microbeSEG: Accurate Cell Segmentation with OMERO Data Management

Tim Scherr[1], Johannes Seiffarth[2,3], Bastian Wollenhaupt[2], Oliver Neumann[1], Marcel P. Schilling[1], Dietrich Kohlheyer[2], Hanno Scharr[4,5], Katharina Nöh[2], and Ralf Mikut[1]

[1] Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, [2] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, [3] Computational Systems Biology (AVT.CSB), RWTH Aachen University, [4] Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, [5] Institute for Advanced Simulation, IAS-8: Data Analytics and Machine Learning, Forschungszentrum Jülich GmbH

Abstract

In biotechnology, cell growth is one of the most important properties for the characterization and optimization of microbial cultures. Novel live-cell imaging methods are leading to an ever better understanding of cell cultures and their development. The key for analyzing acquired data is accurate and automated cell segmentation at the single-cell level. Therefore, we present microbeSEG, a user-friendly cell segmentation tool with OMERO data management. microbeSEG utilizes a state-of-the-art deep learning-based segmentation method and can be used for instance segmentation of a wide range of cell morphologies and imaging techniques, e.g., phase contrast or fluorescence microscopy. The main focus of microbeSEG is a comprehensible, easy, efficient, and complete workflow from creation of training data to final application of the trained segmentation model. We demonstrate that accurate cell segmentation results can be obtained within 45 minutes of user time. Utilizing public segmentation data sets or pre-labeling further accelerates the microbeSEG workflow. This opens the door for accurate and efficient data analysis of microbial cultures.

Introduction

Cell-to-cell heterogeneity induced by intrinsic biological factors and extrinsic environmental fluctuations highly affects microbial cell growth and productivity of microbes.[1–6] State-of-the-art microfluidic lab-on-chip systems with highly precise environmental control and time-lapse imaging enable to investigate causes and consequences of such heterogeneity at the

single-cell level.[7–9] Interesting growth parameters include number of cells over time, or size distributions. To extract such quantities from the acquired image sequences, all cells in all images need to be individually segmented; a process that needs to be highly automated and reliable due to the high numbers of cells (up to several tens of thousands of cells per experiment). However, virtually error-free segmentation of acquired time-lapse images remains a challenge, as contrast and signal-to-noise ratio are typically low due to limited lighting conditions to avoid phototoxic stress.

Cell segmentation tools are therefore often highly specialized to the specific imaging conditions and need to be adapted when acquisition settings or experimental parameters change, or to conform with other cell morphologies or image modalities. When the tools are based on traditional image processing methods, as e.g. ChipSeg,[10] some adaptations can be achieved by manually adjusting parameters. Unfortunately, this often requires some expert knowledge on the underlying method and in some cases even code adaptations. In contrast, deep learning-based methods are not tuned manually and much more versatile, but need to be (re-)trained on annotated data. Due to this versatility deep learning methods nowadays dominate cell segmentation challenges covering multiple object morphologies and imaging techniques, e.g., the 2018 Data Science Bowl.[11]

Training a deep learning-based method involves a pipeline consisting of multiple steps usually covered by a variety of sequentially applied tools or programming steps. A typical workflow starts with training data creation, data handling and loading, model training, model evaluation – these steps may be iterated until the model is suitably accurate – and finally the application of trained models to experiment data. From a user perspective integrating all needed steps in a single, highly automated toolbox is desirable, where especially data management, the training data creation, and the training of the deep learning models need to be user-friendly and time efficient, e.g., by using pre-labeling and subsequent manual corrections, and full automation of routine steps. Thus far, deep learning approaches for the segmentation of biological data either lack a smooth training data creation and retraining workflow, or a data management system like OMERO.[12] Recent examples are Misic for the high-throughput cell segmentation of complex bacterial communities,[13] and Omnipose for robust segmentation of bacteria and other elongated cell shapes.[14]

In this paper, we present microbeSEG, a deep learning-based segmentation tool which uses the freely available OMERO[12] for data management. microbeSEG is a tool tailored to instance segmentation of various cell morphologies and imaging techniques. In contrast to other solutions, microbeSEG covers the whole processing pipeline needed for image analysis: users can easily create training data sets, annotate cells with the jointly developed annotation toolkit ObiWan-Microbi,[15] and train and apply deep learning models. Due to the use

Table 1: microbeSEG accuracy in dependence of annotation time. All models are evaluated on 12 *B. subtilis* and 12 *E. coli* test images. For each setting, the median aggregated Jaccard index AJI+ out of five trained models is shown (default microbeSEG settings). The times include the crop selection times. The AJI+ score ranges from 0 to 1, with 1 indicating perfect segmentation.

Manual annotation time	N_{crops}	N_{cells}	AJI+ (*B. subtilis*)	AJI+ (*E. coli*)	AJI+ (total)
30 min†	13	309	0.585	0.729	0.657
30 min† + 15 min‡	18	732	0.614	0.731	0.673
Omnipose	-	-	0.478	0.519	0.498

† Same training data set.

‡ With pre-labeling using a model trained on the data set annotated in the first 30 minutes.

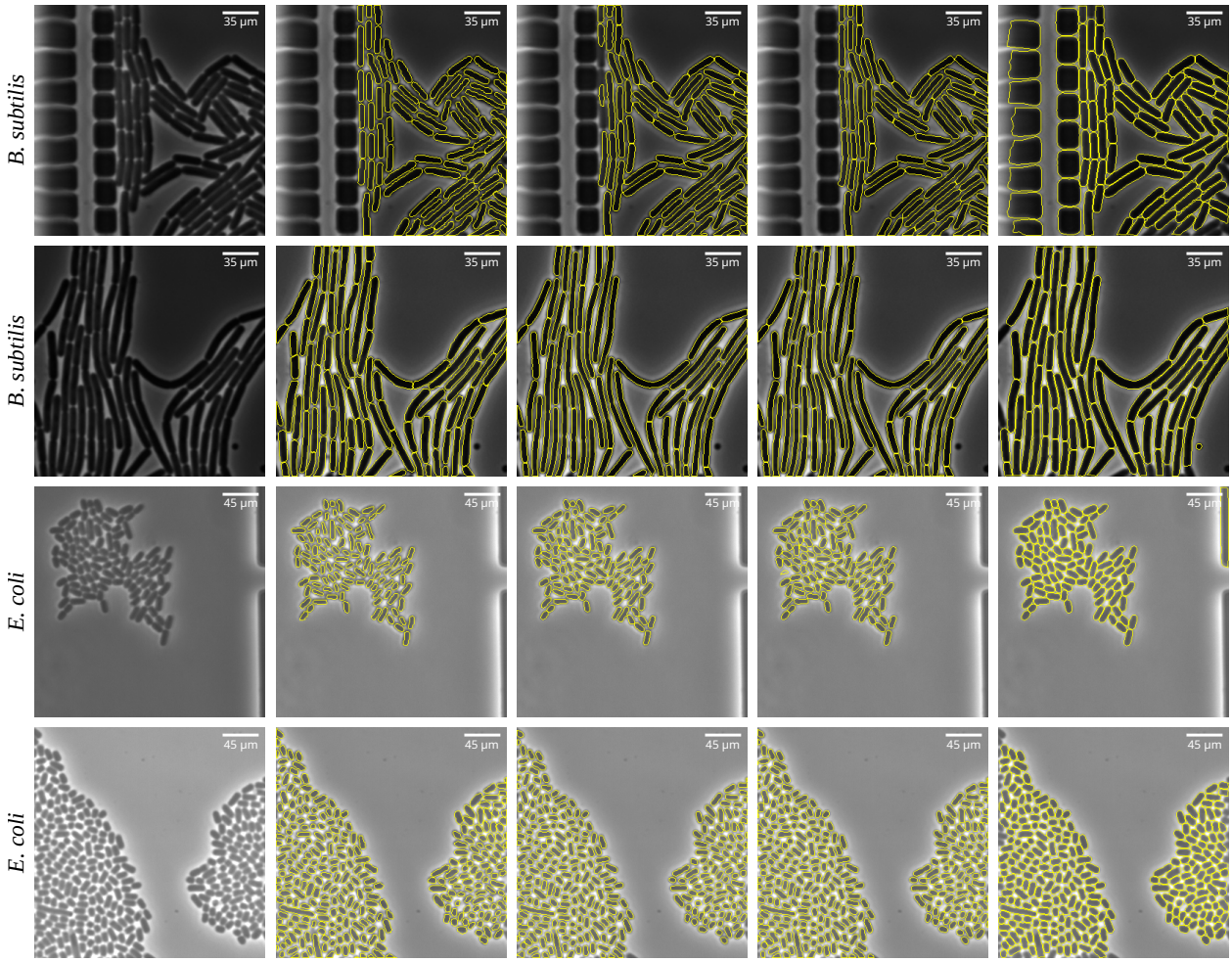
of OMERO and its versatile data importer, supporting over 150 image formats, data management is straightforward and user-friendly. microbeSEG provides powerful deep learning-based cell segmentation in an easy-to-use tool. The open-source software is implemented in Python and publicly available at `https://github.com/hip-satomi/microbeSEG`.

Results

Manual Annotation and Pre-Labeling

Table 1 shows the microbeSEG segmentation accuracy in terms of the aggregated Jaccard index AJI+ in dependence of the annotation time on a test set consisting of two microbe types, i.e., of 721 *B. subtilis* and 1168 *E. coli* cells. The 24 test images of size $320\,\text{px} \times 320\,\text{px}$, 12 for each species, were annotated by three experts, who cross-checked their annotations. The microbeSEG user selected and annotated crops from different experiment data than the experts have used.

microbeSEG produces reasonable segmentation results for the two microbe types within 30 minutes annotation time, even without utilizing public data sets, pre-training or pre-labeling. Fig. 1 shows exemplary segmentations. Switching from completely manual annotation to the correction of pre-labeled image crops after the initial annotation phase results in a higher number of cells annotated per time. Therefore, a model trained on the data set annotated in the first 30 minutes has been used. This increases the scores for both cell types further. For comparison, Omnipose,[14] a deep learning method pre-trained on a large database of bacterial images, yields convincing segmentations of cells, but without re-training suffers from the detection of background structures.



(a) Test image (b) Ground truth (c) microbeSEG (30min ann.) (d) microbeSEG (30min + 15min ann.) (e) Omnipose

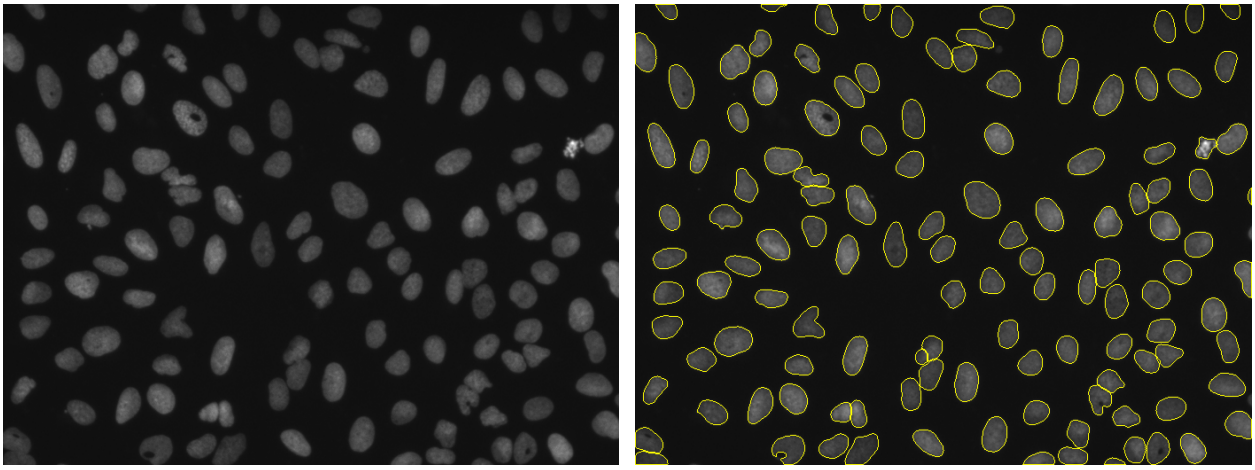
Figure 1: Exemplary images and segmentation overlays for *B. subtilis* and *E.coli*. Shown are the results of the median microbeSEG models reported in Table 1.

Data Import

The import of annotated data sets helps to further decrease manual interaction time, or even make the need for manual annotations or corrections obsolete. Fig. 2 shows qualitative segmentation results of U2OS cells[16] with a microbeSEG model trained on HeLa cell data from the Cell Tracking Challenge.[17,18] Even without annotated U2OS cells, a good segmentation quality is achieved.

Data Analysis

For time-lapse experiments, basic analysis results, i.e., cell counts, mean cell area, mean cell minor/major axis length and total area can be exported after segmentation. Possibly



(a) Image (b) Image-segmentation overlay

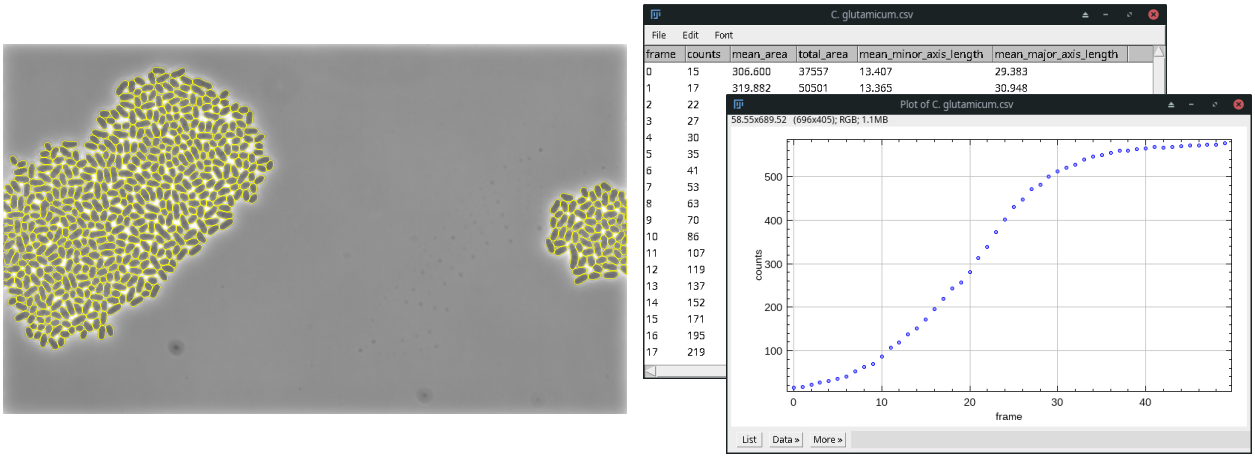
Figure 2: Segmentation of U2OS cells from the BBBC039 data set.[16] The microbeSEG segmentation model has been trained on HeLa cells from the Cell Tracking Challenge.[17,18]

needed manual corrections to reach a virtually error-free segmentation can be applied using ObiWan-Microbi. The results can easily be plotted using Fiji,[19] cf. Fig 3.

Discussion

Utilizing ObiWan-Microbi for training data annotation, microbeSEG covers the whole pipeline from training data creation, training, evaluation and inference. Both tools are designed to work together seamlessly and use OMERO for the data management. The optimized workflow allows producing high-quality segmentation results in reasonable annotation time. Our results indicate that switching to pre-label corrections is a better strategy than annotating manually when suitably accurate models are trained, since more cells can be annotated per time. In our experiments we observed, that it is a good strategy to correct mainly object level errors, i.e., merges, splits, and added or missing cells. The AJI+ scores for *B. subtilis* are mainly limited by size differences and different handling of cell division events, as the qualitative segmentation results show. This is due to different decision boundaries of the microbeSEG user and the single annotators in this experiment, especially for the cell division events of the challenging filamentous *B. subtilis* cells.

The Omnipose segmentation results show that available software solutions may need retraining even when they are pre-trained on a large data set, e.g., to avoid the segmentation of background structures on a microfluidic device. However, such solutions can be utilized to further reduce the manual annotation time. For instance, ObiWan-Microbi includes Omnipose pre-labeling. Prerequisite is that Omnipose works for the present data. A drawback



(a) Image-segmentation overlay (frame 46)

(b) Analysis results opened in Fiji

Figure 3: Exemplary microbeSEG segmentation (a) and analysis results (b) for a growing *C. glutamicum* colony. The results can easily be viewed in Fiji.

is, however, that object shapes may not be matched as good as with manual annotation and the microbeSEG pre-labeling.

Once a large and diverse annotated data set is established, further annotation and model training is required only for significant changes in the experimental setup or for specific cases where the segmentation fails. Thereby, the microbeSEG crop selection with pre-labeling allows adding exactly those new training data where segmentation errors occur.

In summary, we believe that the major concern about deep learning for cell segmentation, that extensive and time-consuming annotation of training data is required, has been disproved. Certainly, microbeSEG can also be applied for the segmentation of cell nuclei or other objects. To further facilitate the data annotation for microbeSEG users, label inspection strategies[20] and training data simulators are of particular interest. A long-term goal is to leverage deep learning for 3D applications, e.g., to study plant-microbe interactions.[21]

Materials and Methods

OMERO

OMERO is an open-source software platform for access and use of a wide range of biological data and provides a unified interface for images, matrices and tables.[12] Over 150 image formats can be imported. Data are organized in projects and data sets. With the OMERO.web client, data can easily be accessed and viewed in the browser.

Segmentation Methods

In microbeSEG, two deep learning-based instance segmentation methods are implemented: (i) a simple semantic-based method with multi-class output, i.e., background, cell interior and cell boundary, and (ii) a distance transform-based method.[22,23] The latter method (ii) has proven its applicability for phase contrast, bright field and fluorescence images for diverse cell morphologies in the Cell Tracking Challenge (http://celltrackingchallenge.net/).[22,24] The simple boundary method is easy to interpret and serves as a baseline.

Training Data Creation

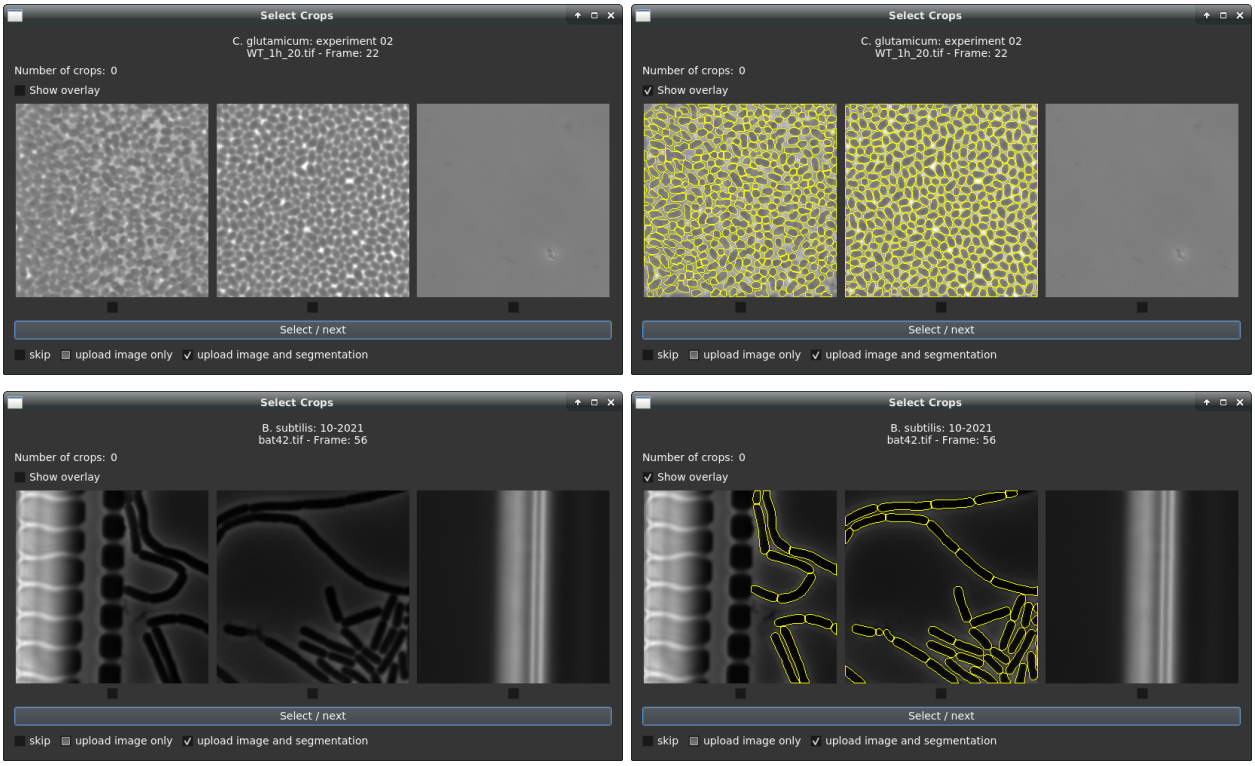
microbeSEG training sets are managed as OMERO data sets. When adding a new training set, a crop size needs to be selected. This is required since annotating whole images or frames is tedious and time-consuming, e.g., for frames with microbial colonies of hundreds or thousands of cells. More efficient is to annotate smaller crops showing less cells. This enables to cover more image and cell features in the same annotation time and results in a more diverse training set. Therefore, crop proposals of selected files can be viewed and added to the training set, cf. Figure 4 (a). When trained models are already available, pre-labeling enables to identify regions where the segmentation fails and improvements are needed, cf. Figure 4 (b). In addition, this approach reduces the annotation time for correctly or partially correctly segmented cells. Selected crops are automatically assigned to a training, validation or test subset. However, there is also an option for explicit assignment. The crops can then be annotated using ObiWan-Microbi. Annotated data sets, e.g., publicly available training data, can be imported. With the OMERO.web client, the training data and the annotations can be viewed as well.

Training

For training on a selected annotated training set, the segmentation method, the batch size, the optimizer and how many models should be trained needs to be specified. Adam[25] and Ranger[26] are available as optimizer. Usually, the default settings are sufficient but the batch size may be reduced if memory availability is limited. A detailed description of the training process is provided by Scherr et al.[22,23]

Evaluation

Trained models can be evaluated on the automatically split test set. Thereby, appropriate parameters are set for the distance method (ii), i.e., a cell size adjustment and a seed extrac-



(a) Without overlay (b) With overlay

Figure 4: Crop creation interface. Automatically proposed crops can be selected and uploaded to OMERO (a). For the pre-labeling, it is possible to upload only the image or the image with its prediction (b). The image only upload is helpful when the pre-label predictions require too much manual corrections.

tion threshold. As evaluation measure, a slightly modified version of the aggregated Jaccard index AJI[27] is used. The modification AJI+ prevents overpenalization by using a one-to-one mapping instead of a one-to-many mapping for the predicted objects.[28,29] The AJI+ score ranges from 0 to 1, with 1 indicating perfect segmentation. The mean AJI+ score over the single test images and its standard deviation of each evaluated model are saved in a csv file.

For the performance evaluation in the results section, the parameters have been adjusted automatically on the test set of the microbeSEG user. The AJI+ scores have been calculated on the external test set annotated by experts, for which different experiments have been used.

Inference

For inference, the best evaluated model with its parameter set is selected automatically. However, the user can also select the model manually. If a not yet evaluated model is selected, standard thresholds are used for the distance method (ii) which may result in (slightly) too large or too small cells for method (ii). The segmentation results can be attached to the

corresponding OMERO image as polygon region of interests and can be viewed with the OMERO.web client.

Data

Data shown in this paper were acquired with an fully automated time-lapse phase contrast microscope setup using a 100x oil immersion objective. Cultivation took place inside a special microfluidic cultivation device.[30] The assembled training and test sets are available online at https://doi.org/10.5281/zenodo.6497715.

Author Contributions

K.N. and H.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

Acknowledgement

We are grateful for the funding by the Helmholtz Association in the programs Natural, Artificial and Cognitive Information Processing (T.S., R.M.), Engineering Digital Futures: Supercomputing, Data Management and Information Security for Knowledge and Action (H.S.), HIDSS4Health - the Helmholtz Information & Data Science School for Health (R.M.), and the Helmholtz Association Initiative and Networking Funds through Helmholtz AI (O.N., R.M.) as well as the Helmholtz Imaging Platform within the project SATOMI (J.S., H.S., D.K., K.N., R.M.). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491111487 (K.N.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

(1) Lara, A.; Galindo, E.; Ramírez, O.; Palomares, L. Living with heterogeneities in bioreactors: Understanding the effects of environmental gradients on cells. *Molecular Biotechnology* **2006**, *34*, 355–381.

(2) Carlquist, M.; Fernandes, R.; Helmark, S.; Heins, A.-L.; Lundin, L.; Sørensen, S. J.; Gernaey, K. V.; Lantz, A. E. Physiological heterogeneities in microbial populations and implications for physical stress tolerance. *Microb Cell Fact* **2012**, *11*.

(3) Takors, R. Scale-up of microbial processes: Impacts, tools and open questions. *Journal of Biotechnology* **2012**, *160*, 3–9.

(4) Delvigne, F.; Goffin, P. Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. *Biotechnology Journal* **2014**, *9*, 61–72.

(5) Lipson, D. A. The complex relationship between microbial growth rate and yield and its implications for ecosystem processes. *Frontiers in Microbiology* **2015**, *6*, 615.

(6) Mariano, G.; Monlezun, L.; Coulthurst, S. J. Dual role for DsbA in attacking and targeted bacterial cells during type VI secretion system-mediated competition. *Cell Reports* **2018**, *22*, 774–785.

(7) Grünberger, A.; Wiechert, W.; Kohlheyer, D. Single-cell microfluidics: Opportunity for bioprocess development. *Current Opinion in Biotechnology* **2014**, *29*, 15–23.

(8) Helfrich, S.; Pfeifer, E.; Krämer, C.; Sachs, C. C.; Wiechert, W.; Kohlheyer, D.; Nöh, K.; Frunzke, J. Live cell imaging of SOS and prophage dynamics in isogenic bacterial populations. *Molecular Microbiology* **2015**, *98*, 636–650.

(9) Rothbauer, M.; Zirath, H.; Ertl, P. Recent advances in microfluidic technologies for cell-to-cell interaction studies. *Lab Chip* **2018**, *18*, 249–270.

(10) de Cesare, I.; Zamora-Chimal, C. G.; Postiglione, L.; Khazim, M.; Pedone, E.; Shannon, B.; Fiore, G.; Perrino, G.; Napolitano, S.; di Bernardo, D.; Savery, N. J.; Grierson, C.; di Bernardo, M.; Marucci, L. ChipSeg: An automatic tool to segment bacterial and mammalian cells cultured in microfluidic devices. *ACS Omega* **2021**, *6*, 2473–2476.

(11) Caicedo, J. C.; Goodman, A.; Karhohs, K. W.; Cimini, B. A.; Ackerman, J.; Haghighi, M.; Heng, C.; Becker, T.; Doan, M.; McQuin, C.; Rohban, M.; Sing, S.; Carpenter, A. E. Nucleus segmentation across imaging experiments: The 2018 Data Science Bowl. *Nat. Methods* **2019**, *16*, 1247–1253.

(12) Allan, C. et al. OMERO: Flexible, model-driven data management for experimental biology. *Nat. Methods* **2012**, *9*, 245–253.

(13) Panigrahi, S.; Murat, D.; Le Gall, A.; Martineau, E.; Goldlust, K.; Fiche, J.-B.; Rombouts, S.; Nöllmann, M.; Espinosa, L.; Mignot, T. Misic, a general deep learning-based method for the high-throughput cell segmentation of complex bacterial communities. *eLife* **2021**, *10*, e65151.

(14) Cutler, K. J.; Stringer, C.; Wiggins, P. A.; Mougous, J. D. Omnipose: A high-precision morphology-independent solution for bacterial cell segmentation. *bioRxiv* **2021**,

(15) Seiffarth, J.; Scherr, T.; Wollenhaupt, B.; Neumann, O.; Scharr, H.; Kohlheyer, D.; Mikut, R.; Nöh, K. ObiWan-Microbi: OMERO-based integrated workflow for annotating microbes in the cloud. https://github.com/hip-satomi/ObiWan-Microbi, 2022; to be submitted.

(16) Ljosa, V.; Sokolnicki, K.; Carpenter, A. Annotated high-throughput microscopy image sets for validation. *Nat. Methods* **2012**, *9*, 637.

(17) Maška, M. et al. A benchmark for comparison of cell tracking algorithms. *Bioinformatics* **2014**, *30*, 1609–1617.

(18) Ulman, V. et al. An objective comparison of cell-tracking algorithms. *Nat. Methods* **2017**, *14*, 1141–1152.

(19) Schindelin, J. et al. Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*.

(20) Schilling, M. P.; Scherr, T.; Münke, F. R.; Neumann, O.; Schutera, M.; Mikut, R.; Reischl, M. Automated annotator variability inspection for biomedical image segmentation. *IEEE Access* **2022**, *10*, 2753–2765.

(21) Peredo, E. L.; Simmons, S. L. Leaf-FISH: Microscale imaging of bacterial taxa on phyllosphere. *Frontiers in Microbiology* **2018**, *8*, 2669.

(22) Scherr, T.; Löffler, K.; Böhland, M.; Mikut, R. Cell segmentation and tracking using CNN-based distance predictions and a graph-based matching strategy. *PLOS ONE* **2020**, *15*, 1–22.

(23) Scherr, T.; Löffler, K.; Neumann, O.; Mikut, R. On improving an already competitive segmentation algorithm for the Cell Tracking Challenge - lessons learned. *bioRxiv* **2021**,

(24) Löffler, K.; Scherr, T.; Mikut, R. A graph-based cell tracking algorithm with few manually tunable parameters and automated segmentation error correction. *PLOS ONE* **2021**, *16*, 1–28.

(25) Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**,

(26) Wright, L. Ranger - a synergistic optimizer. https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer, 2019.

(27) Kumar, N.; Verma, R.; Sharma, S.; Bhargava, S.; Vahadane, A.; Sethi, A. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging* **2017**, *36*, 1550–1560.

(28) Vu, Q. D.; Graham, S. HoVer-Net. https://github.com/vqdang/hover_net, 2021.

(29) Graham, S.; Vu, Q. D.; Raza, S. E. A.; Azam, A.; Tsang, Y. W.; Kwak, J. T.; Rajpoot, N. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis* **2019**, *58*, 101563.

(30) Kaganovitch, E.; Steurer, X.; Dogan, D.; Probst, C.; Wiechert, W.; Kohlheyer, D. Microbial single-cell analysis in picoliter-sized batch cultivation chambers. *New Biotechnology* **2018**, *47*, 50–59.