categories except subcategory A_1 [1].

2.2.2. Rule (r_2) for Elongated Morphotypes

The Group B (spiral), F (unbranched filament) and K (branched filament) morphotypes, as well as the I_1 long prosthecate subcategory, are separated from the remaining morphotype classes using their elongated nature [1]. The length/width ratio of 16:1, as computed from the shape-adaptable formulas described in Section 2.1.1, represents a reliable decision boundary of this metric to differentiate these elongated morphotypes from the remaining morphotype classes:

$$r_2 : \begin{cases} \text{if } \frac{1}{F_6(x)} - 16 \geq 0, & x \in \omega_{\text{elongated}}; \\ \text{if } \frac{1}{F_6(x)} - 16 \leq 0, & x \in (\Omega_1 - \omega_{\text{elongated}}), \end{cases} \quad (6)$$

where $F_6(x)$ represents the width/length ratio, $\omega_{\text{elongated}}$ represents the collection of all elongated morphotypes, and Ω_1 represents the collection of all morphotypes except for A_1 . It should be noted that while some prosthecae morphotypes have a long enough prostheca to be identified as elongated cells, others with short prostheca might not. This varies because the prostheca on some stalk-shaped species elongates as the mother cell produces swarmer daughter cells. This is why the prosthecae category is divided into the long prosthecae and short prosthecae subcategories [1].

2.2.3. Rule (r_3) for Further Classification of Prosthecae Subcategory I_1 , Spiral Group B and Branched Filament Group K

This rule allows for further classification of the $\omega_{\text{elongated}}$ morphotypes [1]. CMEIAS counts the number of poles (*i.e.*, the thin pointed ends) and branch points (*i.e.*, the places where the object splits into two distinct filaments) on each object. It combines this information with the local minimal and maximal angles at the periphery of the object to classify each object by the following rule:

$$r_3 : \begin{cases} \text{if } K(x) > 0 \text{ or } P(x) > 2, & x \in \omega_K; \\ \text{if } K(x) = 0 \text{ and } P(x) = 2, & x \in \omega_B \text{ or } \omega_F; \\ \text{if } K(x) = 0 \text{ and } P(x) < 2, & x \in \omega_{I_1}, \end{cases} \quad (7)$$

where $K(x)$ and $P(x)$ represent the number of branch points and poles in the object x , respectively, ω_K represents objects in Group K , ω_B represents objects in group B , ω_F represents objects in Group F , and ω_{I_1} represents objects in subcategory Group I_1 [1].

2.2.4. Rule (r_4) for Spiral and Unbranched Filament Morphotypes

Objects of the spiral morphotype can be differentiated from those of the unbranched filament morphotype by analyzing whether their waveform properties are repeated [1]. CMEIAS discriminates between these two groups by analyzing local angles along the edges of the cell. Spiral cells will show much greater deviation in local angle than unbranched filaments. Let A_R (“average for ridge”) represent the means of all local angles greater than 180° , and A_V (“average for valley”) represent the means of all local angles less than 180° . The mean of these angles can then be used to discriminate between these two morphotypes, according to the following formula:

$$r_4 : \begin{cases} \text{if } \{(A_R(x) + 120^\circ) < 180^\circ\} \text{ and } \{(A_V(x) - 120^\circ) > 180^\circ, x \in \omega_B; \\ x \in \omega_F \text{ otherwise,} \end{cases} \quad (8)$$

where $A_R(x)$ and $A_V(x)$ represent the values of A_R and A_V , respectively, for object x [1].

2.2.5. CMEIAS k -NN Morphotype Classifier

The next portion of the morphotype classifier was more challenging to develop, as it is difficult to quantify the differences between some morphotypes that have very clear visual distinctions. For example, the ellipsoid and club morphotypes can be distinguished very easily by manual analysis, based on the position of their widest point along their longitudinal axes. However, although human observers can very easily tell these two morphotypes apart (Figure 1), it is not easy to code this distinction into a computer for automated analysis.

CMEIAS operation assumes that these differences can be distinguished using combinations of shape measurements, which led to the decision to use a k -NN *classification system* [1]. This system uses the k -Nearest Neighbor rule, which assigns the morphotype label (Figure 1) of the k nearest samples in the training set to the unknown object [16,17]. Use of this system is the reason for dividing the coccus morphotype category A into three different shape subcategories: A_1 (spheroid), A_2 (ovoid) and A_3 (coccobacillus), as the k -NN *classifier* is more accurate when it treats them as different categories. Since all three subcategories can coexist in the same streptococcal chain, they are recombined later [1].

Use of the k -NN *classification system* necessitated two important decisions: which measurement features to utilize for analysis, and which value of k to use [16,17]. Using as few measurement features as possible in the analysis is desirable for several reasons. In cases where the training sets are small relative to the number of measurement features evaluated, the accuracy of the evaluation will decrease after a threshold number of features are added (known as the “Curse of Dimensionality”). A ratio of roughly 10:1 of the measurement set to the number of measurement features is desirable. This effect is caused by the interdependency of certain measurement features (for example, length and the length/width ratio are by nature highly related). In addition, the more features used, the more redundancy there is in the calculations, resulting in longer computational time (also known as the “Computational Cost”). Narrowing the set down to only the most effective measurement features needed to discriminate between morphotypes allows for more efficient classification.

A Floating Search Method of feature selection analysis [17] indicated that the most effective method of classification used a k -value of three, and 14 out of 15 candidate measurement features: elongation, roundness, compactness, area/bounding box area, width/length ratio, maximum curvature, and eight Fourier descriptors [1]. In a training classification, the accuracy of the classification based on structural judgments (*i.e.*, discrimination of cocci, spirals, unbranched filaments and prosthecae) was 100%. The accuracy of classification based on the k -NN *classifier* was 94.9%. This yielded an overall accuracy of 96.0% [1].

2.2.6. Importance of Fourier Descriptors in the Morphotype Classifier

The Fourier descriptors were the most complex measurement features needed to complete the morphotype classifier, and proved especially necessary for the classification of ovoids, coccobacilli, clubs, curved rods, short prosthecates, U-shaped rods, ellipsoids, and rudimentary branched rods [1]. Fourier descriptors are based on object contour (*i.e.*, the outline of the object) and do not change with scaling, rotation, or translation of that object [18,19]. The outline of the object being measured is recreated using a series of equidistant points (x_k, y_k) , $k = 0, \dots, N-1$, with a constant distance between each pair of neighboring points. Next, the sequence Z_k is examined. This is a sequence in complex space such that $z_k = x_k + jy_k$, $k = 0, \dots, N-1$. Z_k represents the sum of its Fourier transform coefficients, as detailed by the equation:

$$z_k = \sum_{n=0}^{N-1} a_n e^{\frac{-jnk2\pi}{N}}, \quad k = 0, \dots, N-1 \quad (9)$$

where a_n represents the Fourier transformation coefficients, represented by the equation:

$$a_n = \frac{1}{N} \sum_{k=0}^{N-1} z_k e^{\frac{jnk2\pi}{N}}, \quad n = 0, \dots, N-1 \quad (10)$$

and a_0 represents the mean of z_k , $k = 0, \dots, N-1$.

An altered version of the Fourier equations was developed to represent changes to the original vectors [1,18,19]. The equation:

$$z'_k = S z_{k+t} e^{j\phi} + T, \quad n = 1, \dots, N-1 \quad (11)$$

represents a distortion of z_k , or an alteration of the original characteristics (e.g., shape) of z_k . In this equation, S represents the scaling coefficient (the value by which z_k was changed in size), T represents the translation vector in the complex space (the distance that z_k was moved), ϕ represents the rotation angle, and t represents the change in the starting point of z_k . This will also alter the Fourier coefficients as follows:

$$a'_0 = a_0 + T \quad (12)$$

$$a'_n = a_n S e^{\frac{jn2\pi}{N}} e^{j\phi}, \quad n = 1, \dots, N-1 \quad (13)$$

Fourier descriptors are used in the CMEIAS morphotype classifier to reconstruct the outline of the complex shaped objects; the more Fourier descriptors used in the reconstruction, the more accurate the reconstructed outline will be. However, once the required number of Fourier descriptors is utilized, adding more will have a negligible effect on the accuracy of the reconstruction. A reconstruction analysis of the eight morphotypes listed earlier indicated that 8 Fourier descriptors were necessary and sufficient, so the decision was made to use $f_2, f_3, f_4, f_5, f_{N-1}, f_{N-2}, f_{N-3}$, and f_{N-4} in its measurements (designated as FD0-FD7 in the CMEIAS Object Analysis result window) [1].

The accuracies of the fully functional CMEIAS morphotype classifier for properly edited images were as follows [1]: coccus 99.6%; spiral 97.8%; curved rod 95.1%; U-shaped rod 89.2%; regular rod 95.0%; unbranched filament 98.8%; ellipsoid 94.9%; club 96.4%; prosthecate 100.0%; rudimentary branched rod 92.0%; and branched filament 100.0%. The main source of error was found to be due to the biological continuum between similar morphotypes, and therefore a reassign classification feature was included in the released software to accommodate the infrequent errors that occur [1].

2.3. Geospatial Pattern Analysis: CMEIAS Aggregation Cluster Index

CMEIAS computes several metrics that address the spatial distribution of microorganisms within immature biofilm landscapes. The extracted data can be analyzed by geospatial statistics to define the biogeography of microbial cells during biofilm development, and to generate predictive models of their colonization behavior that define the intensity of cell-cell interactions and the spatial scales of their niche [4]. Geostatistical analysis requires that a relevant **Z**-variate be assigned to each georeferenced site sampled. This parameter is a non-binary, quantitative attribute that is continuously distributed in the landscape, and the location of each sampling point (e.g., individual microbe) must be indicated at defined **X**, **Y** Cartesian coordinates relative to the (0, 0) landmark origin within the same domain [4,20]. We created the **CMEIAS cluster index** as the major **Z**-variate for geospatial analyses of microbial biofilm colonization behavior [21]. This unique metric is computed as the inverse of the first nearest neighbor distance [21], and the spatial scale of its ecological influence typically lies in the micrometer range [4]. The magnitude of this index indicates the strength of each microbe's clustered distribution relative to other microbes in its local environment [4] and is a sensor of positive, spatially autocorrelated cell interactions that cooperatively promote bacterial colonization behavior *in situ* [4,21,22]. A high cluster index is assigned to cells in aggregated distributions that assist in positive cell-to-cell interactions, enabling promotion of their growth into microcolony biofilms [4,21,22].

2.4. Color Segmentation Algorithm

Computer-assisted microscopy is underutilized in studies of microbial ecology mainly because segmentation of foreground object pixels is difficult, particularly in complex digital images of real-world samples from the environment (rather than generated in the lab), since their backgrounds often contain significant noise [4,23]. Image segmentation is even more complex when cells are differentially colored to discriminate their ecophysiological and/or phylogenetic characteristics [4,23]. We developed a CMEIAS Color Segmentation tool to address this limitation [23]. It has the unique ability to remove background color pixels whose red, green and blue (RGB) values fall within the range of foreground objects of interest [23].

The image segmentation algorithms address color and spatial relationships of user-selected foreground object pixels [23]. The core of this application uses “instance-based learning” logic to classify the image pixels as foreground or background, based on the local neighborhoods of their color features in the image grid [23]. The information needed to optimize the segmentation output comes from samples of training pixels that accurately denote the color and spatial features of the desired target. Then the linear distances between the specific RGB positions of each image pixel and the training sample pixels are compared, followed by applying a distance-weighted similarity function that

combines the color and spatial distances [23]. Next, the image is edited using these color and spatial comparison analyses related to the location of other pixels within the image [23]. Without these latter steps, background pixels with color similar to the foreground objects would misclassify as foreground, therefore failing to achieve the desired goal. These weighted similarity measurements provide sufficient flexibility to exclude background pixels with RGB values that are similar to foreground pixels in the analysis, and can successfully segment complex images containing multiple regions of interest with different color ranges [23]. This scenario commonly occurs when analyzing microbial communities in environmental samples. Finally, the system displays a new RGB image with the segmented foreground object pixels displayed *in situ* in their original color and the remaining pixels colored white or black (user-specified) to create the noise-free background [23]. Repeats of the protocol and other editing routines featured in the software can be applied if needed to produce the desired final segmentation.

Another important feature of the CMEIAS color segmentation software is the color inclusion threshold that controls the size of the color space that neighboring image pixels can differ from each training pixel and still classify as foreground in the output image [23]. Adjustments of this threshold either expand or narrow the color space of foreground pixels relative to the previous threshold [23].

The overall accuracy of this CMEIAS color segmentation algorithm was >99% when tested at single pixel resolution on 26 complex images of colored microorganisms [23]. Its operation and educational scaffolding instruments provided with the released software were recently favorably reviewed by an independent group [24].

3. Application of These CMEIAS Algorithms for Analysis of *in Situ* Microbial Ecology

3.1. Sample Preparations

Microbial biofilms were colonized on plain glass or polystyrene slides, submerged for four days at one-foot depth in the Red Cedar River (mean temperature of 22 ± 2 °C) located in East Lansing, MI, USA [4]. Grayscale images of the biofilms were acquired using phase-contrast microscopy, then segmented and analyzed by CMEIAS to define the morphology, size and georeferenced location of cells at 0.2 μm resolution [1,4,25]. Statistics were performed using StatistiXL [26], PAST [27], EcoStat [28] and GS+ [29] software.

3.2. CMEIAS Image Analysis of the Microbial Biofilm Assemblages

3.2.1. Morphological Diversity

Diversity defines the community heterogeneity based on the richness and abundance among population classes [30]. This characteristic of a microbial “assemblage” (representative sample of the community) is intrinsically crucial to define its structure/function relationships at all biological organization levels [4]. Ecological theory predicts that biological differences between species, including their trophic level and niche requirements, inevitably create differences in species abundance [30]. Thus, class abundance measurements can help explain how the biological diversity of the assemblage was established [4,30].

Analysis of the richness and morphotaxa abundance can help define the diversity of niche apportionments in the polyphasic taxonomy analysis of microbial communities, particularly when integrated with 16S rDNA-based analyses [1,4,8,30–32]. Morphological information can also reveal *in situ* ecophysiological relationships between community membership and the environment, including rarity and dominance relationships reflecting functional stability, community resilience and ecological succession, indices of community health, allometric scaling of metabolic rates and biomass productivity, allocation and efficient utilization of limiting nutrient resources, and morphologically expressed adaptations after stress-induced perturbations (competition, predation, starvation, eutrophication, antibiosis, *etc.*) [1,4,30–39].

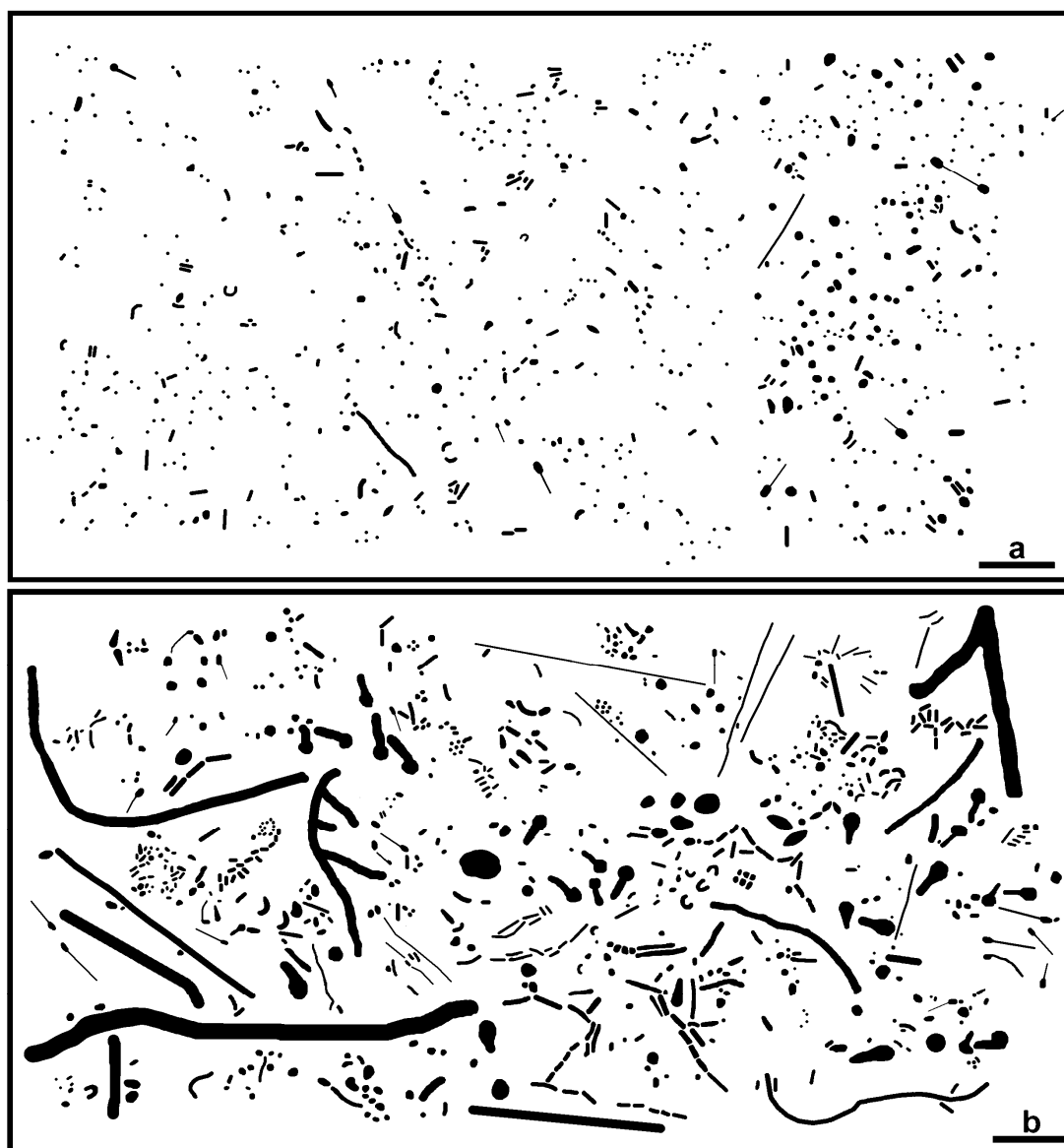


Figure 2. Microbial biofilm assemblages colonized on (a) control glass and (b) polystyrene slides incubated for four days in a flowing freshwater aquatic ecosystem. Bar scales equal 10.0 μm.

The two representative microbial assemblages (Figure 2a,b) differ in morphological diversity, growth physiology and biogeography, reflecting how substratum wettability/hydrophobicity influence

early development of freshwater biofilms before they become confluent. Although derived from the same bacterioplankton community at one-foot depth in the river, the differences in morphological diversity between these microbial biofilm assemblages reflect their enhanced fitness for success in the wettable borosilicate glass *vs.* hydrophobic polystyrene environments, in cell body mass relevant to their allometric scaling relationships and local apportionment of growth-supporting nutrients, and in their local spatial patterns of distribution that distinguish cooperative *vs.* competitive adaptations in their colonization behavior relevant to microbial biofilm ecology.

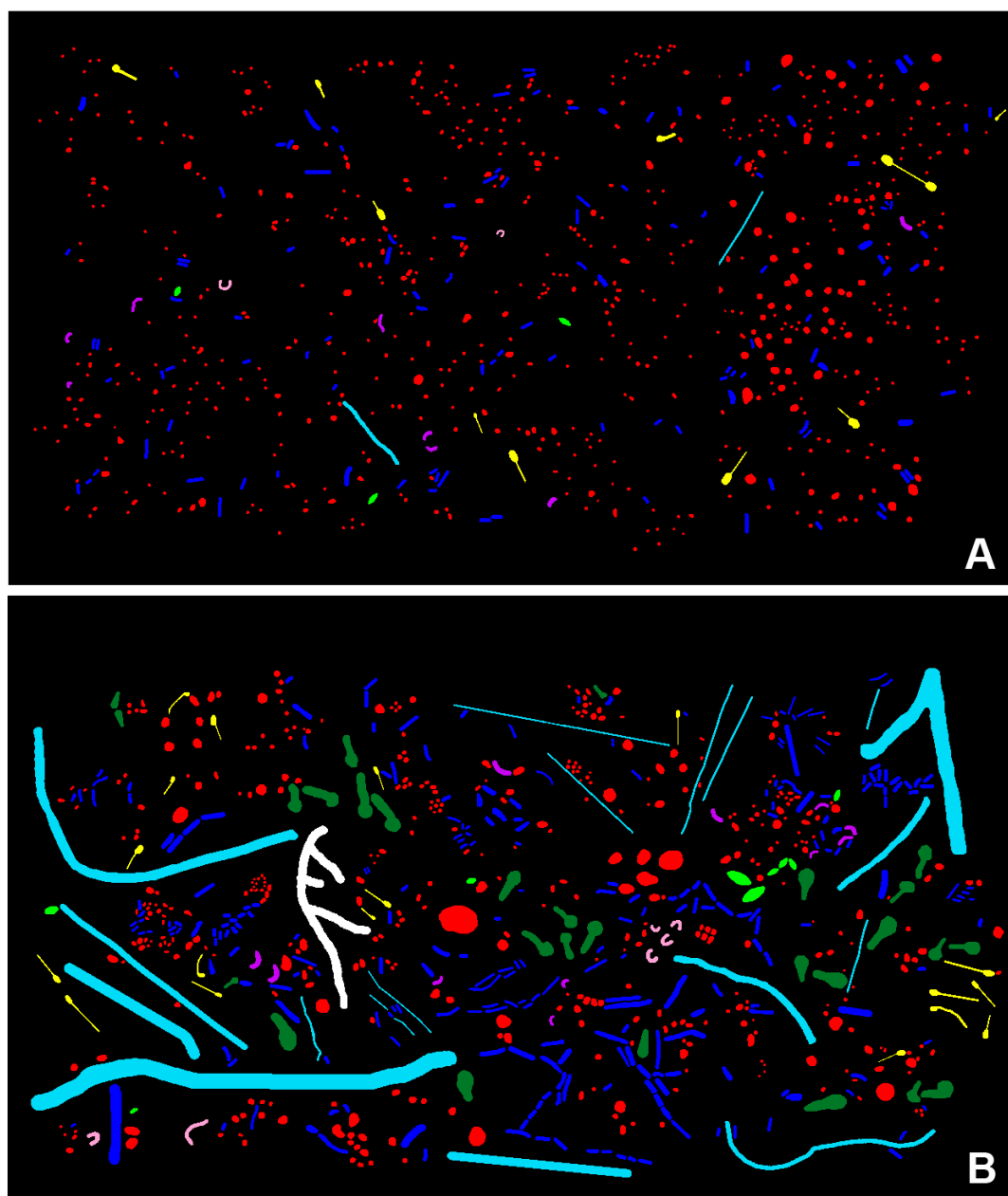


Figure 3. CMEIAS-rendered pseudocolor images showing the morphotype classification of each cell in the freshwater biofilm landscapes developed *in situ* on (a) control glass and (b) polystyrene.

Figure 3a,b is CMEIAS-rendered images of the same biofilm assemblages (Figure 2a,b) with each microbe pseudocolor-coded to indicate its assigned morphotype class. The results indicate that the

biofilm assemblages have seven morphotypes in common, including the numerically dominant regular rods (blue) and cocci (red), and fewer curved rods (purple), prosthecae (yellow), unbranched filaments (aqua), ellipsoids (true green) and U-shaped rods (pink). Additional comparison reveals that the club (olive green) and branched filament (white) morphotypes are only present in the biofilm assemblage developed on polystyrene.

Table 1 indicates CMEIAS data outputs of the richness and abundance for each class of morphotypes in these two biofilm assemblages, reported as differential cell counts and biovolume body mass. The eco-statistical indices computed from these data indicate the heterogeneity in diversity, dominance, and evenness in membership of community morphotaxa [4,25,27,28,30,33].

Table 1. CMEIAS (Center for Microbial Ecology Image Analysis System) analysis of morphological diversity in the river biofilm assemblages on plain glass and polystyrene using differential cell counts or cell biovolume as the metric of population abundance.

Morphotype/Eco-Statistic	Differential Cell Counts		Biovolume (μm^3)	
	Assemblage A (Plain Glass)	Assemblage B (Polystyrene)	Assemblage A (Plain Glass)	Assemblage B (Polystyrene)
Coccus	620	413	104	260
Curved Rod	8	12	4	8
U-shaped Rod	2	5	1	4
Regular Straight Rod	115	255	45	151
Unbranched Filament	2	19	4	568
Ellipsoid	3	9	2	15
Club	0	28	0	171
Prosthecae	10	18	7	7
Branched Filament	0	1	0	50
Totals	760	760	166	1235
Shannon Diversity	0.610	1.160 ^a	1.044	1.480 ^a
Simpson Diversity (1/D)	1.453	2.439 ^a	2.173	3.428 ^a
Berger–Parker’s Dominance	0.816 ^a	0.543	0.623 ^a	0.460
Smith and Wilson Evenness	0.192	0.484 ^a	0.395	0.560 ^a
Simpson Evenness	0.208	0.271 ^a	0.310	0.381 ^a
% Proportional Dissimilarity	27.24%		61.82%	
Bray-Curtis Distance	0.272		0.762	

^a Differences between community pairs are statistically significant at the 5% level [40].

Abundance of populations somewhat reveals their successful competition for limited resources [30], and therefore the metric used for class abundance in communities can influence how that relationship is interpreted. When scored by individual cell counts or cell biovolume, the abundance of morphotype classes in biofilm assemblage **B** were more evenly distributed (hence, have lower dominance), indicating that surface hydrophobicity supports the development of a more diverse microbial biofilm community than does the wettable surface of plain glass. The Simpson diversity index is more influenced by the most abundant class and therefore is higher than the Shannon diversity index [4,30]. Biovolume-weighted measures of abundance result in higher evenness indices since few larger morphotypes will counterbalance smaller, more numerically abundant morphotypes [4,15,30]. Conversely,

the Berger–Parker Dominance index is greater when individual counts are used to measure abundance because it reflects the proportional importance of the most abundant classes. The computed indices between assemblage pairs **A** vs. **B** are significantly different, based on the 10,000-iterated Solow test [40]. Thus, the morphological diversities in these assemblages are more evenly structured when individual biovolume body mass is used as the metric of abundance [4].

Ecological theory addresses important dominance-rarity community relationships and their stability relative to their colonized environment [30,35]. The occurrence of rare taxa indicates a time-resolved ecological succession in which later colonists have more specific requirements that are rarer [30,35,36,41]. The smaller niche apportionment for abundant classes (2/7 and 2/9 for assemblages **A** and **B**, respectively) compared to “rare” classes (<25% relative abundance) frequently occurs when the ecology of the assemblage is dominated by only a few factors [30]. A prominent implication is that “rarity” among class membership in microbial communities is “conditional” [42] without necessarily implying unimportance, since members of rare morphotypes can be very active and significantly contribute to community stability and resilience after environmental perturbation [4,31,32].

Comparison of the β -diversity indices of percent proportional dissimilarity and Berger–Parker distance also indicate that the continuous metric of morphotype-specific biovolume can distinguish the community diversity of these two biofilm assemblages better than does cell counts per morphotype class. These findings highlight the discriminating power of morphotype classification combined with biovolume measurements to analyze the diversity of microbial communities *in situ*.

3.2.2. Filamentous Microbial Morphotypes as Related to Refuge from Protozoan Bacteriivory

Predation is one of the primary determinants of the trophic structure and functioning of ecosystems [43]. This density-dependent activity is unevenly distributed across prey sizes. Size- and morphology-selective predation can affect the richness and abundance of community membership [44–46]. Body size significantly influences prey-selectivity, in relation to the strong evolutionary pressure of predators to maximize their intake of prey resources [33]. In various ecosystems, bacteriivory grazing activities by protozoan nanoflagellates and metazoan predators significantly influence bacterial community membership, mostly because resistance to and refuge from selective bacteriivory is favored by cells that form aggregates and filaments that are larger than the cytosome or lorica mouth opening of the predators, which limits their ability to consume the prey [44–46]. Thus, the proportional abundance of long microbial filaments can indicate the level of selective pressure that phagotrophic predatory stress contributes to shaping the aquatic microbial community [4]. Inspection of the biofilm assemblages (Figure 2a,b) and their pseudocolored morphotype classification (Figure 3a,b) indicate a greater abundance and body size of unbranched filaments (plus one branched filament) in the biofilm assemblage developed on polystyrene. Quantitative image analysis confirmed this differential classification result (Table 1), indicating that the unbranched filament morphotype has a 9.5-fold higher numerical abundance in the assemblage developed on polystyrene, and its biovolume dominates all other morphotypes in that assemblage (46% of total). The cumulative lengths of this selected morphotype were 25.3 vs. 469.7 μm for assemblages **A** and **B**, respectively, and an unpaired Student *t* test statistic indicated that their mean filament lengths (12.7 vs. 24.8 μm ; $df = 18$) were significantly different at $p = 0.006$. These results suggest that bacteriivory was more intense in the biofilm on polystyrene

within the freshwater aquatic ecosystem, and the increased fitness of the elongated filaments amidst the selective predatory stress caused its rise in dominance. Interestingly, the coccus was the only morphotype in the biofilm assemblage on polystyrene that had a lower relative numerical abundance (34% less than in assemblage A), suggesting that the predatory activity may have selectively targeted this morphotype of microbial prey.

3.2.3. Ecophysiological Attributes Linked to Accurate Measures of Biovolume Body Mass

In addition to its importance as a quantitative metric of abundance among morphotype classes, cell biovolume body size can provide insights on several other relevant ecophysiological functions and activities that drive community biodiversity [4,15,30,33,34,37,47]. First, it ranks individuals within the very wide range of body size (sometimes by several orders of magnitude) for the microbial component in ecosystems. Second, the principles of ecological theory amply recognize the importance of this body size metric, as the most abundant species tend to be the smallest in size and the least abundant species tend to be the largest in size. Third, this relationship between abundance and diversity tied to an organism's biomass provides an initial, best estimation of the concentration, apportionment, and acquisition of nutrient resources within the habitat [30]. Fourth, body size determines resource use, how the fractal heterogeneity in food cluster availability and concentration trade-off with body size, and how body sizes of some species hinder their ability to coexist with others [33]. Connected to this issue, the abundance of a community class in part reflects how well it succeeds in competition for limiting resources. Fifth, the distribution of individual body size of organisms is the strength behind the ecological theory of *allometric scaling* in all of biology, which seeks to explain how the biomass of populations in communities is fixed by representative metabolic activities and other significant ecosystem-level processes such as the abundance of incoming energy and nutrients that drive productivity [4,15,30,33,37,47,48]. Allometric scaling relates body size with the concentration and apportionment of nutrient resources in an environment and the consequential metabolic efficiency and growth rate of the organisms in the community [33,37,48].

Biovolume body size is also an important consideration in evaluations of the ecophysiological adaptations to starvation stress in microbial communities [4,15,30,33]. Morphological diversity is more valid in community analysis when cells are actively growing rather than in dormancy, since the latter is normally associated with dwarf cells that have sized down while adapting to nutrient limitations in the environment [4,38,39]. This is because distinctive cell morphologies reflect the shape-determining *murein sacculus* and other components involved in the cell division cycle that are mainly expressed only during active growth. When cells sense nutrient deprivation (e.g., stringent response), their starvation survival adaptations include, *inter alia*, reduction in body size by reductive division resulting in their increased biosurface area-to-volume ratio that improves their nutrient acquisition efficiency and distribution of those resources inside the cell. Optimal cell size is inversely related to resource molecular size when its uptake is limited by diffusive transport and uptake across membranes [34]. Different-sized consumers can coexist when competing for resources of different molecular sizes if the resources inputs and consumers sizes are correctly matched [34]. This information embraces the concept of *autochthonous k strategists* whose small body size and correspondingly increased surface area-to-volume ratio permit slow growth while adapting to resource scarcity and allow them to outcompete

others in environments with low nutrient concentrations [4]. In contrast, *zymogenous r strategists* are copiotrophic with limited carrying capacity and faster growth rates, enabling them to outcompete others when external concentrations of essential nutrients are high.

The spatial density of microbial biovolumes for the representative landscapes of community assemblages **A** and **B** were 15,612 and 119,666 μm^3 biomass/ mm^2 biofilm substratum, respectively. This 7.6-fold higher intensity of microbial biovolume is an indication of how the hydrophobic polystyrene substratum affects the growth physiology and allometric scaling of the biofilm in the river ecosystem. The total carrying capacity of biovolume body mass in the biofilm assemblage developed on polystyrene was higher for all morphotype classes except the prosthecate (Table 1). These data provide compelling evidence indicating that the polystyrene substratum concentrates a higher apportionment of nutrient resources in the freshwater aquatic ecosystem, allowing most cells in the assemblage **B** to grow to larger sizes consistent with a *zymogenous r*-type reproductive strategy. The smaller biovolume of microbes in the assemblage **A** suggests nutritional scarcity and reduced growth rates that trigger starvation stress adaptations, including body size reduction (Table 1) with significantly higher surface area-to-volume ratio (median cell SA/V of 11.43 vs. 9.15 $\mu\text{m}^2/\mu\text{m}^3$ for assemblages **A** and **B**, respectively; Mann–Whitney $Z = -13.38$, $p = 0.000$) that would increase their efficiency in nutrient uptake, consistent with an autochthonous *k*-type reproductive strategy.

3.2.4. Spatial Pattern Analysis and Its Relationship to Microbial Biofilm Ecology

Geospatial patterns of distribution among individuals profoundly affect many important ecological processes at the population, community and ecosystem levels [4,22,49]. Ecological processes change with the spatial scale at which they are observed and measured, and so examining how scale influences our observations and measurements is critical [49]. Many ecological processes (e.g., microbial interactions) are scale-dependent, and therefore their analysis should include a spatial component that defines the scale in which the process occurs [4,49]. Many interactions between individual microbes and their physical, chemical and biological environment occur at μm spatial scales, and therefore quantitative microscopy is the method of choice to explore activities involving cell interactions at this spatial scale. Understanding the significance of spatial scale *in situ* is important because it can explain some of the hyper-extreme diversity found in microbial communities. How community participants partition heterogeneous distributions of the same resource is an important trade-off constraint to explain how multiple species within the community (including microbial predators and prey) can occupy similar niche space [33]. These ecological traits lead to colonization patterns reflecting resource apportionment that allows coexistence of diverse species.

An important objective of spatial analysis is to define the distribution patterns at multiple scales and use these data to construct ecological models that can predict colonization behavior in that landscape [4,49]. Pattern analyses will accept the null hypothesis of *complete spatial randomness* when no interactions influence the resultant spatial colonization pattern [4]. In contrast, rejection of the null hypothesis of random chance indicates coordinated patterns of spatial distribution that signify ecophysiological adaptations in the spatially structured landscapes, and that regionalized interactions have influenced their colonization behavior resulting in the spatial pattern present [4,49].

Non-random spatial patterns also influence crucial events that control their ecophysiology and stability [4,50]. Microbial colonization in nutrient-rich microenvironments produces patterns of distribution that are typically *clustered*, commonly due to cooperative interactions among neighboring cells that promote their growth and flourish as aggregations of local microcolony biofilms [4,22,49,51,52]. An example of this positive relationship occurs in close cross-feeding interactions of syntrophic microbial aggregates that develop in complex methanogenic communities [32]. The scale-dependent heterogeneous fractal variability in partitioning of resources also produces clustered patterns of spatial distribution within landscapes [33]. Microbial colonization in areas of low nutrient availability can have intense competition, leading to poorly productive patterns of *uniformity* that imply negative interactions of conflict resulting in their over-dispersed and self-avoiding colonization behavior [4,22,49,52]. This information is important because spatial heterogeneity resulting from the non-random clustered and uniform patterns between individuals tends to stabilize ecological systems [4,50,51] and helps to explain the very high species diversity in microbial communities [33]. Spatial pattern analyses that reject the null hypothesis of complete spatial randomness are typically followed by other tests that determine whether the pattern is uniform or aggregated, its local vs. regional spatial scale, and its spatial intensity [4,51]. We have found that this trend of nonrandom patterns was dominated by cooperative clustering plus some competitive uniform distributions in microbial biofilms in natural and managed habitats, including on root and leaf surfaces [53–58], freshwater streambed pebbles [59], and in river/lake ecosystems [4,22,52,60]. Thus, spatial analyses of microbial biofilms can reveal many insights on the ecological conditions that trigger the colonization behaviors creating them [4].

The CMEIAS module for spatial pattern assessment includes a variety of measurement features to perform point-pattern, quadrat/lattice and geospatial statistical analyses. The *X*, *Y* Cartesian coordinates of object centroids (means of the *x* and *y* coordinates of every pixel in the object), the first and second nearest neighbor distances, and the empirical distribution function of the first nearest neighbor distance support the point-pattern analysis. The quadrat-lattice analysis module is supported by several landscape ecology, spatial abundance, and point to nearest object metrics, plus CMEIAS Quadrat Maker software [61]. This latter application uniquely transforms the input landscape image into an annotated, color index derivative with optimized grid overlay and column-row labeling of individual quadrats, cuts a copy of the landscape image into quadrats defined by the optimized grid raster, and then saves each individual quadrat image with a file name indicating its unique location within the landscape domain, now ready for stack building and automated image analysis [61].

Several point-pattern spatial aggregation tests are available to obtain a global perspective of the overall spatial pattern of the cells in microbial biofilm landscapes. The Holgate analysis [28,62] rejected the null hypothesis of complete spatial randomness for both biofilm landscapes ($p = 0.001$) and indicated a greater intensity of spatial aggregation for the polystyrene-based biofilm (Holgate's $A = 0.607$, compared to 0.571 for the biofilm developed on control glass). The Ripley K point pattern analysis [27,63] confirmed that clustered aggregation dominated the spatial patterns of cells in both biofilm landscapes, and this pattern was more intense on the polystyrene-based biofilm (maximal and cumulative Ripley's $L(d)$ - d values of 1.30 and 86.76, compared to 0.75 and 61.02 for the biofilm landscape developed on plain glass). The Ripley K analysis also indicated minor patches of uniform spatial patterns among cells in both biofilms (cumulative negative $L(d)$ - d values of -0.67 on plain glass vs. -0.25 on polystyrene), implicating some negative cell-cell interactions that are more intense

in the biofilm developed on plain glass, consistent with their smaller biovolume body mass (Table 1). These differences in spatial distribution indicated by the point-pattern analyses of these two biofilm landscapes set the stage for a more comprehensive spatial analysis of their colonization behavior.

Geostatistics provides a robust method of pattern analysis that examines the dependency among georeferenced points in a landscape in order to examine the continuity of spatial patterns over that entire spatial domain [20]. Its core analysis tests whether a user-defined, continuously distributed regionalized variable, called the “*Z*-variate”, is *spatial autocorrelated*, i.e., has spatial structure [20]. The geostatistical result indicates the irregularity among the *Z*-variate values at each sampling point and the resemblance of that regionalized variable between cell pairs in relation to their separation distance [20]. It reveals whether the *Z*-variate values among neighbors at one location influence that same variable for cell pairs at more distant locations, and also indicates the maximal radial dimension of that spatial correlation [20]. The geospatial patterns are autocorrelated when the regionalized *Z*-variates of close neighbors (e.g., aggregation) are more alike (as in aggregated distributions) than when located further apart [64], indicating that the pattern is controlled by a spatially explicit process rather than occurring randomly and independent of location [4,20]. When the regionalized variable is found to be strongly autocorrelated, its geospatial data can be accurately modeled to create interpolated kriging maps of its intensity at all locations, including those unmeasured in that spatial domain [20].

The main CMEIAS metric used as the *Z*-variate for geospatial autocorrelation of individual cell aggregations in biofilm landscapes is the Cluster Index (inverse of the 1st nearest neighbor distance) [21]. Table 2 summarizes data on this measured index of cell-aggregated patterns in the two biofilm landscapes. Its distribution lacked normality for both biofilms, and the range, median, mode and maximum were higher for cells in the biofilm landscape developed on polystyrene. The Mann–Whitney test confirmed that the distribution of cluster index values was significantly higher for cells in that biofilm landscape.

Table 2. Distribution of the CMEIAS cluster index associated with individual cells in river biofilm landscapes developed on plain glass and polystyrene substrata.

Distribution Statistic	Plain Glass	Polystyrene	Interpretation of Aggregation
Range	1.669	4.767	↑ Polystyrene > Control Glass
Median	0.636	0.725	↑ Polystyrene > Control Glass
Mode	0.976	1.574	↑ Polystyrene > Control Glass
Maximum	1.855	4.881	↑ Polystyrene > Control Glass
Shapiro–Wilk Normality Test <i>W</i> (<i>p</i>)	0.935 (0)	0.849 (0)	Distribution Not Normal
Mann–Whitney Median Test <i>U</i> (<i>p</i>)	330,216 (1.30×10^{-6})		↑ Polystyrene > Control Glass

The geostatistical semivariograms of the cluster index for all cells within the two biofilm landscapes are presented in Figure 4a,b. These plots relate the uncertainty of the regionalized variable under investigation with the distances that this spatial property autocorrelates at measured points in the landscape [4,20]. Table 3 describes four significant geostatistical parameters that are computed from the data to produce these different semivariograms of the autocorrelated cluster index for the biofilm landscapes developed on plan glass vs. polystyrene substrata.

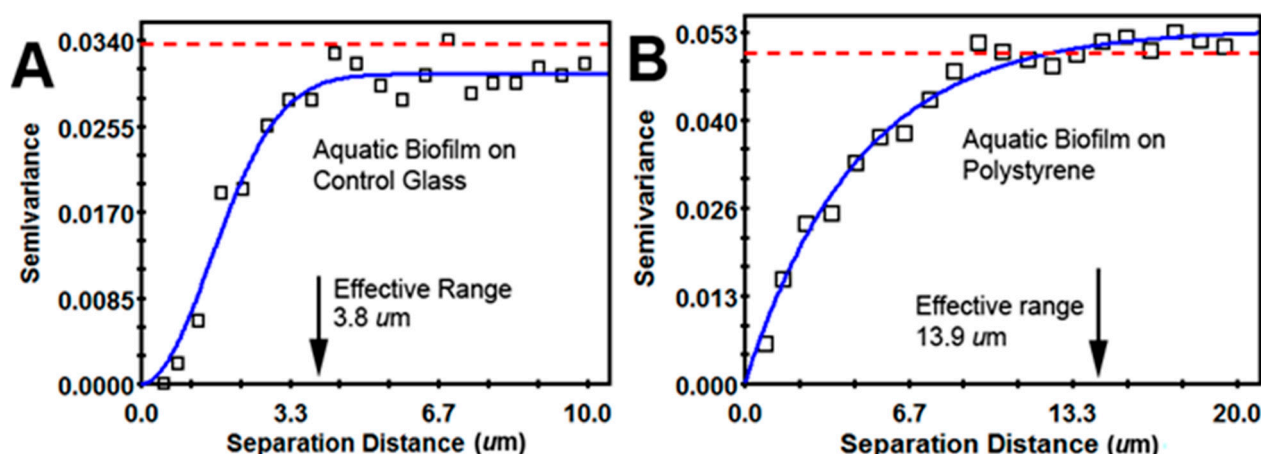


Figure 4. Geostatistical isotropic semivariograms of autocorrelated, clustered colonization behavior of the microbes in the river biofilm landscapes developed on (a) control glass and (b) polystyrene.

Table 3. Geostatistical analysis of the spatial autocorrelation of clustered colonization behavior among individual cells in the aquatic biofilm landscapes on plain glass and polystyrene substrata.

Geostatistical Parameter	Plain Glass	Polystyrene	Interpretation
Residual \sum Squares of the Best-Fit Model	6.85×10^{-5}	6.79×10^{-5}	Strong model fit to semivariance
Nugget Variance	0.00001	0.00001	Sufficient sampling at the proper spatial scale for both biofilm
Moran's I Cluster Index Strength of Autocorrelation	+4.24	+12.06	<p>↑ Polystyrene biofilm has 2.84-fold stronger autocorrelation</p> <p>↑ Polystyrene biofilm 3.66-fold wider radius of influence on clustered colonization behavior between neighboring cells</p>
Effective range of autocorrelated radial separation distances between cells (μm)	3.8	13.9	

The best-fit mathematical models derived from these plots indicate that the spatial cluster indices are isotropic and significantly autocorrelated (Figure 4a,b, Table 3). The low nugget variances (at the Y intercepts) indicate minimal data discontinuity, *i.e.*, the amount of geospatial microstructure of the *Z*-variate is sufficiently represented by the sample size and is measured at the proper spatial scale [20]. The cumulative *Moran's Index* reports the strength of autocorrelated spatial dependence in the cluster index relationships between paired observations in the domain [4,20]. When positive, the cumulative Moran's Index indicates that the clustered colonization behavior has significantly more spatial structure than would occur if the underlying spatial process were random in these two biofilm landscapes. From the microbial ecology perspective, this result implies active cell-cell interactions that promote their aggregated spatial patterns [4]. This cooperative colonization behavior of the microbes is significantly more intense in the biofilm developed on the polystyrene substratum. The *effective range of separation distance* between autocorrelated sampling points provides a profoundly important estimate of the greatest distance that cells within the landscape can influence their neighbor's

aggregation behavior, enabling them to develop microcolony biofilms [4]. That 3.66-fold wider radial distance defines a circle of influence that covers a 13.38-fold larger substratum area for cells in the biofilm colonized on polystyrene. Examples of ecophysiological processes that can cause this positive aggregation behavior include syntrophic cross-feeding, secretion of signals that promote microcolony development, localized degradation of extracellular metabolic wastes, and production of a protective biofilm matrix of extracellular polymers that sequester limiting nutrients, stabilize external transforming DNA, provide refuge from predator bacteriovory, and restrict diffusion of antimicrobials [4]. These are colossal benefits for microbial lifestyles within biofilms!

Figure 5A,B show the two-dimensional kriging maps of the spatial variability in cluster index that is built after identifying the best-fit geostatistical autocorrelation model. Its unique benefit is the ability to provide statistically defendable interpolation of continuous predicted values of the *Z*-variate within the landscape, even where it has not been measured [20,29]. The kriging map includes *isopleth lines* whose curvature connects points of equal cluster index values [20]. The resultant map of these isopleth contours and pseudocolored bin scale of the regionalized *Z*-variate reveal the relative strength of intensity gradients of neighboring cell-cell interactions that promote aggregated colonization behavior within spatially defined clusters, and estimates that parameter at all landscape locations. These figures provide compelling evidence that the geospatially autocorrelated clustering behavior was stronger and more abundant when the microbes colonized the polystyrene substratum.

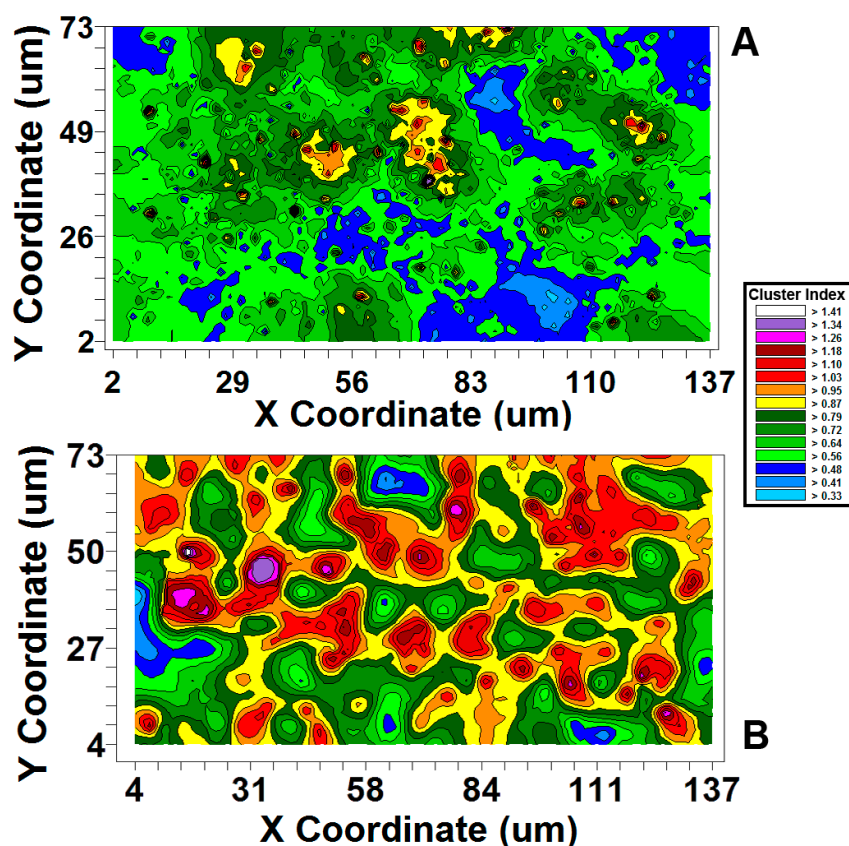


Figure 5. Kriging maps of the spatially autocorrelated intensity of aggregated colonization behavior of the microbial assemblages in the freshwater biofilm landscapes developed on (a) plain glass and (b) polystyrene. The legend shows the stepped scale of pseudocolored bins for the range of CMEIAS cluster index values in each landscape.

3.2.5. Color Segmentation Tool for Cell-Cell Communication/Gene Expression Studies

An interesting application of color-discriminated microbial ecophysiology utilizes engineered strains to report cell-cell communication that triggers the expression of specific genes [4,65]. This process occurs when external signal molecules accumulate sufficiently to affect neighboring cells [4,22,65]. Earlier studies [65] using laser scanning confocal fluorescence microscopy and CMEIAS image analysis indicated that two individual cells (one a producer and the other a sensor) are sufficient to fulfill the quorum requirement for *N*-acylhomoserine lactone (AHL)-negotiated communication while they colonize plant roots. The “calling distance” between pairs of communicating cells during colonization had a mode of 4–5 μm and extended out to a maximum of 78 μm [65] (equivalent to two people conversing while at opposite ends of a soccer field). Further geospatial analysis indicated that this process of cell-cell communication is controlled more by cell location within diffusion gradients of the signal molecules made by their neighbors than by a quorum requirement of high population density [65]. Thus, bacterial communication is more common than previously anticipated by the dogma that it only occurs at very high population densities [4,65]. Spatial scale and location really matter!

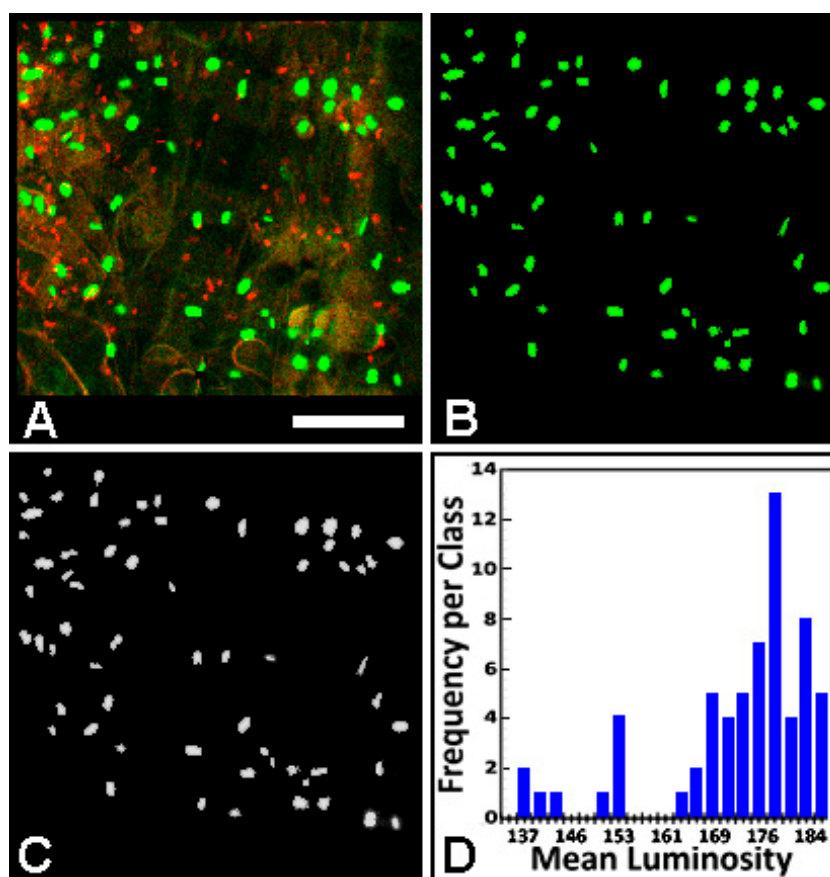


Figure 6. *In situ* variation in intensity of cell communication between *Pseudomonas putida* reporter cells colonized on a wheat rhizoplane. (a) Red-fluorescent “signal source” cells communicating with green-fluorescent “signal sensor” cells via secreted *N*-acylhomoserine lactones; (b) Green color-segmented and (c) grayscale converted images of the communication response in Figure 6a; and (d) Frequency distribution of the mean luminosity of Gfp fluorescence per sensor cell indicating the *in situ* variation in cell communication intensity at single-cell resolution. Bar scale is 15 μm .

In images for that study [65], the autofluorescent root substratum was represented by background pixels with the same color as the fluorescent reporter bacteria but at less luminosity and density, making it difficult to segment the foreground cells before image analysis [4]. Figure 6a is a fluorescent micrograph from the same study [65] showing a wheat rhizoplane landscape being colonized by two reporter strains of *Pseudomonas putida*, including red-fluorescent cells of the “AHL-signal source” strain (constitutively expressed *rfp* derivative of wild type IsoF) and cells of the “AHL-signal sensor” strain (with a mutation blocking its ability to synthesize AHLs and containing a reporter plasmid with a Gfp-promoter fusion in its AHL-regulated promoter) that fluoresce green (expressing the green fluorescent protein, Gfp) when activated by sufficient extracellular AHL signal. With optimal setting of the color tolerance variable, the CMEIAS Color Segmentation software application [23] was able to accurately eliminate the red and green background pixels, and segmented the 63 green sensor bacteria expressing Gfp in response to the AHL signal source (Figure 6b), producing the processed image (Figure 6c) that was analyzed to quantify the “soft whisper to loud shouting” variation in signal communication intensity at single-cell resolution (Figure 6d).

4. Summary and Concluding Statements

This review describes several unique algorithms and computational features of CMEIAS image analysis software, followed by examples of their application in ecological studies of microorganisms during their colonization of (non)biological surfaces. The analyses are intended to reveal various insights on the ecophysiology of individual microbial cells within biofilms at spatial scales relevant to the physical, chemical and biological dimensions of their ecological niches *in situ*. In the first example, we describe the algorithm to compute cell length, which is adaptable to the shape of the cell being measured. This adaptability feature is especially important to increase the accuracy of measuring cell length when its medial axis is irregular, branched and/or has curvature. An example of its use is to analyze the increase in length and relative abundance of elongated cells within aquatic microbial communities as an indicator of the intensity of protozoan bacterivory and bacterial adaptations to avoid that biological stress. In the second example, we describe CMEIAS’s most accurate algorithms to compute cell biovolume. These algorithms also utilize shape-adaptable formulas to achieve very high accuracy in computing cell biovolume for a wide range of microbial morphotypes in two-dimensional images. Important examples of their use are to study the connections between individual body size and niche apportionment enabling microorganisms to cohabitate with each other within the same biofilm landscape, as well as an organism’s adaptive response to nutritional enrichment or starvation conditions. In the third example, we describe the unique, mathematically-based algorithms of pattern recognition featured in CMEIAS to accurately classify all of the morphotypes in the microbial community under investigation, and examples that use this hierarchical supervised classifier to measure the impact that environmental perturbation (in this case, increased substratum hydrophobicity) has on community diversity, evenness, dominance and rarity relationships among morphotaxa within natural aquatic biofilms and their enhanced fitness for success in the particular habitat. In the fourth example, we describe some of the unique CMEIAS spatial measurement attributes and computing tools that allow for unprecedented analyses of the *in situ* biogeography of microbes and their colonization behavior within biofilm communities. The data of these CMEIAS

attributes extracted from images can be evaluated using point-pattern, quadrat-lattice and geostatistical methods that are designed to assess the degree of coaggregation or overdispersed uniformity indicating cooperative *versus* competitive colonization behaviors that are neither random nor can be explained by chance, and quantify the real-world spatial scale and local intensity of their autocorrelated cell-cell interactions. In the fifth and final example, we describe a stand-alone CMEIAS software system with unique algorithms designed to accurately segment foreground objects in complex color images when the RGB ranges of their pixels overlap those of the background. We then show an example that uses this color segmentation tool to process images of bacterial reporter/sensor strains colonized on plant roots so the variation in intensity of their cell-cell communication can be accurately measured *in situ*.

The algorithms to automatically measure the roundness-adaptive length and width metrics needed to compute cell biovolume, the various pattern recognition computations of the morphotype classifier, the *X*, *Y* Cartesian coordinates to analyze the spatial relationships of neighboring cells, and the mean luminosity metric to analyze the intensity of gene expression are already featured in the CMEIAS version 1.28 currently available at the project website [5]. Also, the CMEIAS Color Segmentation software available at the website can perform all of the color segmentation algorithms described here. The next release version of CMEIAS-IT image analysis software will include the automatic calculation of cell biovolume, the first and second nearest neighbor distances, the local spatial cluster index and many other new image processing and analysis features [25]. All of these unique algorithms incorporated into CMEIAS image analysis software can be used to strengthen microscopy-based approaches for understanding the ecophysiology of microorganisms at single-cell resolution and the *in situ* spatial scales at which they are important.

Acknowledgments

Development of CMEIAS image analysis software and its applications to quantify microbial colonization of roots and other biofilms has been supported by the US-Egypt Science & Technology Joint fund project 58-3148-1-140, the Michigan State University Center for Microbial Ecology, the NSF Long-Term Ecological Research program at the MSU Kellogg Biological Station, and Michigan AgBioResearch. We thank Stephan Gantner for providing images of *Pseudomonas putida* sensor strains reporting cell-to-cell communication on plant roots, and the CMEIAS team [25] for their assistance in development of this software.

Author Contributions

Preparation of this review document was a combined effort of both authors. BCN prepared the section on the mathematical rules of pattern recognition, most of the conclusion, and the reference list. FBD computed the data, prepared the figures and wrote the remaining sections of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest. When fully tested and documented, copyrighted CMEIAS software components are freely available at <http://cme.msu.edu/cmeias/>.

References

1. Liu, J.; Dazzo, F.B.; Glagoleva, O.; Yu, B.; Jain, A. CMEIAS: A computer-aided system for the image analysis of bacterial morphotypes in microbial communities. *Microb. Ecol.* **2001**, *41*, 173–194.
2. Daims, H.; Lucker, S.; Wagner, M. Daime, a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* **2006**, *8*, 200–213.
3. Daims, H.; Wagner, M. Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 237–248.
4. Dazzo, F.B.; Klemmer, K.J.; Chandler, R.; Yanni, Y. *In situ* ecophysiology of microbial biofilm communities analyzed by CMEIAS computer-assisted microscopy at single-cell resolution. *Diversity* **2013**, *5*, 426–460.
5. CMEIAS Center for Microbial Ecology Image Analysis System Project Website. Available online: <http://cme.msu.edu/cmeias/> (accessed on 22 April 2013).
6. Dazzo, F.B.; Mateos, P.; Orgambide, G.; Philip-Hollingsworth, S.; Squartini, A.; Subba-Rao, N.S.; Pankratz, H.S.; Baker, D.; Hollingsworth, R.; Whallon, J.; *et al.* The infection process in the Rhizobium-legume symbiosis and visualization of rhizoplane microorganisms by laser scanning confocal microscopy. In *Trends in Microbial Ecology*; Guerrero, R., Pedros-Alio, C., Eds.; Spanish Society for Microbiology: Barcelona, Spain, 1993; pp. 259–262.
7. Schlöter, M.; Borlinghaus, R.; Bode, W.; Hartmann, A. Direct identification and localization of *Azospirillum* in the rhizosphere of wheat using fluorescence-labeled monoclonal antibodies and confocal scanning laser microscopy. *J. Microsc.* **1993**, *171*, 173–177.
8. Assmus, B.; Hutzler, P.; Kirchhof, G.; Amann, R.; Lawrence, J.R.; Hartmann, A. *In situ* localization of *Azospirillum brasilense* in the rhizosphere of wheat with fluorescently labeled, rRNA-targeted oligonucleotide probes and scanning confocal laser microscopy. *Appl. Environ. Microbiol.* **1995**, *61*, 1013–1019.
9. Chi, F.; Shen, S.H.; Cheng, H.P.; Jing, Y.X.; Yanni, Y.G.; Dazzo, F.B. Ascending migration of endophytic rhizobia, from roots to leaves, inside rice plants and assessment of benefits to rice growth physiology. *Appl. Environ. Microbiol.* **2005**, *71*, 7271–7278.
10. Lawrence, R.; Korber, D.R.; Neu, T.R. Analytical imaging and microscopy techniques. In *Manual of Environmental Microbiology*, 3rd ed.; Hurst, C.J., Crawford, R.L., Garland, J., Lipson, D., Mills, A., Stetzenback, L.J., Eds.; American Society for Microbiology Press: Washington, DC, USA, 2007; pp. 40–68.
11. Lewandowski, Z.; Beyenal, H. *Fundamentals of Biofilm Research*; CRC: Boca Raton, FL, USA, 2007.
12. Krambeck, C.; Krambeck, H.; Overbeck, J. Microcomputer assisted biomass determination of planktonic bacteria on scanning electron micrographs. *Appl. Environ. Microbiol.* **1981**, *42*, 142–149.
13. Sieracki, M.; Viles, C. Enumeration and sizing of microorganisms using digital image analysis. In *Digital Image Analysis of Microbes: Imaging, Morphometry, Fluorometry and Motility Techniques and Applications*; Wilkinson, M.H.F., Schut, F., Eds.; Wiley: Chichester, UK, 1998; pp. 175–198.

14. Sieracki, M.; Reichenbach, S.; Webb, K. Evaluation of automated threshold selection methods for accurately sizing microscopic fluorescent cells by image analysis. *Appl. Environ. Microbiol.* **1989**, *55*, 2762–2772.
15. Folland, I.; Trione, D.; Dazzo, F.B. Accuracy of biovolume formulas for CMEIAS computer-assisted microscopy and body size analysis of morphologically diverse microbial populations and communities. *Microb. Ecol.* **2013**, *68*, 596–610.
16. Duda, R.; Hart, P. *Pattern Classification and Scene Analysis*; John Wiley & Sons: New York, NY, USA, 1973; pp. 103–132.
17. Jain, A.K.; Zongker, D. Feature selection: Evaluation, application, and small sample performance. *Inst. Electr. Electron. Eng. Trans. Pattern Anal. Mach. Intell.* **1993**, *19*, 153–158.
18. Rosin, P.; Venkatesh, S. Extracting natural scales using Fourier descriptors. *Pattern Recognit.* **1993**, *26*, 1383–1393.
19. Yu, B.; Yuan, B. A consistent attributed graph-based hand-drawn circuit diagram reading system. *Chin. J. Electron.* **1995**, *4*, 1–11.
20. Robertson, G.P. Geostatistics in ecology: Interpolating with known variance. *Ecology* **1987**, *68*, 744–748.
21. Dazzo, F.B.; Joseph, A.R.; Goma, A.B.; Yanni, Y.G.; Robertson, G.P. Quantitative indices for the autecological biogeography of a *Rhizobium* endophyte of rice at macro and micro spatial scales. *Symbiosis* **2003**, *35*, 147–158.
22. Dazzo, F.B. CMEIAS-aided microscopy of the spatial ecology of individual bacterial interactions involving cell-to-cell communication within biofilms. *Sensors* **2012**, *12*, 7047–7062.
23. Gross, C.A.; Reddy, C.; Dazzo, F.B. CMEIAS color segmentation: An improved computing technology to process color images for quantitative microbial ecology studies at single-cell resolution. *Microb. Ecol.* **2010**, *59*, 400–414.
24. Wiesmann, V.; Franz, D.; Held, C.; Munzenmayer, C.; Palmisano, R.; Wittenberg, T. Review of free software tools for image analysis of fluorescence cell micrographs. *J. Microsc.* **2014**, *257*, doi:10.1111/jmi.12184.
25. Dazzo, F.B.; Liu, J.; Gross, C.; Reddy, C.; Monosmith, C.; Philips, N.; Radek, C.; Klemmer, K.J.; Ji, Z.; Smith, P.; *et al.* CMEIAS v3.10: Advanced computational tools of image analysis software designed to strengthen microscopy-based approaches for understanding microbial ecology. In Proceedings of the 2013 All Scientists Meeting, Kellogg Biological Station Long-Term Ecological Research Program, Michigan State University, East Lansing, MI, USA, 2013; Available online: <http://lter.kbs.msu.edu/abstracts/434> (accessed on 2 March 2015).
26. Roberts, A.; Withers, P. *StatistiXL, Version 1.10*; Broadway-Nedlands: Kalamunda, Australia, 2012.
27. Hammer, O.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 1–9.
28. Towner, H. *EcoStat, Version 1.03*; Trinity Software Inc.: Campton, NH, USA, 1999.
29. *GS+*, *Version 7.0*; Gamma Design Software: Plainwell, MI, USA, 2004.
30. Magurran, A.E. *Measuring Biological Diversity*; Blackwell Publishing: Malden, MA, USA, 2004.
31. Hashsham, S.; Fernandez, A.; Dollhopf, S.; Dazzo, F.B.; Hickey, R.; Tiedje, J.M.; Criddle, C.S. Parallel processing of substrate correlates with greater functional stability in methanogenic bioreactor communities perturbed by glucose. *Appl. Environ. Microbiol.* **2000**, *66*, 4050–4057.

32. Fernandez, A.; Hashsham, S.; Dollhopf, D.; Raskin, L.; Glagoleva, O.; Dazzo, F.B.; Hickey, R.; Tiedje, J.M.; Criddle, C.S. Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. *Appl. Environ. Microbiol.* **2000**, *66*, 4058–4067.
33. Ritchie, M.E. *Scale, Heterogeneity, and the Structure and Diversity of Ecological Communities*; Princeton University Press: Princeton, NJ, USA, 2010.
34. Yoshiyama, K.; Klausmeier, C.A. Optimal cell size for resource uptake in fluids: A new facet of resource competition. *Am. Nat.* **2008**, *171*, 59–70.
35. Lambshead, P.J.D.; Platt, H.M.; Shaw, K.M. The detection of differences among assemblages of marine benthic species based on an assessment of dominance and diversity. *J. Nat. Hist.* **1983**, *17*, 859–874.
36. Gaston, K.J. *Rarity*; Springer: Danvers, MA, USA, 1994.
37. West, G.B.; Woodruff, W.H.; Brown, J.H. Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2473–2478.
38. Baker, R.M.; Singleton, F.L.; Hood, M.A. Effects of nutrient deprivation on *Vibrio cholerae*. *Appl. Environ. Microbiol.* **1983**, *46*, 930–940.
39. Holmquist, L.; Kjelleberg, S. Changes in viability, respiratory activity, and morphology of the marine *Vibrio* sp. strain S14 during starvation of individual nutrients and subsequent recovery. *FEMS Microbiol. Ecol.* **1993**, *12*, 215–224.
40. Solow, A.R. A simple test for change in community structure. *J. Anim. Ecol.* **1993**, *2*, 191–193.
41. Tilman, D.; Knops, J.; Wedin, D.; Reich, P.; Ritchie, M.; Siemann, E. The influence of functional diversity and composition on ecosystem processes. *Science* **1997**, *277*, 1300–1302.
42. Shade, A.; Hogan, S.; Klimowicz, A.K.; Linske, M.; McManus, P.S.; Handelsman, J. Culturing captures members of the soil rare biosphere. *Environ. Microbiol.* **2012**, *14*, 2247–2252.
43. Schmitz, O. Predator diversity and trophic interactions. *Ecology* **2007**, *88*, 2415–2426.
44. Jürgens, K.; Arndt, H.; Zimmermann, H. Impact of metazoan and protozoan grazers on bacterial biomass distribution in microcosm experiments. *Aquat. Microb. Ecol.* **1997**, *12*, 131–138.
45. Hahn, M.W.; Moore, E.R.; Hofle, M.G. Bacterial filament formation, a defense mechanism against flagellate grazing, is growth rate controlled in bacteria of different phyla. *Appl. Environ. Microbiol.* **1999**, *65*, 25–35.
46. Gonzalez, J.M.; Sherr, E.B.; Sherr, B.F. Size-selective grazing on bacteria by natural assemblages of estuarine flagellates and ciliates. *Appl. Environ. Microbiol.* **1990**, *56*, 583–589.
47. Petchey, O.I.; Long, Z.T.; Morin, P.J. The consequences of body size in model microbial ecosystems. In *Body Size: The Structure and Function of Aquatic Ecosystems*; Hildrew, A., Raffaelli, D., Edmonds-Brown, R., Eds.; Cambridge University Press: Cambridge, UK, 2007; pp. 245–265.
48. Brown, J.H. Toward a metabolic theory of ecology. *Ecology* **2004**, *45*, 1771–1789.
49. Turner, M.G. Landscape ecology: The effect of pattern on process. *Annu. Rev. Ecol. Syst.* **1989**, *20*, 171–197.
50. Hassell, M.P.; May, R.M. Stability in insect host-parasite models. *J. Anim. Ecol.* **1973**, *42*, 693–726.
51. Perry, J. Measures of spatial patterns for counts. *Ecology* **1998**, *79*, 1008–1017.

52. Dazzo, F.B. CMEIAS digital microscopy and quantitative image analysis of microorganisms. In *Microscopy: Science, Technology, Applications and Education*; Microscopy Book Series #4; Mendez-Vilas, A., Diaz, J., Eds.; Formatex Res Center: Badajoz, Spain, 2010; pp.1083–1090.
53. Yanni, Y.G.; Rizk, R.Y.; Abd El-Fattah, F.K.; Squartini, A.; Corich, V.; Giacomini, A.; de Bruijn, F.; Rademaker, J.; Maya-Flores, J.; Ostrom, P.; *et al.* The beneficial plant growth-promoting association of *Rhizobium leguminosarum* bv. trifolii with rice roots. *Aust. J. Plant Physiol.* **2001**, *28*, 845–870.
54. Dazzo, F.B. Applications of quantitative microscopy in studies of plant surface microbiology. In *Plant Surface Microbiology*; Varma, A., Abbott, L., Werner, D., Hampp, R., Eds.; Springer-Verlag: Berlin, Germany, 2004; pp. 503–550.
55. Dazzo, F.B.; Schmid, M.; Hartmann, A. Immunofluorescence microscopy and fluorescence *in situ* hybridization combined with CMEIAS and other image analysis tools for soil- and plant-associated microbial autecology. In *Manual of Environmental Microbiology*, 3rd ed.; Garland, J., Hurst, C., Lipson, D., Mills, A., Stetzenbach, L., Crawford, R., Eds.; American Society for Microbiology Press: Washington, DC, USA, 2007; Chapter 59, pp. 712–733.
56. Dazzo, F.B.; Yanni, Y.G. CMEIAS: An improved computing technology for quantitative image analysis of root colonization by rhizobacteria *in situ* at single-cell resolution. In *Molecular Microbial Ecology of the Rhizosphere*; DeBruijn, F., Ed.; John Wiley & Sons: New York, NY, USA, 2013; Volume 2, Chapter 69, pp. 733–742.
57. Dazzo, F.B.; Gantner, S. Rhizosphere. In *Topics in Ecological and Environmental Microbiology*; Schmidt, T., Schaechter, M., Eds.; Elsevier Press: Waltham, MA, USA, 2012; pp. 467–479.
58. Dazzo, F.B.; Ganesan, I.; Yanni, Y. Spatial ecology of rhizobacterial colonization on roots analyzed by CMEIAS computer-assisted microscopy at single-cell resolution. In *Trends in Soil Ecology*; Singh, D.P., Singh, H., Eds.; Studium Press: Houston, TX, USA, 2014; pp. 67–92.
59. Fukuda, M.; Matsuyama, J.; Katano, T.; Nakano, S.; Dazzo, F.B. Assessing primary and bacterial production rates in biofilms on pebbles in Ishite Stream, Japan. *Microb. Ecol.* **2006**, *52*, 1–9.
60. Ji, Z.; Card, K.; Dazzo, F.B. CMEIAS JFrad: A new computing toolkit to discriminate the fractal geometry of landscape architecture and spatial patterns of individual cells in microbial biofilms. *Microb. Ecol.* **2014**, doi:10.1007/s00248-014-0495-1.
61. Dazzo, F.B.; Gross, C.A. CMEIAS Quadrat Maker: A digital software tool to optimize grid dimensions and produce quadrat images for landscape ecology spatial analysis. *J. Ecosyst. Ecogr.* **2013**, *3*, 1–4.
62. Holgate, P. Some new tests of randomness. *J. Ecol.* **1965**, *53*, 261–266.
63. Ripley, B.D. Tests of “randomness” for spatial point patterns. *J. R. Stat. Soc.* **1979**, *41*, 368–374.
64. Ettema, C.H.; Wardle, D.H. Spatial soil ecology. *Trends Ecol. Evol.* **2002**, *17*, 117–183.
65. Gantner, S.; Schmid, M.; Dürr, C.; Schuegger, R.; Steidle, A.; Hutzler, P.R.; Langebartels, C.; Eberl, L.; Hartmann, A.; Dazzo, F.B.; *et al.* *In situ* spatial scale of calling distances and population density-independent *N*-acylhomoserine lactone mediated communication by rhizobacteria colonized on plant roots. *FEMS Microb. Ecol.* **2006**, *56*, 188–194.