Quantitative Inference of Bacterial Motility Behavior

Xiaomeng Liang*, Lin-Ching Chang, Arash Massoudieh
School of Engineering
The Catholic University of America
Washington, DC, 20064, USA
74liang@cardinalmail.cua.edu

Nanxi Lu, Thanh H. Nguyen
Dept. of Civil and Environmental Engineering
Univ. of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

***Abstract*— This paper proposes a framework for automatic bacteria motile trajectories detection and motility behavior clustering. The input data is a sequence of images which contains bacteria motility information. The traditional experimental methods to identify the trajectories, and segment them to "run" and "tumble" mode are time consuming and subjective. The proposed method processes bacteria motility movies and extracts statistical features of runs and tumbles which drastically saves time, human labor, and minimizes human error. The statistics will be used in simulations to model bacteria motility. The methodology can be replicated with similar format of experimental microscopic images.**

Keywords—bacteria motility; trajectories detection; image processing; pattern recognition

I. INTRODUCTION

In recent years, technological advancement has been altering the world, and a lot of digital information has been fetched and stored in new data formats. With the development and availability of the new microscopes, it becomes feasible to film living bacteria movement. With the videos made, it is promising to classify the modes of motion, and calculate the diffusion coefficient of individual bacteria [2]. With quantitative analysis of image processing, the trajectories yields from the process provide full of important information in studying bacteria motility behaviors. However, how to extract and utilize the information in the data becomes challenging.

Microbes are abundant and have various impacts on human life. For example, microbial contamination of aqueous system has been recognized as one of the major risks to water resources and human health in terrestrial and coastal waters as well as groundwater. Pathogens can find their way to enter streams, lakes, beaches and groundwater systems as a result of accidental or intentional human activities. Bacteria have some beneficial aspects as well. For example, bioremediation, e.g. by injecting bacteria into soil, sediment, groundwater to metabolically degrade chemical contaminants in soil matrix (bioaugmentation), is one among the promising ways for in-situ cleanup of contaminated sites. Understanding bacterial motility behavior also have implications in water filtration in water treatment, more effectively using bacteria to decompose or generate energy from waste or to evaluate the risk of spread of antibiotic resistance [10]. Due to all these reasons it is important to understand how bacteria move in aquatic environments. One crucial step to fill in the gaps in understanding bacterial movement in aquatic environments is to understand the migration of motile and non-motile bacteria [1].

In general, bacteria transport can be described by the traditional advection- dispersion equation considering hydrodynamic, retention and biological growth and decay. The most basic mechanism is hydrodynamic transport due to advection in aqueous system mixed with small scale dispersion. Then retention arises by factors including attachment and detachment when bacteria interact with solid immobile phase [3]. Biological elimination is a process depends on the physical, chemical and biological environment. However, for motile bacteria, motility caused by flagellar movement can play a significant role in the movement and the spread of bacteria at small scales. Motile bacteria are equipped with spinning tail-shaped features called flagella, which enable them to swim. Flagellated motility is the best known bacterial translocation mechanism. Flagellated bacteria swim to modify their routes towards favorable environments, such as spatial and temporal changes in light, temperature, and substrate availability [4]. Several bacteria species, including one of most commonly seen ones, e.g. Escherichia coli, have self-propulsion ability by performing a sequence of "runs" and "tumbles" (Fig. 1) [6]. Running keeps the motion in approximately straight paths while tumbling changes the direction and the bacteria remain stationary. It is crucial to gain a quantitative understanding of bacterial motion behavior in order to predict the collective fate transport behavior of bacterial communities in aqueous environments. In previous studies, micro-scale models and macro-scale models have been built to simulate bacteria transport. However, the task of evaluating the modeled results is usually performed manually by microscopic observation. It involves long-term observation, which is time consuming, and easily affected by human errors [5]. The need to eliminate these limitations gives rise to the motivation of integrating automation in tracking and analysis bacteria motions.

Nowadays, interdisciplinary research has great potential for innovation if in collaboration with experts from different areas. Extracting bacteria motility features from Microscopic images

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.97

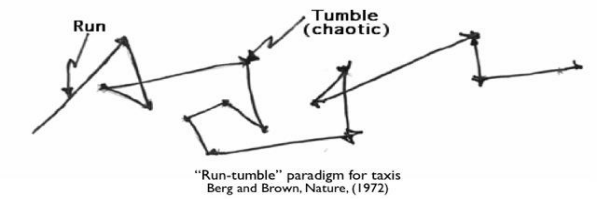


Fig. 1. Bacteria motility classified into run and tumble behavior

is among one of the applications and it combines expertise in natural science, applied mathematics and computer science. This paper presents an automatic pipeline for bacteria motile trajectory detection and bacteria motility behavior clustering (i.e., run and tumble mode). The extracted bacteria motility statistics which later on can be applied in simulation models for bacteria trajectory. This methodology can be replicated for study different strains of bacteria with similar format of the experimental microscopic images input. We used a soil indigenous bacterial strain *Azotobacter vinelandii* [22]as our model bacterium for studying its transport and fate has been our continuous research effort to understand the interplay of bacterial motility, transport, and horizontal gene transfer [10, 16, 17].

II. METHODOLOGY

To evaluate the bacteria motility using image processing technique, the first step is to reconstruct the observed bacteria motion as accurate as possible, then the following cluster analysis will be based on the detected trajectories. The tasks include detecting sequential images of bacteria and linking the detected results to follow the traces. Through this paper, each stack of 1,000 consecutive pictures is called a movie. Each image in a movie is called a frame. A reconstruction, an ordered series of point locations over the recording time of the individual frames, is called a (discrete) trajectory. And cells, particles, colloids and points are equivalent to bacteria [7]. After the trajectories of each movie are acquired, statistical features will be extracted, and similar behaviors of trajectory segments are grouped together. Evaluation of the bacteria motility can be divided into three tasks:

- Detecting Bacteria trajectory
- Clustering analysis on run and tumble behaviors
- Extracting the parameters determining the moving behavior

The main focus of this paper is on tasks one and two.

A. Data and data preparation

The images were taken using an inverted Axio Observer microscope (Carl Zeiss, Oberkochen, Germany) and a camera (Andor Technology iXon 897, Belfast, UK) controlled by Solis software (Andor Technology). Each stack of 1,000 pictures with a duration of 31 second was obtained with a 0.4 μm per pixel size resolution. *A. vinelandii* cells were grown on modified (no molybdenum) Burk's medium [18] plates with addition of 0.013 M ammonium acetate at 30 °C for 2 days and then in liquid media of modified (no molybdenum, no iron) Burk's medium [18] with addition of 0.013 M ammonium acetate shaking at 170 rpm for 18 to 20 hours [8, 16]. We followed the growth conditions used in previous studies [8, 16, 17] to facilitate comparability across studies to investigate the effects of motility.

B. Bacteria trajecroty detection

Bacteria detection was implemented using Particle Tracker plugin under ImageJ [7, 19, 20, 21], particle detection function initially developed by MOSAIC group, Sep, 2014 release [7]. The bacteria profile in this study is similar to colloidal spheres, with the 2nd order moment intensity, combined with the zero order moment. Visual inspection indicates that the particles were well segmented from the background and unwanted cells by the Particle Tracker plugin [7, 9].

The linking algorithm identifies points corresponding to the same physical particle in subsequent frames and links the positions $\{C^t\}_{t=1}^{T}$ into trajectories. This involves finding a set of associations between the point location matrices $\{C^t\}_{t=1}^{T}$ such that a predefined cost function is minimized. The present implementation is based on a particle matching algorithm [7] using a graph theory technique to determine optimal associations between two sets of points.

There are five main user defined parameters in this plugin, the first three for particle segmentation and detection, and another two for trajectory linking [7]:

1) Radius: Approximate radius of the particles in the images in units of pixels. The value should be slightly larger than the visible particle radius but smaller than the smallest inter-particle separation. 2) Cutoff: The score cut-off for the non-particle discrimination. 3) Percentile: The percentile (r) that determines which bright pixels are accepted as particles. All local maxima in the upper rth percentile of the image intensity distribution are considered candidate particles. 4) Displacement: The maximum number of pixels a particle is allowed to move between two succeeding frames.5) Link Range: The number of subsequent frames that is taken into account to determine the optimal correspondence matching.

C. Clustering analysis on run and tumble behavior

- Clustering feature extraction

In order to perform the clustering analysis, local velocity was chosen as the cluster feature input. It is defined as the Euclidean distance to origin normalized by the total time-lapse. A C++ code has been developed to analyze and extract features from the trajectories of each cell obtained from the particle tracking module.

The cell displacement from origin (the point on the trajectory where the state of the motion is intended to be determined) was obtained through trajectory analysis and was calculated using the following expression:

$$d_j = \sqrt{\left[\left(x_{j,t} - x_{j,0}\right)^2 + \left(y_{j,t} - y_{j,0}\right)^2\right]} \qquad (1)$$

where $x_{j,t}$ and $y_{j,t}$ are the coordinates of cell j in time step t, $x_{j,0}$ and $y_{j,0}$ are the initial location of cell j, and $d_{j,t}$ is the displacement to origin for cell j at time t. In preparing the

input data for motion modes clustering analysis, local velocity of cells at each location was calculated. A moving average inspired method, as shown in the equation (2), was used to minimize the intra-group variance and maximize the inter-group differences:

$$V_t = \frac{d_j - d_{j-n}}{w * timelapse\ per\ frame} \qquad (2)$$

where w is the window size for smoothing, and V_t is the local velocity of previous n data. The window size was determined by an exhausted search on various sizes (i.e., 3 ~ 15), and the window size 5 was chosen. The output of the C++ code (i.e., the local velocity for each trajectory segment) is then used as the input of a MATLAB(TM) module for further analysis on modes clustering and statistical analysis, e.g. distribution fitting and parameter estimation.

- Evaluation of optimal number of clusters

From the trajectory analysis of the motile bacteria, at any given time, some of the cells showed a clearly tumbling behavior, some showed actively running while other cells look like in transition. To further extract the statistical parameters of actively running bacteria, a cluster analysis was performed to separate the trajectory segments of actively running and non-running cells, based on the local velocity magnitude obtained from each trajectory. In order to determine the optimal number of clusters, we perform an evaluation of clusters analysis using a gap criterion. The gap criterion is derived from the common graphical approach which plots the error measurement versus several proposed cluster numbers. By locating the "elbow" of the plot, the most dramatic decrease in error measurement can be located. The gap criterion utilized this concept by estimating the number of clusters with the largest gap value. The gap value is computed using equation (3) which is also defined in [12]:

$$Gap_n(K) = E_n^*\{log(W_k)\} - lo\,g(W_k) \qquad (3)$$

where n is the sample size, k is the number of clusters being evaluated, and W_k is the pooled within cluster dispersion measurement which can be calculated using equation (4):

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \qquad (4)$$

where n_r is the total number of data points in cluster r and D_r is the sum of the pairwise distances for all points in cluster r. According to previous studies on bacteria motility [1], run and tumble are the two travelling behaviors, therefore, the number of clusters to be evaluated are given from 1 to 6.

- K-means method for clustering analysis

K-means method was first applied to group different motility behaviors within the same strains, which is described in [13]:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x_i \in S_i} \left\| x_j - \mu_i \right\|^2 \qquad (5)$$

where k is the number of cluster centroids, x is a one dimensional vector containing the local velocities for the trajectory of each cell, S indicates the subset of different clusters, and μ is the centroid for each cluster. The goal is to minimize the sum of the distances of each vector feature point to the corresponding cluster centroids. A MATLAB(TM) procedure 2 phase k-means algorithm [12] was used to perform this clustering task.

- Fuzzy c-means method for clustering analysis

Due to the fact that mode transitions for bacteria motility are a spectrum rather than a binary transition, a similar technique, Fuzzy c-means, has been tested. With Fuzzy c-means, feature data are clustered into given cluster numbers with every feature datum in the dataset belonging to every cluster to a certain degree. For example, a point that lies in the middle of two cluster centroids may have a relatively equal probability for both groups, while a point that lies closer to one centroid would have a higher probability of belonging to one group and a comparatively lower probability of belong to another group. A MATLAB(TM) Fuzzy c-means algorithm was applied for clustering [14].

- Naive Bayes Classifier for clustering analysis

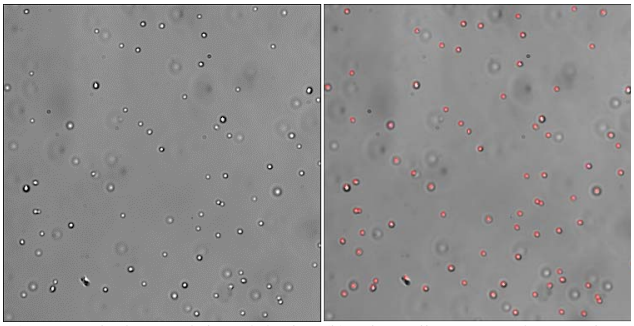
Since both k-means and Fuzzy c-means methods are unsupervised machine learning techniques, a supervised learning technique, Naive Bayes Classifier, was used for validation. A training set of 20 trajectories containing run segments and tumble segments from the detected trajectories were manually selected to train the naive Bayes model [15].

III. RESULTS AND DISCUSSION

A. Particle tracking results

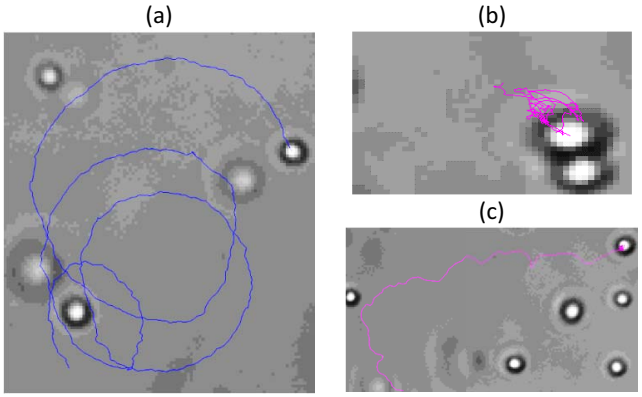
Fig. 2 shows an example of cell segmentation result using Particle Tracker plugin in ImageJ. The radius w was set to 3, cutoff size was set to 0, and r was set to 0.3%. Bacteria are successfully captured in the images and shown as circle particles with a bright core. The smaller, in-focus particles are closer to the surface. These are the particles being detected and whose behaviors will be examined in this paper. The larger and shadier circles are off-surface and off-focus ones; their motile behaviors may not fall into the same categories as the near surface ones. Therefore, the off-focus particles should be studied separately, and will be investigated in the near future.

The default values for the trajectory linking algorithm yields a reasonable result. As shown in Fig. 3, different swimming behaviors of bacteria are extracted. From the resulting trajectories, it is clear that there are running and tumbling modes for the same strain of bacteria. Run and tumble modes can also switch with each other randomly, one particle can be running in one frame and start tumbling in the next frame. Overall, there are 523 total trajectories detected in the Azotobacter vinelandii strain out of the 1,000 images.



(a) A sample frame of the original image from a video sequence (b) The cell segmentation result from (a)

Fig. 2. A sample segmentation result using Particle Tracker plugin in ImageJ



(a) Bacteria running in a circular fashion; (b) Bacteria tumbling; (c) Bacteria runs and then tumbles

Fig. 3. Particle tracking results

B. Clustering Analysis Results

Based on the local velocity calculation, the result turns out to be most reasonable when using window size 5. In Fig. 4, we plot the measured local velocity using a few selected trajectories. These plots show that for the same strain of bacteria, the velocity is not always constant and can vary from time to time. For example, in the first row of the plot (a, b, and c), sometimes the velocity is higher than other times, they represent a behavior changing between running and tumbling. In the plots in the second row (d, e, and f), comparatively, the velocity features are higher and the trajectory length are shorter compared to other plots. These cases represent a running behavior of bacteria. The shorter trajectories are due to bacteria running outside the field of view and therefore they could no longer be captured by the camera. Meanwhile, in the 3rd row of Fig.4 (g, h, and i), the measured velocity features are relatively small and they represent a tumbling behavior. With the 523 trajectories detected and a window size of 5 for the local velocity calculation, we obtained in total 17,437 trajectory segments which will be used for clustering.

With the gap criterion, the number of clusters was determined as 2, which matches the previous experimental observations. Table 1 shows the computed gap values when the number of clusters was set to 1-6. Note that, although the maximum gap value occurred when the number of clusters equals 2, the gap values for 2 or 3 clusters are very close, thus, there might be a transition mode in between run and tumble. Another study will be conducted in the near further to investigate the effect on a possible transition mode.

Table 2 shows the results of two unsupervised clustering algorithms, K-means, and Fuzzy c-means, with cluster number equals to 2. From row 1 and row 2, the mean value for tumble cluster is around 6-7 $\mu m/s$ and the run cluster is around 57-58 $\mu m/s$. It is clear that the two modes of motion have been separated. On the other hand, 20 trajectories are selected to perform a manual segmentation on run and tumble clusters in order to train the Naive Bayes model to validate the unsupervised method. As shown in row 3, the centroid for the tumble cluster is 6.2 whereas the centroid for the run cluster is 54, slightly to the right of the unsupervised method. Nevertheless, it is safe to conclude that the unsupervised clustering methods are doing a good job. Results show that run makes up roughly a quarter of the total movements, while tumble makes up almost 3 quarters of the movements, which indicates tumbles plays a bigger role, comparatively.

IV. CONLCUSION

In this paper, an automatic pipeline is presented for bacterial motility behavior analysis. Image processing and pattern recognition techniques were employed to detect the trajectories and cluster the two modes of motion in bacteria migration. The proposed pipeline can save laborious work in manually observing the trajectories and make statistical feature extraction of bacteria motility more accurate.

The proposed framework provides a systematic approach to process a very large number of bacteria images, and to extract knowledge and insights of bacteria motility that leads to a better understanding of bacteria behavior. Once the mystery of bacteria migration is solved, the beneficial aspects of bacteria could be further studied and applied in engineering, medicine, and other fields.

The proposed processing pipeline can be rapidly replicated in similar studies with different bacteria strains. Parameters of model inputs might vary in other studies, but the work flow should stay the same. Moreover, statistical features extracted can be applied in a forward trajectory model to simulate bacteria motility. Further studies could also be done in understanding the transition spectrum between run and tumble modes, the different variations of trajectory types within the same strain, such as, the differences in mostly run or mostly tumble, etc.

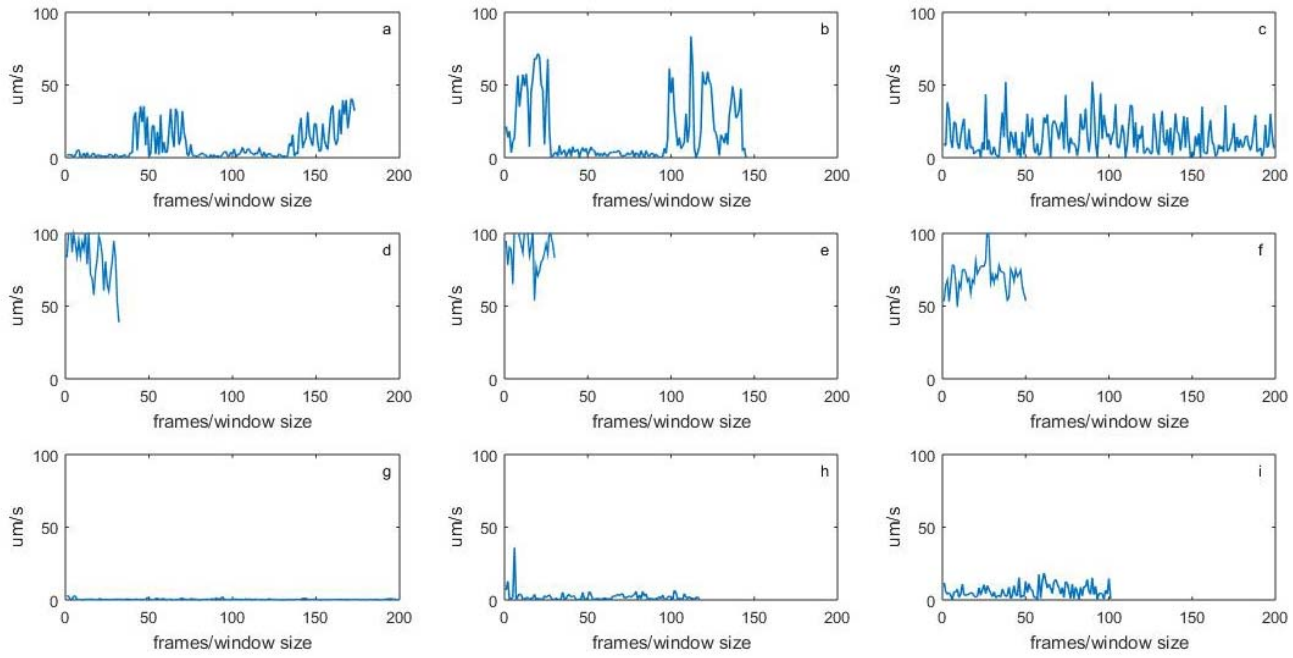


Fig. 4. Local Velocity distribution along the frame using window size 5

TABLE I. OPTIMAL NUMBER OF CLUSTERS EVALUATION RESULT

Inspected K	**1**	**2**	**3**	**4**	**5**	**6**
Criterion Values	1.5807	1.6422	1.6307	1.5996	1.5894	1.5663

TABLE II. THE CLUSTER MEAN AND STANDARD DEVIATION RESULT USING VARIOUS CLUSTERING METHODS

Method	**Cluster 1 – Tumble (μm)**			**Cluster 2- Run (μm)**		
	Mean	***Std. Dev***	***%***	***Mean***	***Std. Dev***	***%***
K means	6.8890	8.6668	24.87%	57.5016	19.0902	75.12%
Fuzzy c- means	6.0377	NA	24.80%	58.5037	NA	75.20%
Naive Bayes	6.2793	5.5194	NA	54.7152	21.5188	NA

ACKNOWLEDGMENT

The study was funded by NSF grant number 1114274 titled "Collaborative Research: Horizontal Gene Transfer in Porous Media: Experiments and Modeling".

REFERENCES

[1] Berg, Howard C. E. coli in Motion. Springer Science & Business Media, 2008.

[2] Goulian, Mark, and Sanford M. Simon. "Tracking single proteins within cells."Biophysical journal 79, no. 4 (2000): 2188-2198.

[3] Zajac, Mateusz, Jeffrey Law, Dragana Donna Cvetkovic, Macarena Pampillo, Lindsay McColl, Cynthia Pape, Gianni M. Di Guglielmo, Lynne M. Postovit, Andy V. Babwah, and Moshmi Bhattacharya. "GPR54 (KISS1R) transactivates EGFR to promote breast cancer cell invasiveness." PLoS One 6, no. 6 (2011): e21599.

[4] Darnton, Nicholas, Linda Turner, Kenneth Breuer, and Howard C. Berg. "Moving fluid with bacterial carpets." Biophysical journal 86, no. 3 (2004): 1863-1870.

[5] Langdon, Simon P. Cancer cell culture: methods and protocols. Vol. 88. Springer Science & Business Media, 2004.

[6] Berg, Howard C., and Douglas A. Brown. "Chemotaxis in Escherichia coli analysed by three-dimensional tracking." Nature 239, no. 5374 (1972): 500-504.

[7] Sbalzarini, Ivo F., and Petros Koumoutsakos. "Feature point tracking and trajectory analysis for video imaging in cell biology." Journal of structural biology151, no. 2 (2005): 182-195.

[8] Lu, Nanxi, Tara Bevard, Arash Massoudieh, Changyong Zhang, Alice C. Dohnalkova, Julie L. Zilles, and Thanh H. Nguyen. "Flagella-mediated differences in deposition dynamics for Azotobacter vinelandii in porous media."Environmental science & technology 47, no. 10 (2013): 5162-5170.

[9] Crocker, John C., and David G. Grier. "Methods of digital video microscopy for colloidal studies." Journal of colloid and interface science 179, no. 1 (1996): 298-310.

[10] Lu, Nanxi, Arash Massoudieh, Xiaomeng Liang, Tamir Kamai, Julie L. Zilles, Thanh H. Nguyen, and Timothy R. Ginn. "A kinetic model of gene transfer via natural transformation of Azotobacter vinelandii." Environmental Science: Water Research & Technology 1, no. 3 (2015): 363-374.

[11] MATLAB version 8.5.0. Natick, Massachusetts: The MathWorks Inc., 2015.

[12] Tibshirani, Robert, Guenther Walther, and Trevor Hastie. "Estimating the number of clusters in a data set via the gap statistic." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63, no. 2 (2001): 411-423.

[13] Spath, H. Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples. Translated by J. Goldschmidt. New York: Halsted Press, 1985.

[14] Bezdec, J.C., Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

[15] Mitchell, Tom M. "Machine learning. WCB.", McGraw Hill, 1997.

[16] N. Lu, A. Massoudieh, X. Liang, D. Hu, T. Kamai, T. R. Ginn, T. H. Nguyen. “Swimming mtility prevents bacterial attachments to surfaces.” Journal of Environmental Quality, no.4 (2015): 1366-1375.

[17] Massoudieh, Arash, Nanxi Lu, Xiaomeng Liang, Thanh H. Nguyen, and Timothy R. Ginn. "Bayesian process-identification in bacteria transport in porous media."Journal of contaminant hydrology 153 (2013): 78-91.

[18] Strandberg, G. W., and P. W. Wilson. "Formation of the nitrogen-fixing enzyme system in Azotobacter vinelandii." Canadian Journal of Microbiology 14, no. 1 (1968): 25-31.

[19] Abràmoff, Michael D., Paulo J. Magalhães, and Sunanda J. Ram. "Image processing with ImageJ." Biophotonics international 11, no. 7 (2004): 36-42.

[20] Schneider, Caroline A., Wayne S. Rasband, and Kevin W. Eliceiri. "NIH Image to ImageJ: 25 years of image analysis." Nature methods 9, no. 7 (2012): 671-675.

[21] Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997-2015.

[22] Setubal, João C., Patricia dos Santos, Barry S. Goldman, Helga Ertesvåg, Guadelupe Espin, Luis M. Rubio, Svein Valla et al. "Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes." Journal of bacteriology 191, no. 14 (2009): 4534-4545.