Fast and simple tool for the quantification of biofilm-embedded cells sub-populations from fluorescent microscopic images

Short running title: tool for the quantitative analysis fluorescent images

Mikhail I Bogachev[1,2], Oleg A Markelov[1], Elena Trizna[2], Diana Baydamshina[2], Pavel Zelenikhin[2], Regina Murtazina[2], Airat R Kayumov[2,*]

[1]Biomedical Engineering Research Center, St. Petersburg Electrotechnical University, St. Petersburg, Russia

[2]Institute of Fundamental Medicine and Biology, Kazan Federal University, Russia

* Corresponding author

E-mail: kairatr@yandex.ru

Abstract

Fluorescent staining is a common tool for both quantitative and qualitative assessment of pro- and eukaryotic cells sub-population fractions by using microscopy and flow cytometry. However, direct cell counting by flow cytometry is often limited, for example when working with cells rigidly adhered either to each other or to external surfaces like in bacterial biofilms or adherent cell lines and tissue samples. An alternative approach is provided by using fluorescent microscopy and confocal laser scanning microscopy (CLSM), which enables the evaluation of fractions of cells subpopulations in a given sample. To facilitate the quantitative assessment of cell fractions in microphotographs, we suggest a simple two-step algorithm that combines the cell selection based and the statistical approaches. Based on a series of experimental measurements performed on bacterial and eukaryotic cells under various measurement conditions, we show explicitly that the suggested approach effectively accounts for the fractions of different cell sub-populations (like the live/dead staining in our samples) in all studied cases that are in good agreement with manual cell counting on microphotographs and flow cytometry data. This algorithm is implemented as a simple software tool that includes an intuitive and user-friendly graphical interface for the initial adjustment of algorithm parameters to the microscopic imaging conditions as well as for the sequential analysis of homogeneous series of similar microscopic images without further user intervention. The software tool entitled *BioFilmAnalyzer* is freely available online at http://kpfu.ru/eng/strau/laboratories/molecular-genetics-of-microorganisms-lab/software/biofilmanalyzer-v10.

Introduction

One of the key issues in both pro- and eukaryotic cell studies is the quantitative characterization of cellular subpopulations like the estimation of the fractions of either live or dead cells in a given population, differentiation of bacterial species in mixed biofilms or eukaryotic cell types in culture. There are two common experimental approaches to these issues, namely the flow cytometry and the fluorescent microscopy. In both methods the cells are stained with fluorescent dyes which specifically differentiate the cells of interest. Thus, Syto9/PI, DioC6/PI, AO/PI, CFDA/PI, Calcein AM/PI, Hoechst/PI and many other combinations of dual staining are widely used to differentiate viable and non-viable cells [1-3]. Normally, the first dye is biochemically modified by viable cells followed by the production of the green-fluorescent product. The second dye like the propidium iodide or ethidium bromide penetrates through the damaged membrane of dead cells forming complexes with nucleic acids and providing red fluorescence. While many researchers report that the estimation of viable cells fraction by using vital staining often exhibits significant differences in comparison with the values obtained by using classical microbiological methods [4], fluorescent staining remains a fast and easy approach to the quantification of (non-)viable cells.

While flow cytometry provides with a much more accurate assessment of the cell subpopulation fractions [5, 6], it has principal limitations that significantly narrow its application area [7]. In particular, cells being adhered to each other and to external surfaces should be suspended prior to their infusion into a cytometer that appears difficult when, for example, bacterial biofilms or strongly adherent cells are analyzed, or the original structure of the cell colonies, cell complex or tissue structure should be preserved. Moreover, flow cytometer is normally unable to detect particles <0.500 μm [8]. Finally, currently available flow cytometry systems require considerable amount of maintenance and highly skilled operators.

Fluorescent microscopy is largely free of above limitations and provides a reasonable

71 alternative to the cytometric measurements. However, in the presence of adherent and/or spore-like
72 cells they largely overlap leading to the limitations of direct cell selection and counting algorithms in
73 the microscopic images. The situation gets even more complicated when the cells are not equidistantly
74 stained, image quality and color balance varies in different fields of view. Manual counting is usually
75 possible, while it requires significant efforts from experts increasing the lab personnel workload
76 drastically. Thus, automatic or semi-automatic analysis of cells seems to be a fast and easy approach for
77 the microscopic data quantification. In the last two decades, a number of methods and computer-
78 assisted algorithms have been developed to resolve the cell counting issue implemented in a number of
79 both commercial and free software tools [9-13]. Existing software solutions include cell counting and
80 classification algorithms [14], estimation of their parameters from microscopic imaging [15], 3d
81 reconstructions from confocal microscopy data [16] and several other more specific applications.
82 However, automatic microscopic image analysis remains challenging in the presence of adherent and/or
83 spore-like cells that are common conditions in biofilm studies. Automatic counting methods are usually
84 based either (i) on detection, selection and counting of discrete objects, or (ii) on the statistical analysis
85 of the image properties that avoid direct counting approach and estimate some effective characteristics
86 from the statistical properties of the entire image [17]. While the detection methods fail under cell
87 overlapping conditions, the statistical assessment methods are unable to differentiate between various
88 types of cells. Among few exceptions that largely overcome the above limitations, a very recently
89 designed software tool for quantification of live/dead cells in a biofilm based on a series of image
90 transformation could be mentioned [18].

91 Another common disadvantage of many automatic image analysis tools in practical settings is
92 often, though may sound surprising, their excessive use of automation. Quite often image analysis
93 algorithms use complex transformations with parameters that can be hardly controlled by a user who is
94 not an expert in digital image processing. As a result, there is little or no feedback between the

95 algorithm and its end user. Thus, the user deals with a kind of black box design, where an image is
96 inserted and a value comes out, without being able to cross-check the performance of the algorithm at
97 some intermediate steps. Despite the increasing complexity and performance of image processing tools,
98 the variety of cell structures and microscopic imaging conditions to our opinion is still too broad to
99 fully rely upon automation in all cases.

100 Here we suggest an algorithm based on a simple combination of the object counting and
101 statistical approaches and easy-to-use cell-counting software tool, *BioFilmAnalyzer*, freely available at
102 http://kpfu.ru/eng/strau/laboratories/molecular-genetics-of-microorganisms-
103 lab/software/biofilmanalyzer-v10. Following preliminary threshold-based filtering and segmentation of
104 the image, an effective number of cells is calculated under partially-manual control by the investigator.
105 Based on a series of experimental measurements performed in bacterial cells of *S.auerus* and *B.subtillis*
106 exhibiting different shapes as well as eukaryotic cells, we show explicitly that the suggested approach
107 effectively account for the fractions of live/dead cells in all studied cases. The validity of the
108 *BioFilmAnalyzer* based cells live/dead fractions quantification was assessed by comparison with the
109 results of manual counting performed by several experts in visual microscopic image analysis and
110 cytometric measurements.

111

112

113 # Materials and methods

114 ## Bacterial strains, cell lines and fluorescent microscopy

115 The fluorescent microscopic images of bacterial cells obtained in previous works were used
116 (BMRI, Acta, SR, BIOnano). Briefly, *Staphylococcus aureus* (ATCC® 29213™) and *Bacillus subtilis*
117 *168* grown in 35-mm TC-treated polystyrol plates (Eppendorf) for 48 hours under static conditions at
118 37°C to obtain rigid biofilm structures were further subjected to differential live/dead fluorescent
119 staining.

120 The human colon adenocarcinoma Caco-2 cells (RCCC) were cultured in DMEM supplemented
121 with 10% FBS, 2 mM L-glutamine, 100 μg ml^{-1} penicillin and 100 μg ml^{-1} streptomycin. The cells
122 were seeded in 24-well plates at the density of 30000 cells per well and allowed to attach overnight.
123 The cells were cultured at 37 ºC and 5% CO_2 until 70% confluence and camptothecin (Sigma-Aldrich)
124 was added in final concentration of 6 μM. After 24 h of exposition the cells were subjected to
125 fluorescent staining and analyzed with flow cytometry and fluorescent microscopy.

126 The viability of biofilm-embedded cells was evaluated by staining for 5 min with the Acridine
127 orange (Sigma) at final concentration of 0.12 μg/ml (green fluorescence) 3,3′-Dihexyloxacarbocyanine
128 iodide (DioC6) (Sigma) at final concentration of 0.02 μg/ml (green fluorescence) and propidium iodide
129 (Sigma) at final concentration of 3 μg/ml (red fluorescence) to differentiate between bacteria with
130 intact and damaged cell membranes (live and dead cells). The eukaryotic cells were stained with DioC6
131 (0.02 μg/ml) and propidium iodide (3 μg/ml). The microscopic imaging was performed using either
132 Carl Zeiss Observer 1.0 microscope or Carl Zeiss CLSM780 with 40-100× magnification.

133

134

135

Flow cytometry

Cytometric analysis was performed using BD FACSCanto II flow cytometer. The Caco-2 cells were stained as described above. Following data analysis was carried out using FACSDiva software.

Image analysis methodology

Image analysis starts with the preliminary detection of pixels with given color channel intensity (red, green or overall) exceeding a given threshold. Fig 1A shows a sample of original microscopic image, while Fig 1B shows the result of threshold-based selection of the red channel according to $I_R - I_G > T$, where I_R and I_G are the intensities in the red and in the green channels, respectively, $T = 30$ is the threshold setting. In the next step, the preprocessed image is segmented into separate non-overlapping objects by adjusting horizontally, vertically or diagonally neighboring pixels above the threshold to the same object. Additionally, objects with a given set of properties that determine the cell subpopulation are selected. Fig 1C exemplifies the results of selection by size, with objects only in a given size range between $s_{\min}$ = 500 and $s_{\max}$ = 1000 pixels being highlighted, where the color encodes each object.

In the following, we count the selected objects, calculate their total area, and obtain the average size of the typical cell from the studied subpopulation. Next we determine the effective number of cells by dividing the total area above the threshold (Fig 1B) by the average size of selected cells (Fig 1C) according to

$$N_R^{\text{eff}} = \frac{S_R}{\overline{S}_R\left[s_{\min} \ldots s_{\max}\right]} \text{ и } N_G^{\text{eff}} = \frac{S_G}{\overline{S}_G\left[s_{\min} \ldots s_{\max}\right]},$$

where N_R^{eff}, N_G^{eff} are the effective numbers of cells in the red and green channels, respectively; S_R, S_G are the total area of selected cells belonging to the red and green channels, respectively;

$\overline{S}_R\left[s_{\min}\ldots s_{\max}\right]$, $\overline{S}_G\left[s_{\min}\ldots s_{\max}\right]$ are the average area of cells with a size range between $s_{\min}$ and $s_{\max}$.

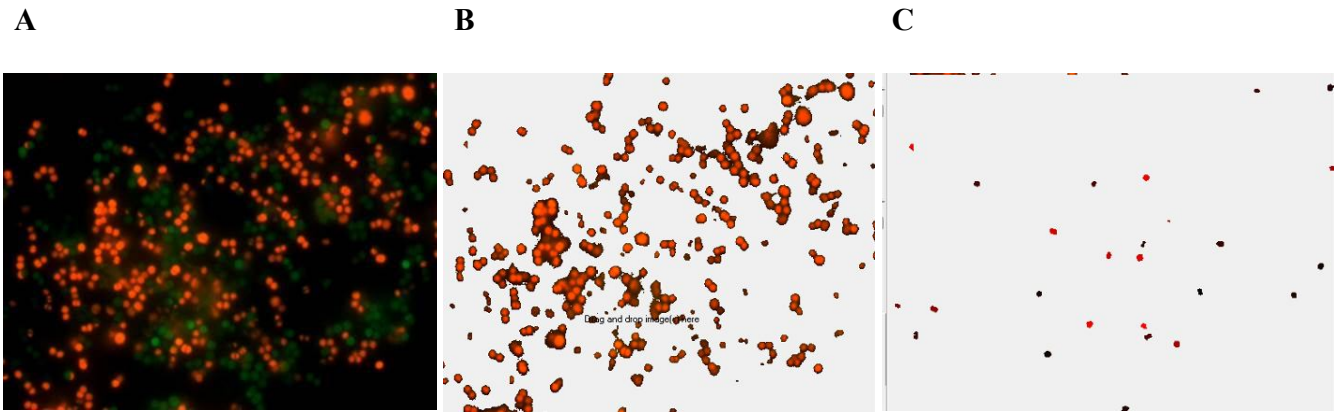


Fig 1. Consecutive steps in cell quantification by *BioFilmAnalyzer*. (A) original image containing overlapped red and green channels, (B) selected red channel data after threshold-based filtering, (C) selected cells of size between $s_{\min}$ and $s_{\max}$ that are used to determine the effective single cell size with each separate cell shown by another color as determined by the automatic segmentation algorithm.

While T, $s_{\min}$ and $s_{\max}$ have to be appropriately chosen, their choice has to be performed once in a series of similar experiments considering similar cell types. Accordingly, so far their manual choice by expert seems to be the easiest solution option, since the feedback from the first samples being tested allows for a more specific adjustment of these parameters, to avoid potentially bizzare results that may arise in the case of their blind application with no feedback. In the following, for the entire series of experiments the parameters are fixed and no further manual adjustment is required. Thus a series of images representing different fields of view under identical conditions in simply passed through an algorithm. When the differences in the conditions are minor and do not change significantly the microscopic images, but influence only some of their parameters like the live/dead cell fractions, such

as testing antimicrobials with gradually changing concentrations, several series of images can be submitted without further adjustment of the algorithm parameters. The implemented software solution organizes the results of calculations in a table, that could be exported, and represent them in a graph.

Statistical assessment

Here we used simple linear regression without intercept term, i.e. $y = kx$. In the suggested regression model the ideal case corresponds to k=1 or simply $y = x$ that would mean perfect agreement between automatic and manual expert counting. The quality of the results is characterized by two independent coefficients. The first one is the standard coefficient of determination R^2. $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, where SS_{tot} is the total sum of squares $SS_{tot} = \sum_i \left(f_i - \overline{y} \right)^2$ and SS_{res} is the residual sum of squares $SS_{res} = \sum_i \left(y_i - f_i \right)^2$, here $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ is the mean value of all analyzed data points and f_i is calculated regression point. The R^2 coefficient indicates how well the analyzed data set is replicated by the regression model.

In this investigation it is also important how close is the regression model to the ideal case $y = x$. Therefore we introduce a similarly designed metric of how well the model $y = kx$ is close to the ideal counting line $y = x$ which is called L^2. Its definition is similar to R^2 besides the calculation is done for regression points f_i. SS_{res} for L^2 is calculated as $SS_{res} = \sum_i \left(f_i - x_i \right)^2$, where x_i is the corresponding abscissa value for the current f_i, and SS_{tot} is the same as in R^2. In ideal case when observational

regression line follows $y=x$, $L^2 = 1 - \frac{\sum_i (f_i - x_i)^2}{\sum_i (f_i - \overline{f})^2} = 1 - \frac{0}{\sum_i (f_i - \overline{f})^2} = 1$. Thus both R^2 and L^2 coefficients range from 0 to 1.

Results and Discussion

Software description

BioFilmAnalyzer software can be used to count any bacterial or eukaryotic fluorescent stained cells from two-dimensional microphotographs provided either as single images or as a series of images obtained under similar conditions such as quality and magnification (for example, several views of the same sample taken with fixed camera settings). This in-house algorithm was implemented in C++ programming language and is compatible with Windows (XP versions and higher). The software tool has a simple user-friendly environment for the image analysis with the two-step algorithm allowing manual parameter adjustments by end user at each step. The logic of the image analysis is based on the preliminary adjustment of the algorithm parameters by using one or several images that the investigator finds more or less representative for the studied cohort in terms of imaging conditions. In the first step, simple drag & drop of a single image into the program window initiates its instant processing including threshold based detection of cells according to a specified rule based on the exceedance of a given threshold by either a certain color channel, difference between color channels or overall intensity. The threshold value T which is the only algorithm parameter in this first step can next be adjusted by the investigator by its increasing in cases of strongly autofluorescent background or by its decreasing in low contrast images until the background noise is eliminated. For image series exhibiting strongly non-homogeneous color distributions in the studied color channels, instead of the manual adjustment of the threshold value T for each of the images individually, preliminary image color normalization often appears a faster alternative. Since color normalization is a simple and standard image preparation procedure that is straightforward and thus can be applied consecutively to a series of images using many commercial or free image processing software tools (e.g., ImageMagick, Gimp etc.), we do not focus further on this issue. However, as we show below, since our algorithm is robust against moderate variations of the threshold T around its optimized value for a given color distribution, the color

223 normalization appears unnecessary for image series with moderate variations of color distributions and
224 may be required only rarely under considerable variations of imaging conditions. The view can be
225 easily switched between the original and the processed images by a single click on the upper image
226 panel. In the second step, the effective cell size should be adjusted such that only or nearly only single
227 cells appear in the lower image panel, which shows the processed image after a simple segmentation
228 procedure. Increasing of the lower limit for the cell size helps to eliminate some noise bursts, while
229 increasing of the upper limit eliminates large patches of adherent cells from the effective cell size
230 statistics. With the cell size window representing only a simple criteria that often appears insufficient,
231 whenever necessary, further elimination of anomalous segments can be done manually by selecting and
232 double-clicking over them in the lower image panel. Since the parameters including sizes of different
233 sub-population of cells can differ from one another (e.g., non-viable eukaryotic cells are commonly
234 smaller than viable cells), the effective cell size should be re-adjusted for each studied sub-population
235 of cells. For this reason, the analysis of each color channel should be performed separately, and thus
236 with the exception of the intensity based analysis rule, all other options analyze color channels
237 individually. Thus they perform similarly for single or overlapped color channel images. Finally, the
238 effective number of cells is determined as the total area of the image exceeding the threshold T shown
239 in the upper panel divided by the effective cell size determined from the lower image panel. Once a
240 reasonable set of algorithm parameters in found, a series of up to 100 images can be dropped onto the
241 program window for fully automated analysis with the same set of parameters as determined from the
242 first one or several representative images without further user intervention. Finally, the results can be
243 exported to the MS Excel for the validation of the results and following statistical analysis.

244 In comparison with some other freely available image processing and analysis software [18-20],
245 our image-processing procedure is semi-automatic: settings are modified by manually adjusting the
246 analysis parameters and each step is controlled with an expert to evaluate the analysis quality. Being

247 optimized once for a first image in a series the software analyzes other images automatically over
248 coffee-break.

249 Software validation by bacterial cells counting

250 We analyzed the efficiency of the proposed algorithm using the fluorescent images of coccal
251 (*Staphylococcus aureus*) and rod (*Bacillus subtilis*) cell morphologies containing different fraction of
252 viable (green-stained) cells. Since the cell suspension can be easily analyzed with flow cytometry, we
253 focused on the analysis of adherent, biofilm-embedded cells. For that, we used previously obtained
254 series of microphotographs of *S.aureus* in 48-72 h old biofilms treated with different antimicrobials
255 [21, 22]. Cells were stained with DioC6 and propidium iodide to differentiate the viable and non-viable
256 cells, and non-viable cells fraction was quantified with *BioFilmAnalyzer* software. As a representative
257 example, twelve images with overlapped red and green channels containing different fraction of viable
258 cells are shown in Fig 2. Since the image brightness, contrast and saturation vary from image to image
259 depending on the staining quality, microscope settings and sample itself, for each microphotograph
260 shown in Fig 2 the thresholds T (ranging from 25 to 55) and effective cell sizes were chosen
261 individually. Next to validate the performance of the algorithm when the intensity threshold is chosen
262 quite arbitrarily without careful manual adjustment to the imaging conditions, these microphotographs
263 were analyzed consequently for different analysis thresholds (15, 20, 30, 45 and 60) and the results
264 obtained by the automatic cell counting were plotted as a linear function $y=kx$ of the manual cell
265 counting (Fig 3 A-C). While the number of both red (non-viable, Fig 3A) and green (viable, Fig 3B)
266 cells decreases at higher thresholds, their fractions (i.e., live/dead ratio) remained similar for each
267 threshold value exhibiting no significant differences with the manual evaluation data (compare thick
268 dashed line corresponding to the ideal fit of manual and automatic counting, Fig 3C).

269 Table 1 shows the regression coefficients k for the linear regressions $y=kx$ and their coefficients
270 of determination R^2. Since by definition there should be no systematic shift between the automatic and

271 the manual count, i.e. in the absence of viable or non-viable cells the respective number of cells equals
272 zero, we used the simplest linear regression model without intercept. Since the ideal counting
273 corresponds to k=1 or simply to the line given by $y=x$, we also calculated another coefficient denoted
274 L^2 that determines the deviation of the obtained regression line $y=kx$ from the ideal counting line $y=x$.
275 Of note, the automatic cell enumeration by *BioFilmAnalyzer* software exhibited the best fit with
276 manual count at the analysis threshold of 45 (for images present on Fig 2). For further details on the
277 statistical analysis of our results, we refer to the Materials and Methods section at the end of this paper.
278 The individual adjustment of the threshold and effective cell size is obviously possible only
279 when a small number of images should be analyzed. For a more accurate quantification, a series of 10
280 or more images from the same sample normally should be analyzed. The *BioFilmAnalyzer* software
281 allows analysis with constant settings of threshold and cell size of 2 and more images when they are
282 dragged & dropped simultaneously onto the program window. To estimate the performance of the
283 software when the images with different quality are analyzed without individual optimization of the
284 intensity threshold, the series of 115 microphotographs randomly taken from different experiments
285 were analyzed consequently at different thresholds (15, 20, 30, 45 and 60) with fixed cell size ranges
286 used to determine the effective cell sizes. The results were compared against manual cell counting data
287 (Fig 3 D-F). Similarly to previous results obtained for 12 images with accurate settings of both
288 thresholds and cell sizes, the live/dead ratio remains almost similar for each threshold value and the
289 obtained regression fit is in an excellent agreement with the manual cell counting data indicated by L^2
290 being very close to 1.0 at thresholds in the range of 15-30 (Table 1).
291

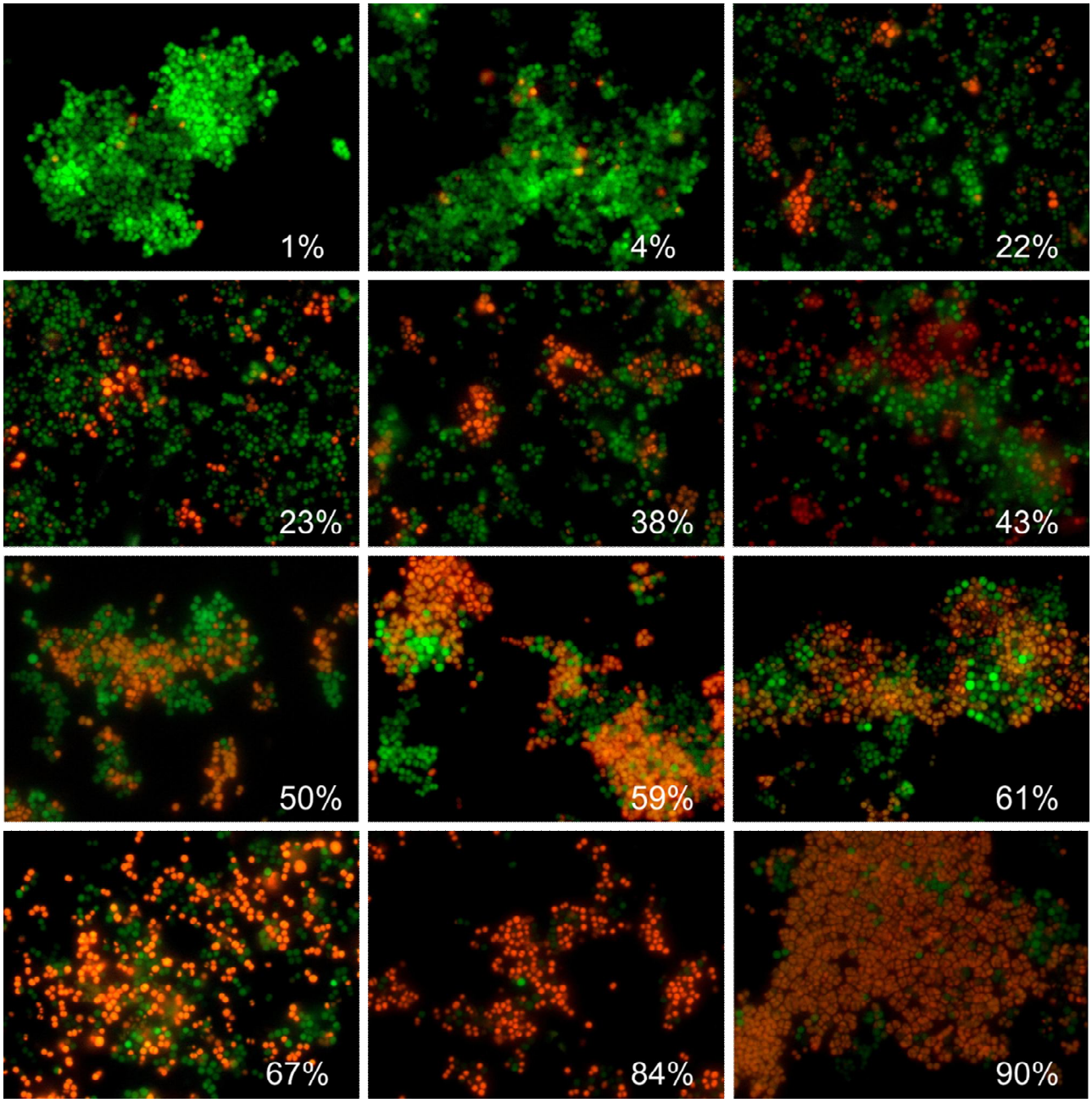
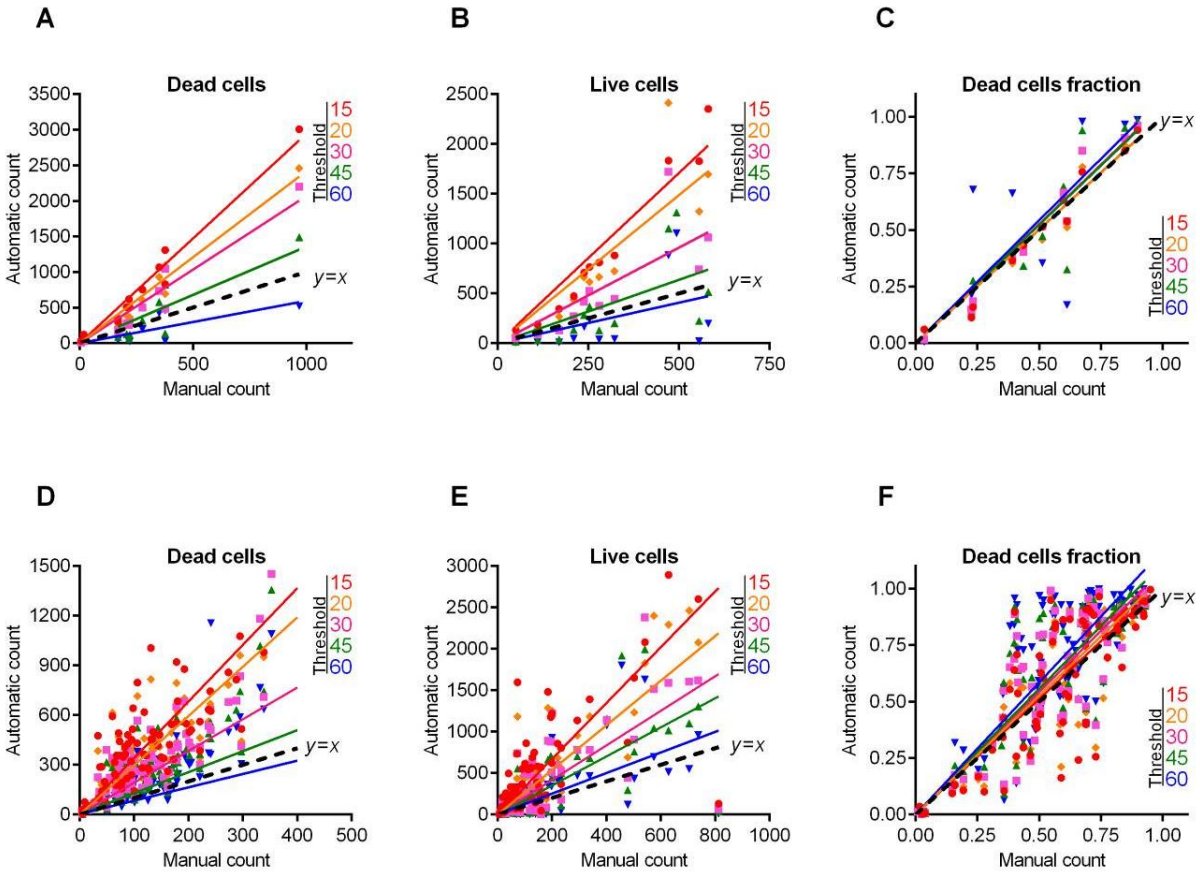


Fig 2. Evaluation of the *S.aureus* biofilm-embedded non-viable cells fraction by using *BioFilmAnalyzer* software exemplified for 12 microscopic images. The percentage of red-stained cells quantified by the *BioFilmAnalyzer* software is shown in each panel.



298

299 **Fig 3. *S.aureus* cells count and live/dead ratio dependence on the image analysis threshold *T*.** Full
300 lines show the linear regression lines, while dashed line shows the ideal counting line as determined by
301 the manual analysis performed by several experts in visual microscopic image analysis. Panels A-C
302 show data for the 12 images presented in Fig 2. Panels D-F show data for the 115 microscopic images
303 taken randomly from different experiments with various imaging conditions.

304

305

306 **Table 1. Regression coefficients *k*, the coefficients of determination R^2 and the accuracy**
307 **coefficient L^2 indicating the correspondence between the automatic and the manual counting for**
308 **the *S.aureus* live/dead ratios on fluorescent images.**

	12 images			115 images		
Threshold	k	R^2	L^2	k	R^2	L^2
15	1,01	0,95	1,00	1,02	0,63	1,00
20	1,00	0,94	1,00	1,05	0,65	0,99
30	1,05	0,94	0,99	1,08	0,66	0,98
45	1,05	0,73	0,99	1,11	0,72	0,96
60	1,09	0,59	0,98	1,17	0,67	0,92

309

310

311 Next, similar analysis was performed on microscopic images of *B.subtilis* biofilm-embedded
312 cells treated with different antimicrobials [23]. Like in the previous example, Fig 4 shows 12
313 representative images with different fraction of viable cells quantified with *BioFilmAnalyzer* software.
314 Fig 5 A-C is designed similar as Fig 3 and shows the interdependence of the image analysis threshold
315 and cell number calculated by *BioFilmAnalyzer*. Similar to *S.aureus* microphotographs, the cell number
316 calculated automatically increased at high threshold with the best fit with manual count at threshold
317 value of 45, while the live/dead ratio remained unchanged with L^2 exceeding 0.96 at thresholds up to
318 45, suggesting that the performance of the algorithm does not depend on the fine tuning of the
319 threshold and on the cell shape. Fig 5 D-F shows the analysis of 50 randomly chosen images of
320 live/dead stained *B.subtilis* cells. Similarly, the statistical analysis does not reveal significant
321 differences between automatic and manual live/dead ratio estimation (Table 2).

322 While the absolute cell counting depends on the image analysis threshold which should be
323 adjusted manually and therefore has a factor of subjectivity, the live/dead ration quantification does not
324 depend on the threshold settings in range of 15-60 color intensity units (on the 0..255 scale) exhibiting
325 linear function with R^2 values exceeding 0.9 and regression coefficients of k ranging between 0.9 and
326 1.1 (Tables 1 and 2). This fact allows performing the automatic analysis of multiple images with
327 constant settings optimized for first image in a series.

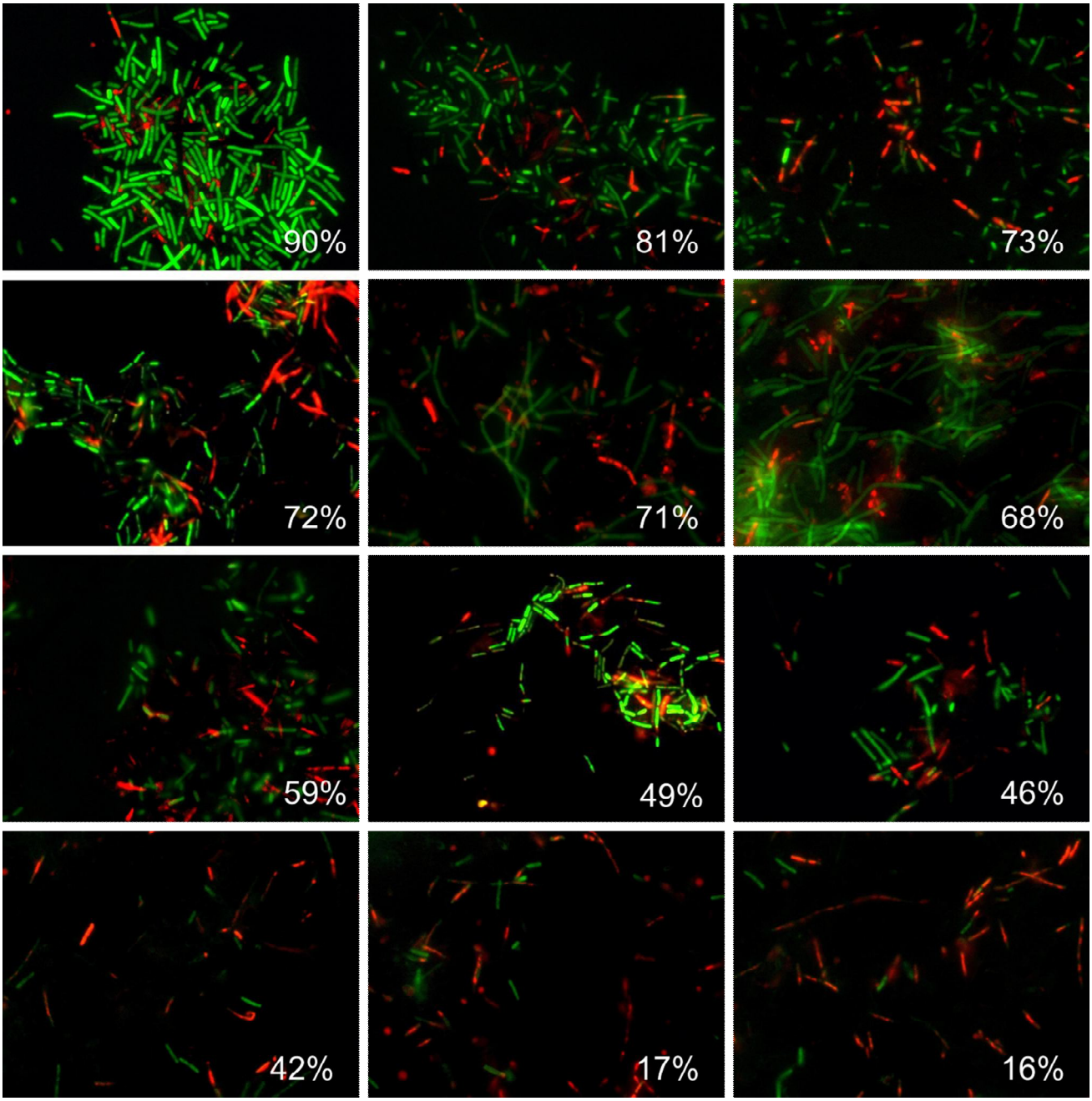


Fig 4. Evaluation of the *B.subtilis* biofilm-embedded non-viable cells fraction by using *BioFilmAnalyzer* software exemplified for 12 microscopic images. The percentage of red-stained cells quantified by the *BioFilmAnalyzer* software is shown in each panel.

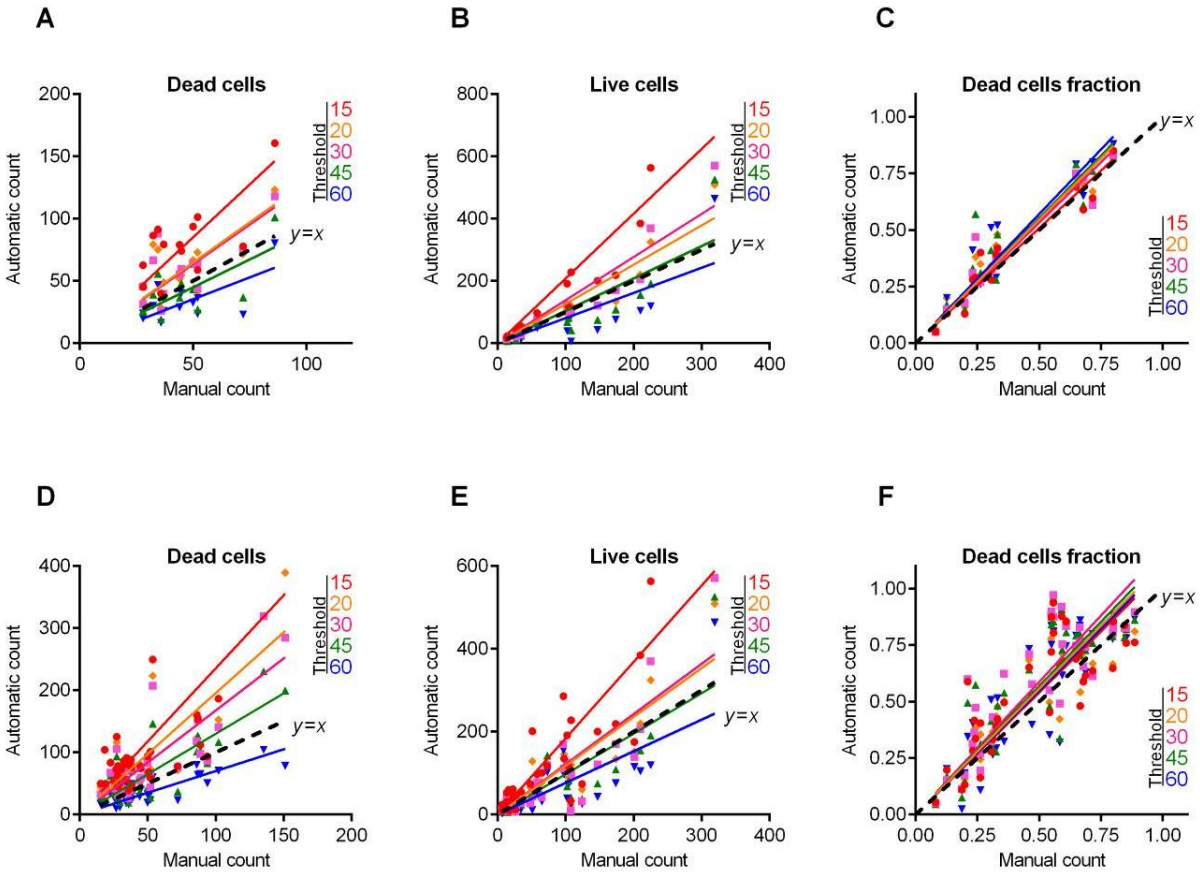


Fig 5. *B.subtilis* cells count and live/dead ratio dependence on the image analysis threshold *T*. Full lines show the linear regression lines, while dashed line shows the ideal counting line as determined by the manual analysis performed by several experts in visual microscopic image analysis. Panels A-C show data for the 12 images presented in Fig 4. Panels D-F show data for the 50 microscopic images taken randomly from different experiments with various imaging conditions.

343 **Table 2. Regression coefficients *k*, the coefficients of determination R^2 and the accuracy**
344 **coefficient L^2 indicating the correspondence between the automatic and the manual counting for**
345 **the *B.subtilis* live/dead ratios on fluorescent images**

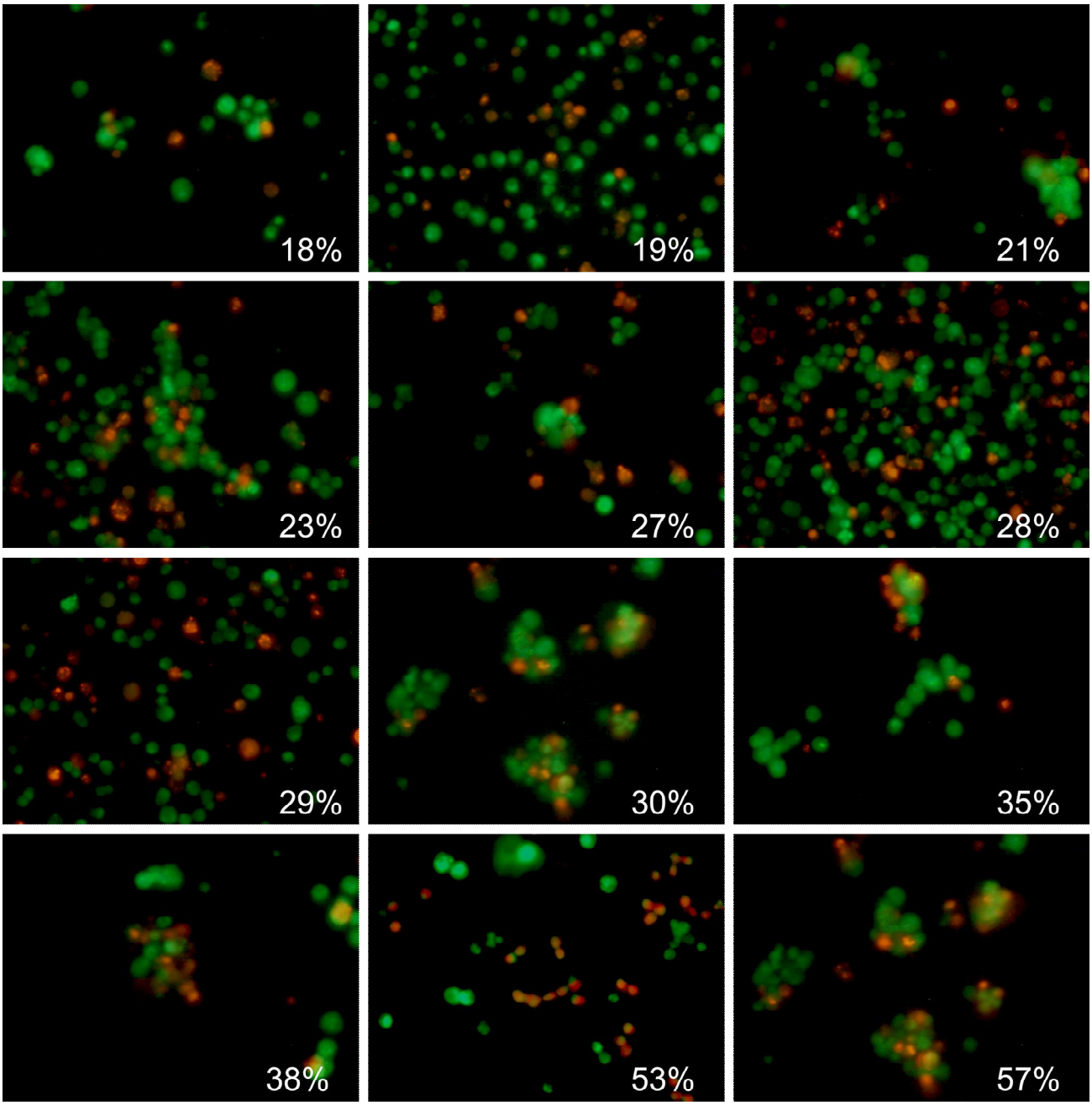
	12 images			50 images		
Threshold	k	R^2	L^2	k	R^2	L^2
15	1,02	0,92	1,00	1,08	0,72	0,97
20	1,09	0,93	0,97	1,11	0,74	0,95
30	1,07	0,91	0,98	1,17	0,75	0,90
45	1,10	0,84	0,96	1,14	0,74	0,93
60	1,14	0,90	0,93	1,10	0,74	0,96

346

347 ## Eukaryotic cells counting

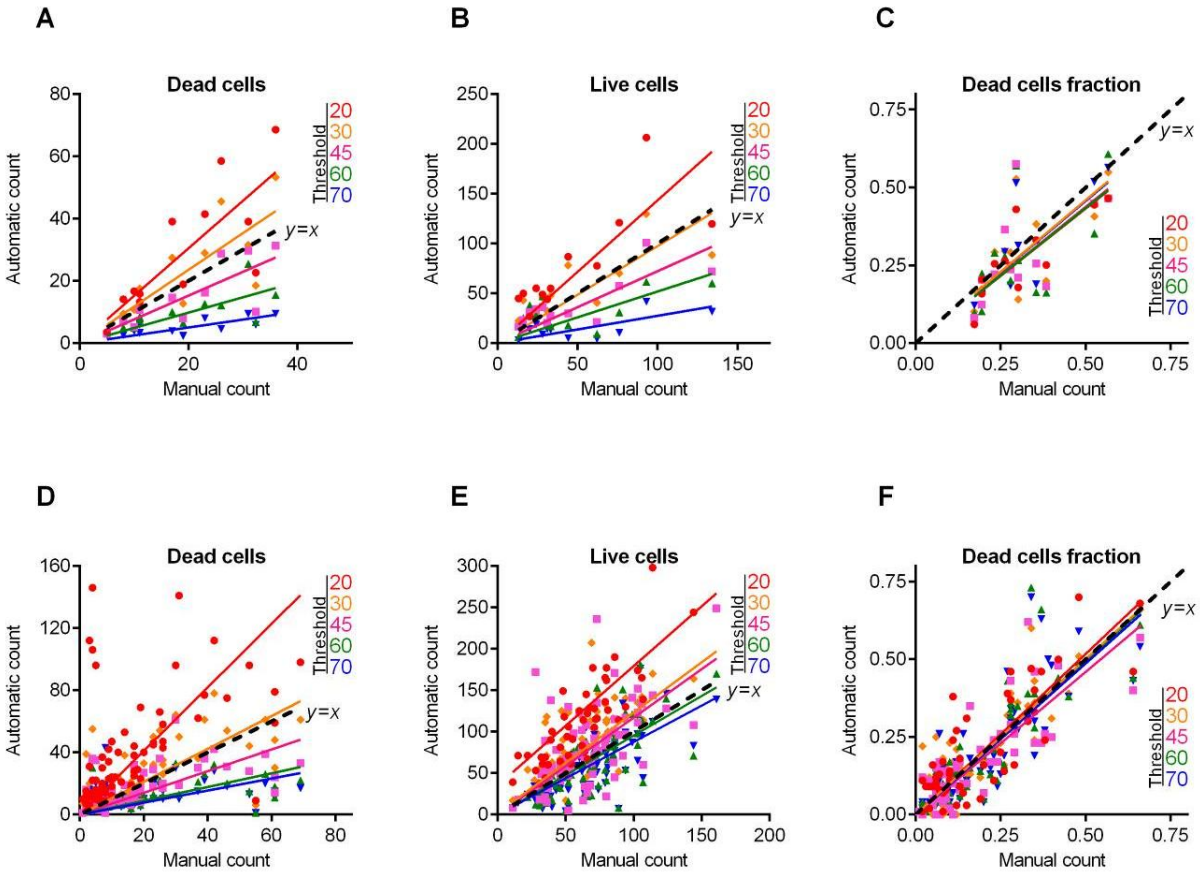
348 The performance of the *BioFilmAnalyzer* software was also analyzed for the eukaryotic cells.
349 For that, Caco-2 cells were treated with different concentrations of camptothecin and analyzed after
350 24h of exposition. Fig 6 and 7C show the fractions of viable cells quantified with *BioFilmAnalyzer*
351 software on microphotographs with overlapped green and red channels. Here, the best fit of manual and
352 automatic count of either live or dead cells was observed with the analysis threshold T=30 (Fig 7),
353 while higher background level led to artificial results at T=15 (not shown). In contrast to bacterial cells,
354 the accuracy was lower (L^2 over 0.80), probably, because of discrepancies in cell sizes and non-evenly
355 staining of cells. In contrast, when the software performance was evaluated on a series of 87 randomly
356 chosen images (Fig 7 D-F), the automatic estimation of live/dead fractions fits with manual one with
357 high confidence level (L^2 equals 1.0, see Table 3). This effect could be attributed to the larger effective
358 sizes and thus also smaller average number of cells in each field of view for eukaryotic cells compared
359 to bacterial cells that in turn requires analyzing more images in order to obtain similar statistics.

360



361

362 **Fig 6. Evaluation of the Caco-2 non-viable cells fraction by using *BioFilmAnalyzer* software**
363 **exemplified for 12 microscopic images.** The percentage of red-stained cells quantified by the
364 *BioFilmAnalyzer* software is shown in each panel



365

366

367 **Fig 7. Caco-2 cells count and live/dead ratio dependence on the image analysis threshold *T*.** Full
368 lines show the linear regression lines, while dashed line shows the ideal counting line as determined by
369 the manual analysis performed by several experts in visual microscopic image analysis. Panels A-C
370 show data for the 12 images presented in Fig 6. Panels D-F show data for the 87 microscopic images
371 taken randomly from different experiments with various imaging conditions

372

373

374 **Table 3. Regression coefficients *k*, the coefficients of determination R^2 and the accuracy**
375 **coefficient L^2 indicating the correspondence between the automatic and the manual counting for**
376 **the Caco-2 live/dead ratios on fluorescent images**

	12 images			87 images		
Threshold	k	R^2	L^2	k	R^2	L^2
15	nd	nd	nd	nd	nd	nd
20	0,86	0,69	0,80	1,03	0,73	1,00
30	0,92	0,53	0,93	1,00	0,76	1,00
45	0,88	0,46	0,83	0,92	0,71	0,97
60	0,87	0,40	0,81	0,99	0,73	1,00
70	0,91	0,54	0,92	0,97	0,71	1,00

377

378 To verify the accuracy of the overall procedure including the microscopy and automatic cells
379 counting, the treated cells were detached from the wells by trypsin treatment, stained with DioC6 and
380 ethidium bromide and aliquoted. One half of the sample was analyzed with flow cytometry to evaluate
381 the fraction of necrotic cells, while the other half was subjected to microscopy and quantified by
382 *BioFilmAnalyzer*. Fig 8 shows fractions of the non-viable cells in 4 independent repeats quantified with
383 flow cytometry versus automatic analysis of series of 10 microphotographs from each sample with
384 separate green and red channels (representative examples of images are shown in the figure).

385

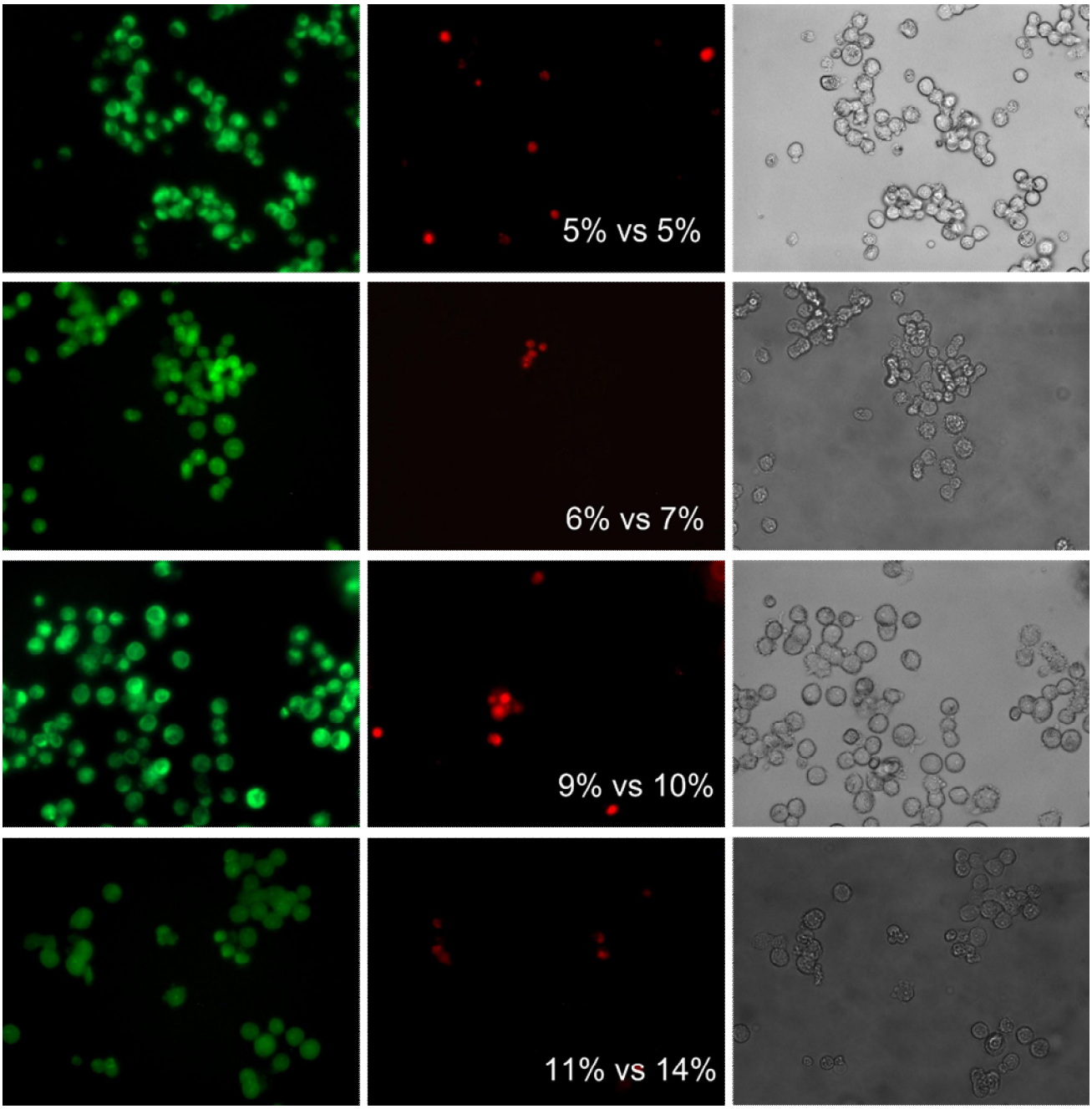


Fig 8. The fraction of the non-viable eukaryotic cells quantified by flow cytometry versus automatic analysis with *BioFilmAnalyzer*.

Quantification of confocal images with Z-stacks

Finally, we used the same algorithm and software to analyze the confocal images with Z-stacks, including 4 confocal images of *S.aureus* cells treated with different antimicrobials from our recent work [24]. For that, we analyzed the raw series of 2D images that were used previously to reconstruct 3D images in [24], where the fraction of live/dead cells were evaluated by both automatic and manual expert counting (Fig 9).

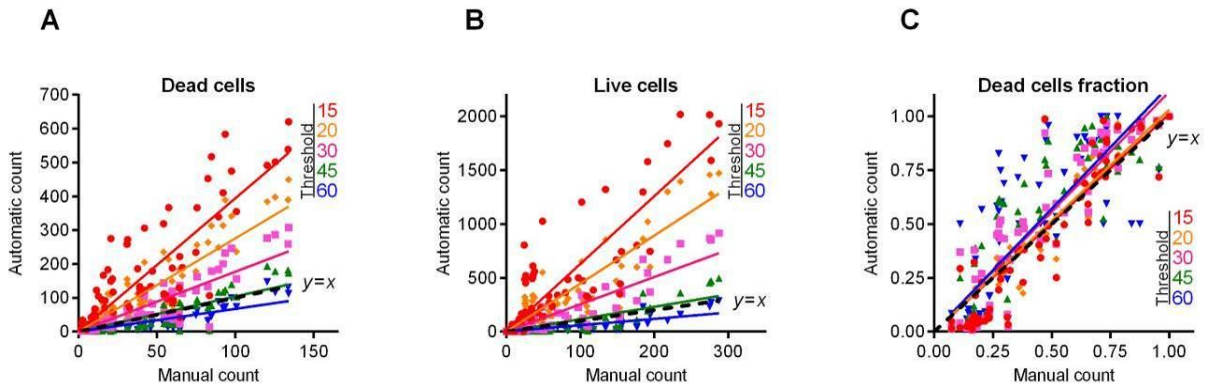


Fig 9. Automatic vs manual cell count cell count and live/dead ratio dependence for different image analysis thresholds *T*. Full lines show the linear regression lines, while dashed line shows the ideal counting line as determined by the manual analysis performed by several experts in visual microscopic image analysis. Four different confocal images containing between 12 and 23 Z-stacks each were taken for the analysis.

Typically 2d images obtained by confocal layer scanning technique are lower quality in comparison with single layer fluorescent microscopic images considered above. There are generally several contributing effects including defocus aberration of the cells that appear close but nevertheless displaced against the focal point due to the limited depth of focus as well as motion blur due to the

408 sample shift during operation. Both effects lead to blurred images, the problem that is partially resolved
409 in the 3d reconstruction algorithms by averaging or smoothening filters that improve the overall image
410 quality at the cost of its effective resolution. The question is, whether raw 2d images obtained by
411 confocal layer scanning technique can be used for the cell sub-population quantification using our
412 algorithm, and whether this would require some preliminary filtering to reduce blurring effects. For the
413 latter, two standard image filtering techniques, namely the Gaussian and the Sobel filters have been
414 tested. The Sobel filtering aims on edge detection in the images using a discrete differentiation operator
415 which computes an approximation of the gradient of the image intensity function. Therefore this kind
416 of image preparation might be helpful to make edges of cells more stepwise and thus to reduce the
417 dependence of the performance of the cell counting algorithm on the choice of the threshold that may
418 then appear anywhere within this step. Alternatively, blurring may be treated as effective additive noise
419 that could be reduced with simple Gaussian filter. Exhibiting a Gaussian impulse response, such filter
420 decreases the overall noise level in the image.

421 Figure 10 shows an example of the 2d confocal Z-stack microscopic image before and after
422 Gaussian and Sobel filtering (upper panel) and the regression functions of automatic cells count at
423 different thresholds as a function of the manual cell count (lower panel) before and after preliminary
424 filtering of the images (obtained for the entire cohort of studied confocal images). The figure shows
425 that, while the image appears visually less blurred, there is no significant improvement on cells fraction
426 count according to the regression analysis results (see also Table 4). Furthermore, preliminary filtering
427 leads to the overestimation of the red-stained cell count in some samples analyzed this way also
428 corrupting the overall sub-population fraction estimates. Therefore, we find that due to the general
429 robustness of the sub-population fraction estimation against the variations of the threshold T (and thus
430 also its relation with the quantile of the image color distribution), 2d images obtained by confocal layer
431 scanning technique can be analyzed without preliminary image filtering.

432

433 **Table 4. Regression coefficients *k*, the coefficients of determination R^2 and the accuracy**
434 **coefficient L^2 indicating the correspondence between the automatic and the manual counting for**
435 **the *S.aureus* live/dead ratios on confocal images.**

	No filters			Gaussian			Sobel		
Threshold	k	R^2	L^2	k	R^2	L^2	k	R^2	L^2
15	1,01	0,81	1,00	1,02	0,85	1,00	1,08	0,87	0,98
20	1,03	0,81	1,00	1,04	0,83	0,99	1,10	0,84	0,97
30	1,11	0,82	0,95	1,15	0,76	0,92	1,17	0,77	0,92
45	1,14	0,77	0,93	1,25	0,71	0,78	1,24	0,79	0,87
60	1,14	0,68	0,93	1,32	0,60	0,69	1,30	0,78	0,78

436

437

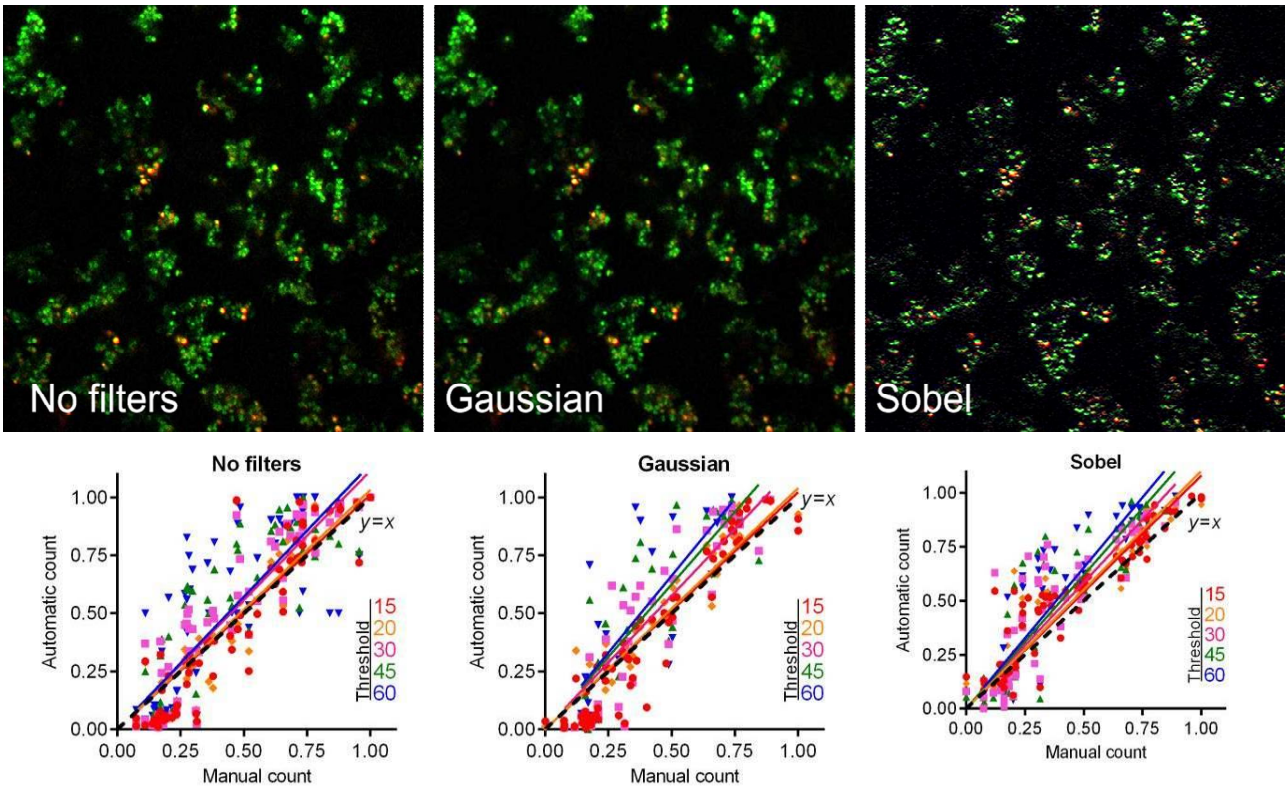


Fig 10. The live/dead ratio dependencies for different image analysis thresholds *T*. Full lines show the linear regression lines, while dashed line shows the ideal counting line as determined by the manual analysis performed by several experts in visual microscopic image analysis. Four confocal images containing between 12 and 23 Z-stacks were taken for the analysis, before and after being subjected to either Gaussian or Sobel filtering, as indicated.

449 # Conclusion

450 To summarize, our results indicate that the sub-population fraction estimates obtained from
451 fluorescent-stained cells imaging data by *BioFilmAnalyzer* are very close to the results obtained by
452 other techniques such as expert manual counting and flow cytometry. The two-step algorithm
453 implemented in the *BioFilmAnalyzer* that normalizes the stained image area in the units of the effective
454 single cell size that are determined under partial manual control by the investigator performs largely
455 independently of the cell shape and imaging conditions providing feasible results for cells aggregated
456 in clusters. Moreover, in most cases no preliminary preparation or filtering of raw images is required.

457 Thus, we suggest that the proposed algorithm implemented as the *BioFilmAnalyzer* software
458 can be used for numerous applications. First, preliminary evaluation of cell counts and live/dead ratios
459 can be quickly obtained without expertise in image processing. Second, the analysis of surface-
460 adherent bacterial or eukaryotic cells without their resuspension and maintenance of the native
461 distribution pattern on the surface is possible. Third, the quantification of cellular sub-populations from
462 2d confocal layer images is also possible. Fourth, the software allows quantification of the cells sub-
463 populations expressing fluorescent proteins in during long-time incubation. Finally, because no further
464 user intervention is required after few initial adjustment and cross-check procedures which usually take
465 a couple of minutes using few representative images, further processing can be done over coffee-break
466 by automated analysis of a series of images.

467 We believe that the suggested algorithm and software would be useful in saving time and efforts
468 in cells sub-population quantification from fluorescent microscopy data that is commonly required in
469 various biomedical, biotechnological and pharmacological studies, especially for the analysis of
470 biofilm-embedded cells where the performance of conventional techniques based on cell counting or
471 flow cytometry is strongly limited. Both the algorithm and the *BioFilmAnalyzer* can be freely

downloaded at http://kpfu.ru/eng/strau/laboratories/molecular-genetics-of-microorganisms-lab/software/biofilmanalyzer-v10, utilized and redistributed.

Acknowledgments

The financial support of this work was provided by the Russian Science Foundation (project No. 15-14-00046, AK), by the Ministry of Education and Science of the Russian Federation (St. Petersburg Electrotechnical University, assignment 2.5475.2017/6.7, MB) and by the Program of competitive development of Kazan Federal University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability: All data are within the manuscript. The algorithm proposed in this manuscript is implemented as a simple software tool entitled *BioFilmAnalyzer* is freely available online at http://kpfu.ru/eng/strau/laboratories/molecular-genetics-of-microorganisms-lab/software/biofilmanalyzer-v10

489 # References

490 1. Tawakoli PN, Al-Ahmad A, Hoth-Hannig W, Hannig M, Hannig C. Comparison of different
491 live/dead stainings for detection and quantification of adherent microorganisms in the initial oral
492 biofilm. Clinical Oral Investigations. 2013;17(3):841-50. doi: 10.1007/s00784-012-0792-3. PubMed
493 PMID: WOS:000316749300019.
494 2. McMullan BJ, Desmarini D, Djordjevic JT, Chen SCA, Roper M, Sorrell TC. Rapid
495 Microscopy and Use of Vital Dyes: Potential to Determine Viability of Cryptococcus neoformans in the
496 Clinical Laboratory. Plos One. 2015;10(1). doi: 10.1371/journal.pone.0117186. PubMed PMID:
497 WOS:000348821400028.
498 3. Atale N, Gupta S, Yadav UCS, Rani V. Cell-death assessment by fluorescent and nonfluorescent
499 cytosolic and nuclear staining techniques. Journal of Microscopy. 2014;255(1):7-19. doi:
500 10.1111/jmi.12133. PubMed PMID: WOS:000339710500002.
501 4. Netuschil L, Auschill TM, Sculean A, Arweiler NB. Confusion over live/dead stainings for the
502 detection of vital microorganisms in oral biofilms - which stain is suitable? Bmc Oral Health. 2014;14.
503 doi: 10.1186/1472-6831-14-2. PubMed PMID: WOS:000330071900001.
504 5. Gerstner AOH, Mittag A, Laffers W, Dahnert I, Lenz D, Bootz F, et al. Comparison of
505 immunophenotyping by slide-based cytometry and by flow cytometry. Journal of Immunological
506 Methods. 2006;311(1-2):130-8. doi: 10.1016/j.jim.2006.01.012. PubMed PMID:
507 WOS:000237527200013.
508 6. Mital J, Schwarz J, Taatjes DJ, Ward GE. Laser scanning cytometer-based assays for measuring
509 host cell attachment and invasion by the human pathogen Toxoplasma gondii. Cytometry Part A.
510 2006;69A(1):13-9. doi: 10.1002/cyto.a.20202. PubMed PMID: WOS:000234383700002.
511 7. Adan A, Alizada G, Kiraz Y, Baran Y, Nalbant A. Flow cytometry: basic principles and
512 applications. Critical Reviews in Biotechnology. 2017;37(2):163-76. doi:
513 10.3109/07388551.2015.1128876. PubMed PMID: WOS:000392639500003.
514 8. van der Pol E, Hoekstra AG, Sturk A, Otto C, van Leeuwen TG, Nieuwland R. Optical and non-
515 optical methods for detection and characterization of microparticles and exosomes. Journal of
516 Thrombosis and Haemostasis. 2010;8(12):2596-607. doi: 10.1111/j.1538-7836.2010.04074.x. PubMed
517 PMID: WOS:000285110700003.
518 9. Heydorn A, Nielsen AT, Hentzer M, Sternberg C, Givskov M, Ersboll BK, et al. Quantification
519 of biofilm structures by the novel computer program COMSTAT. Microbiology-Sgm. 2000;146:2395-
520 407. PubMed PMID: WOS:000089985400007.

521 10. Beyenal H, Donovan C, Lewandowski Z, Harkin G. Three-dimensional biofilm structure
522 quantification. Journal of Microbiological Methods. 2004;59(3):395-413. doi:
523 10.1016/j.mimet.2004.08.003. PubMed PMID: WOS:000224843300009.
524 11. Merritt JH, Kadouri DE, O'Toole GA. Growing and analyzing static biofilms. Curr Protoc
525 Microbiol. 2005;Chapter 1:Unit 1B. doi: 10.1002/9780471729259.mc01b01s00. PubMed PMID:
526 18770545; PubMed Central PMCID: PMCPMC4568995.
527 12. Ji L, Piper J, Tang JY. Erosion and dilation of binary images by arbitrary structuring elements
528 using interval coding. Pattern Recognition Letters. 1989;9(3):201-9. doi: 10.1016/0167-
529 8655(89)90055-x. PubMed PMID: WOS:A1989U514000008.
530 13. Adams R, Bischof L. SEEDED REGION GROWING. Ieee Transactions on Pattern Analysis
531 and Machine Intelligence. 1994;16(6):641-7. doi: 10.1109/34.295913. PubMed PMID:
532 WOS:A1994NR97200008.
533 14. Nattkemper TW, Twellmann T, Ritter H, Schubert W. Human vs. machine: evaluation of
534 fluorescence micrographs. Computers in Biology and Medicine. 2003;33(1):31-43. doi: 10.1016/s0010-
535 4825(02)00060-4. PubMed PMID: WOS:000180493700003.
536 15. Tchoukalova YD, Harteneck DA, Karwoski RA, Tarara J, Jensen MD. A quick, reliable, and
537 automated method for fat cell sizing. Journal of Lipid Research. 2003;44(9):1795-801. doi:
538 10.1194/jlr.D300001-JLR200. PubMed PMID: WOS:000185522300024.
539 16. Daims H, Lucker S, Wagner M. daime, a novel image analysis program for microbial ecology
540 and biofilm research. Environmental Microbiology. 2006;8(2):200-13. doi: 10.1111/j.1462-
541 2920.2005.00880.x. PubMed PMID: WOS:000234647600003.
542 17. Lempitsky V, Rother C, Roth S, Blake A. Fusion Moves for Markov Random Field
543 Optimization. Ieee Transactions on Pattern Analysis and Machine Intelligence. 2010;32(8):1392-405.
544 doi: 10.1109/tpami.2009.143. PubMed PMID: WOS:000278858600004.
545 18. Klinger-Strobel M, Suesse H, Fischer D, Pletz MW, Makarewicz O. A Novel Computerized Cell
546 Count Algorithm for Biofilm Analysis. Plos One. 2016;11(5). doi: 10.1371/journal.pone.0154937.
547 PubMed PMID: WOS:000375676800095.
548 19. Abramoff MD, Magalhães PJ, Ram SJ. Image processing with ImageJ2004; 11(7):[36-42 pp.].
549 20. Selinummi J, Seppala J, Yli-Harja O, Puhakka JA. Software for quantification of labeled
550 bacteria from digital microscope images by automated image analysis. Biotechniques. 2005;39(6):859-
551 63. doi: 10.2144/000112018. PubMed PMID: WOS:000234104900011.
552 21. Kayumov AR, Nureeva AA, Trizna EY, Gazizova GR, Bogachev MI, Shtyrlin NV, et al. New

553 Derivatives of Pyridoxine Exhibit High Antibacterial Activity against Biofilm-Embedded
554 Staphylococcus Cells. Biomed Research International. 2015:10. doi: 10.1155/2015/890968. PubMed
555 PMID: WOS:000367944100001.
556 22. Trizna E, Latypova L, Kurbangalieva A, Bogachev MI, Kayumov A. 2(5H)-Furanone
557 Derivatives as Inhibitors of Staphylococcal Biofilms. Bionanoscience. 2016;6(4):423-6. doi:
558 10.1007/s12668-016-0258-1. PubMed PMID: WOS:000390587700036.
559 23. Trizna EY, Khakimullina EN, Latypova LZ, Kurbangalieva AR, Sharafutdinov IS, Evtyugin
560 VG, et al. Thio Derivatives of 2(5H)-Furanone As Inhibitors against Bacillus subtilis Biofilms. Acta
561 Naturae. 2015;7(2):102-7. PubMed PMID: WOS:000355757100013.
562 24. Baidamshina DR, Trizna EY, Holyavka MG, Bogachev MI, Artyukhov VG, Akhatova FS, et al.
563 Targeting microbial biofilms using Ficin, a nonspecific plant protease. Scientific Reports. 2017;7. doi:
564 10.1038/srep46068. PubMed PMID: WOS:000398641900002.
565
566
567