ColiCoords: A Python package for the analysis of bacterial fluorescence microscopy data

Jochem H. Smit[1¤]*, Yichen Li[1], Eliza M. Warszawik[2], Andreas Herrmann[2,3,4], Thorben Cordes[1,5]*

1 Molecular Microscopy Research Group, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands, 2 Department of Polymer Chemistry, Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands, 3 DWI – Leibniz Institute for Interactive Materials, Forckenbeckstr. 50, 52056 Aachen, Germany, 4 Institute of Technical and Macromolecular Chemistry, RWTH Aachen University, Worringerweg 2, 52074 Aachen, Germany, 5 Physical and Synthetic Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Großhadernerstr. 2-4, 82152 Planegg-Martinsried, Germany

¤ Current address: Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory for Molecular Bacteriology, KU Leuven, Herestraat 49, Gasthuisberg Campus, B-3000 Leuven, Belgium
* jhsmit@gmail.com (JHS); cordes@bio.lmu.de (TC)

OPEN ACCESS

Abstract

Single-molecule fluorescence microscopy studies of bacteria provide unique insights into the mechanisms of cellular processes and protein machineries in ways that are unrivalled by any other technique. With the cost of microscopes dropping and the availability of fully automated microscopes, the volume of microscopy data produced has increased tremendously. These developments have moved the bottleneck of throughput from image acquisition and sample preparation to data analysis. Furthermore, requirements for analysis procedures have become more stringent given the demand of various journals to make data and analysis procedures available. To address these issues we have developed a new data analysis package for analysis of fluorescence microscopy data from rod-like cells. Our software ColiCoords structures microscopy data at the single-cell level and implements a coordinate system describing each cell. This allows for the transformation of Cartesian coordinates from transmission light and fluorescence images and single-molecule localization microscopy (SMLM) data to cellular coordinates. Using this transformation, many cells can be combined to increase the statistical power of fluorescence microscopy datasets of any kind. ColiCoords is open source, implemented in the programming language Python, and is extensively documented. This allows for modifications for specific needs or to inspect and publish data analysis procedures. By providing a format that allows for easy sharing of code and associated data, we intend to promote open and reproducible research. The source code and documentation can be found via the project's GitHub page.

1 Introduction

Fluorescence microscopy has become a crucial tool in studying bacterial cell biology [1–8]. It is minimally invasive, allows for the study of living bacteria in a controlled environment, and enables the motion and sub-cellular topologies of any proteinaceous factor [9–24] or nucleic acids [25–28] to be monitored. Through the multitude of genetically programmable fluorescent protein probes [29–36] and commercially available dyes [37–39, 39–41] with different conjugation capabilities [42–46], virtually all components of bacterial cellular machinery can be studied with high specificity and spatio-temporal resolution. Even the detection of single fluorescent probes has become a routine experiment, revealing both dynamics and heterogeneity.

However, with the introduction of fully automated [47] and autonomously operating microscopes [48, 49], images can be acquired at increasingly high acquisition rates [50–52]. A typical experiment conducted overnight spanning the dimensions of position, time and channels can easily generate thousands of images. This explosion of available multidimensional data has made analysis the new bottleneck in single-cell fluorescence microscopy studies. Furthermore, in the spirit of reproducibility and open availability of scientific results, more and more peer-reviewed journals require deposition of source data and evaluation methods—a process that is hampered by non-standardized evaluation routines and lab-specific software, often based on commercial platforms.

The exception is current options for bacterial image analysis such as SuperSegger [53, 54], Oufti [55] and MicrobeJ [56]. Although these data analysis tools are able to tackle a lot of the problems inherent to (bacterial) live-cell image analysis, we have identified several drawbacks which we aimed to address with ColiCoords. These drawbacks are related to the platform, structure and philosophy of the data analysis routines, rather than their functionality for analysing images.

First, with available tools, it is difficult for users themselves to either inspect the exact mathematical procedures applied to their data or customize them. The source code of these analysis packages are freely available and licensed under the GNU General Public License. Modification of Oufti and SuperSegger, however, require a Matlab license, which is not available in many institutions. MicrobeJ, on the other hand, can be freely modified, but users face a daunting task due to the shear size of the project and the lack of docstrings.

ColiCoords is written in the freely available language Python and the source code has been released under the MIT license. The code is available on GitHub, is extensively documented with both docstrings and online documentation, and its basic principles are described in this paper. Being hosted on GitHub together with Continuous Integration for testing makes ColiCoords an ideal platform for other users to contribute and modify its code.

Second, most other currently available analysis options are based on a graphical user interface (GUI). We would argue that GUIs are inherently limiting and obfuscate the exact data analysis procedure. While GUI-based analysis is intuitive, it limits the use to the available graphical elements, with little flexibility. Importantly, due to the large number of permutations in the order of operations that can be executed via a GUI, the exact data analysis procedure is typically not documented in detail and thus hard to reproduce.

In contrast, ColiCoords offers a basic foundation for single cell data analysis where users can apply all functionality from Python itself, the SciPy [57] ecosystem, and community developed image analysis procedures. The workflow of ColiCoords is implemented in Jupyter Notebooks [58, 59], which act as a lab journal page, and in which all steps of data analysis and its associated parameters can be executed and described, and the results plotted in interactive graphs. This type of interactive analysis workflow is being adopted rapidly in the scientific
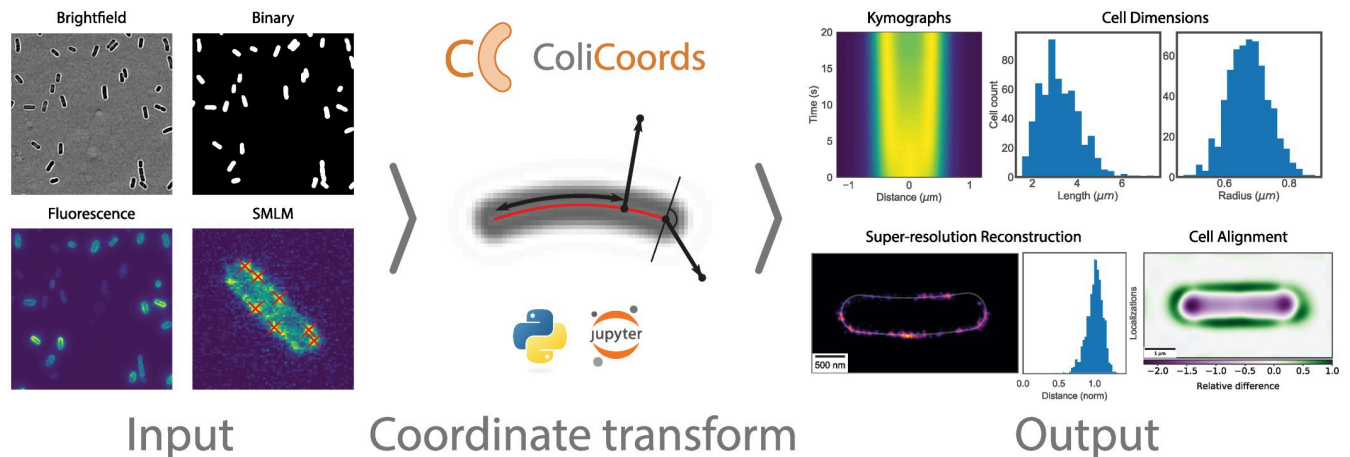
Fig 1. Overview of workflow pipeline using ColiCoords. Input data is either image data or sparse data (localizations). Input images need to be segmented to identify cell location and orientation. Third party options for image segmentation include Ilastik [63], CellProfiler [62, 70] or Keras [65]/Tensorflow [64]. SMLM data have to be reconstructed by external software such as DAOSTORM [66, 67], ThunderSTORM [68], QuickPALM [69] or others prior to use. ColiCoords can then be used to transform the Cartesian coordinates of the input data to cellular coordinates. The transformed data can be used to generate output graphs, such as kymographs, histograms of the cell's dimensions, axial distributions or to align cells.

community [60]. An example from the microscopy community is the smFRET analysis package FRETBursts [61].

In addition, ColiCoords is open-ended, meaning it can be used in a pipeline of analysis where users can freely choose pre- and post-processing options (Fig 1). The input data can be either image-based (brightfield, phase contract, fluorescence) or sparse data (super-resolution localizations). Users are therefore free to choose preprocessing options for tasks such as segmentation [55, 62–65] and super-resolution reconstruction [66–69]. Colicoords is out of the box compatible with High-performance computing (HPC) as operations on a per-cell basis can be computed in parallel.

The central feature of bacterial image analysis in ColiCoords is transforming input Cartesian coordinates (image pixel coordinates or super-resolution localizations) to cellular coordinates. The coordinate system can be derived from different sources, including binary images, brightfield images or even super-resolution membrane markers. With the coordinate system in place, ColiCoords offers several different visualization and analysis methods, including alignment of cells, radial, longitudinal or angular (poles) distributions, kymographs and super-resolution reconstruction. (Fig 1).

2 ColiCoords principles and features

2.1 Coordinate definitions

The general principle of the coordinate system is shown in Fig 2. Here, a brightfield image of an *E. coli* cell is shown with the coordinate system overlaid. By establishing a per-cell coordinate system every pair of Cartesian coordinates (x_p, y_p) from either pixels or fluorescent foci can be mapped to cell coordinates (l_c, r_c, ϕ); Fig 2A. The transformation is based on a second-degree polynomial $(p(x), x_l \leq x \leq x_r)$ (red line) which corresponds to the midline of the cell. The parameters describing $p(x)$ are initially based on guess values which can be refined afterwards. To transform a point (x_p, y_p), the first step is to find the point (x_c, y_c) on $p(x)$ which is closest to (x_p, y_p), perpendicular to $p(x)$. This is done by minimizing the squared distance
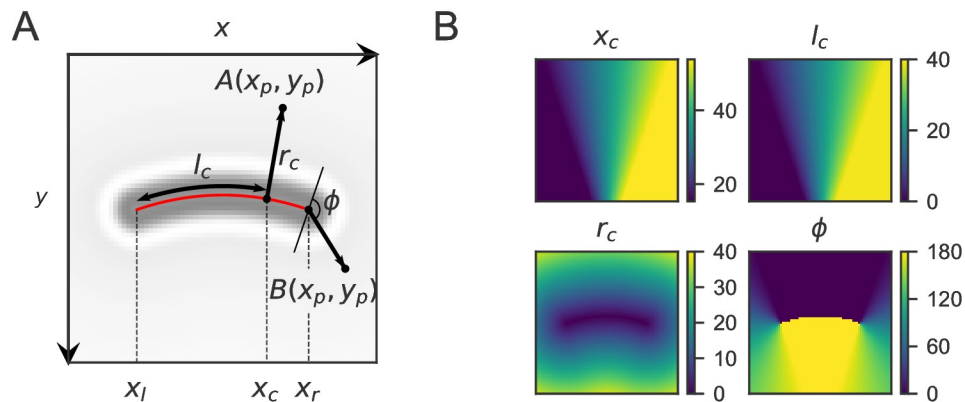
Fig 2. General description of the in-cell coordinate system. (A) Brightfield image of an *E. coli* cell with coordinate system overlayed. Every point with coordinates x_p, y_p can be transformed to coordinates l_c, r_c, ϕ. (B) Images showing the values of x_c as well as cellular coordinates values l_c, r_c, ϕ.

between both points:

$$r^2 = \left(x_c - x_p\right)^2 + \left(p(x_c) - y_p\right)^2 \tag{1}$$

$$\frac{\partial r^2}{\partial x_c} = 0 \tag{2}$$

Solving Eq 2 gives a cubic equation which is solved analytically to ensure fast transformation of many data points. To account for points B at the poles of the cells the coordinate x_c is restricted to the domain $[x_l, x_r]$. The cellular coordinates (l_c, r_c, ϕ) can then be calculated from x_c and $p(x)$. The longitudinal coordinate l_c is given by the arc length along $p(x)$ from x_l to x_c:

$$l_c = \int_{x_l}^{x_c} \sqrt{1 + \left(\frac{\mathrm{d}p(x)}{\mathrm{d}x}\right)^2} \, dx \tag{3}$$

This expression also gives the full length of the cell when x_c is substituted by x_r. The radial coordinate r_c is simply the euclidean distance between $(x_c, p(x_c))$ and (x_p, y_p):

$$r_c = \sqrt{\left(x_c - x_p\right)^2 + \left(p(x_c) - y_p\right)^2} \tag{4}$$

The third coordinate is the angle ϕ (in degrees) and it uniquely defines points A along the body of the cell by distinguishing between the top and bottom of the cell, as well as defining the position of points B at the poles. The value of ϕ is 0 at the top of the cell and runs from 0 to 180 along the right pole, where ϕ is defined as the angle between the line perpendicular to $p(x)$ at x_r and the line from $(x_c, p(x_c))$ to (x_p, y_p). For points below the midline of the cell, the value of ϕ is 180, which then runs from 180 back to 0 along the left pole.

Note that the top area of the cell, as it is displayed in Panel (A) of Fig 2, is given by $y_p < p(x_p)$, since the origin of the Cartesian coordinates is at the top left, where the postive y-axis is running down. By this definition, the coordinate of the midpoint of the top-left pixel is (0.5, 0.5), in line with the coordinate definitions in the image analysis software ImageJ [71, 72].

2.2 Preprocessing and optimization

To analyse fluorescence microscopy data with ColiCoords some preprocessing of the input data is required. Notably, ColiCoords is not designed to segment or detect cells in the raw microscopy images. Binary (labelled) input images are required for ColiCoords processing to identify the location of cells in the images. However, these binaries are only needed to localize cells and initialize the coordinate system, and not to form the final optimized coordinate system, and so consequently a high accuracy of segmentation is not required. Binary or segmented images can be derived from brightfield, differential interference contrast (DIC), phase contrast or fluorescence images by established tools such as CellProfiler [62], Ilastik [63], Oufti [55] or Convolutional Neural Networks (e.g. implemented by TensorFlow [64] or Keras [65]). An implementation of the U-Net segmentation architecture [73] is included in ColiCoords together with examples on how to use it.

Fluorescence images should be (optionally) background or illumination corrected and different channels such as brightfield or DIC should be aligned if they are acquired with different optics. ColiCoords allows the processing of any kind of image data as well as sparse data, such as single-molecule tracking and single-molecule localization microscopy (SMLM) data. Based on the provided binary images, single cells are automatically cut out from all image data, the positions of the localizations are selected and transferred to the coordinates of the cropped image. The orientation of the cell is calculated using the binary image by calculating the image moments [74]. All data elements are then rotated to orient the cells horizontally. The result of this process is a collection of single-cell Python-objects where all data are organized on a cell-by-cell basis, and every cell has its own coordinate system which can be used to perform calculations, analysis or visualizations on the data element of choice. These sets of cells can then be indexed and selected based on user-defined criteria (shape, size, fluorescent signal etc.) and further analysed or inspected through interactive plots within the Jupyter Notebook environment.

After initialization of the cell objects, the coordinate system needs to be refined to allow it to more accurately describe the cell's shape. ColiCoords implements the optimization of the coordinate system based on different data sources by providing several objective functions which, when minimized, give the best matching coordinate system for the respective data source. Any image-type data (e.g. brightfield) can be used provided there is an anisotropic signal along the length of the cell, as well as localization (SMLM) data provided the localizations describe the outline of the cell.

Fig 3 shows the iterative optimization process. In Panel (A) the ground truth binary image (derived from measurement) is shown together with the cell's coordinate system. The coordinate system is described by a total of 6 parameters, 3 coefficients a_0, a_1, a_2 which define the second degree polynomial $p(x)$, the left and right endpoints x_l and x_r, and the cell's radius r. In the figure, the polynomial $p(x)$ is shown together with the isodistance line from $p(x)$ with distance r. As can be seen from the figure, the initial coordinate system does not accurately describe the cell's shape. To refine the coordinate system, first a radial distance image is calculated (Panel (B)). This image is then thresholded with the initial guess value of r, which gives a new binary image (Panel (C), black). Comparison of this image with the ground truth binary image gives the χ^2 value, which is minimized to optimize the fit of the coordinate system. The fitting in ColiCoords is done via the package Symfit [75], which provides an API to the minimizers implemented by `scipy.minimize` [57].

Other microscopy images can be used to optimize the coordinate system, where the only constraint is that the signal needs to be isotropic along the length of the cell. This is shown for binary, brightfield and fluorescence images in Fig 4. The figure shows the three different
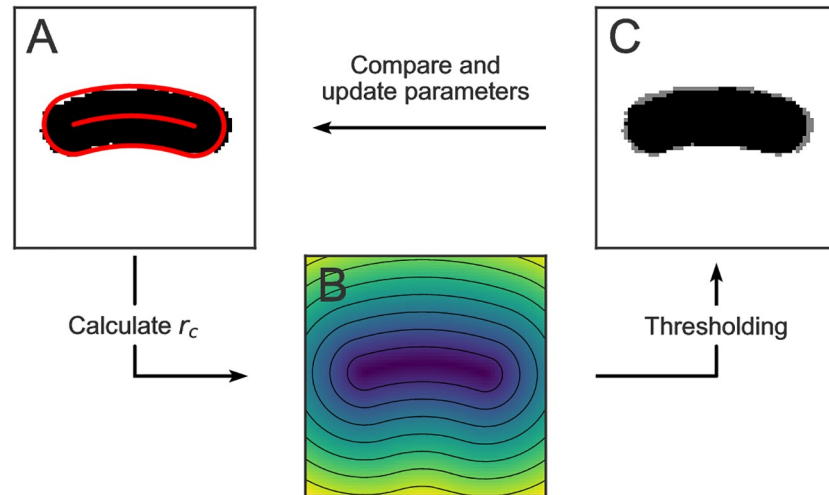
Fig 3. Optimization of the cell's internal coordinate system. (A) Ground-truth binary image with cell midline and outline calculated from initial guess parameters. (B) Radial distance image calculated from initial guess coordinates. (C) Calculated binary image (black) obtained by thresholding the distance image superimposed on the ground truth image (grey).

https://doi.org/10.1371/journal.pone.0217524.g003

Fig 4. Iterative optimization of the coordinate system. Optimization based on binary, brightfield and fluorescence images are shown.

https://doi.org/10.1371/journal.pone.0217524.g004

Fig 5. Iterative optimization of the coordinate system based on STORM super-resolution data of the membrane marker LacY-eYFP. (A) Top: STORM reconstruction on top of the *E. coli* binary image (grey) with initial guesses coordinate system in red. Bottom: Coordinate system after optimization based on STORM localizations. (B) STORM reconstruc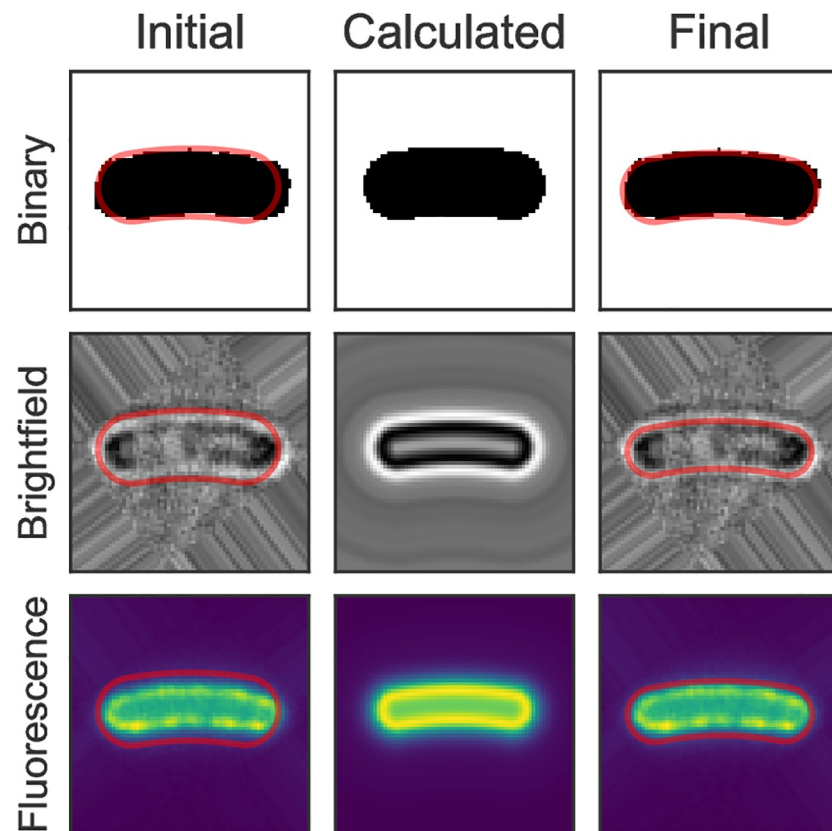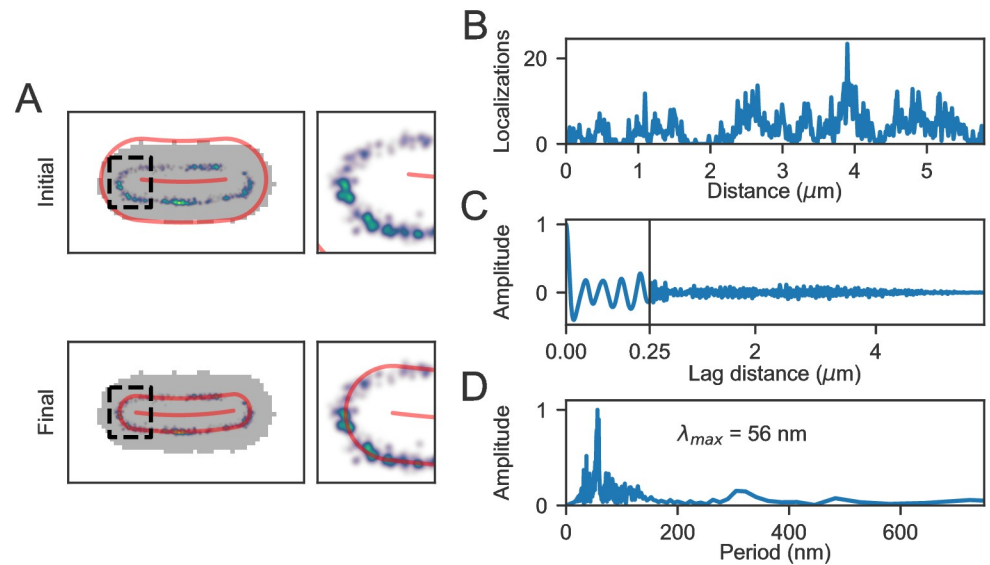tion along the perimeter of the cell showing localizations as a function of distance along the membrane. (C) Spatial autocorrelation function of (B). (D) Fourier transform of (C) showing the largest amplitude at a periodicity of 56 nm.

channels measured for a single cell. The binary is derived from the brightfield and is used for initial guesses of the coordinate system. This initial guess coordinate system is shown in red in the left column. This coordinate system is used to calculate a synthetic image, averaged along the angular coordinate (middle). A comparison between this image and the measurement gives a measure for how well the current parameters describe the shape of the cell. The optimal parameters can be found via iterative optimization (right). For brightfield- or fluorescence data-based optimization the radius r of the cell is determined in a second step by determining the half-maximum point of pixel intensity.

Finally, ColiCoords can optimize the coordinate system based on SMLM data of a membrane marker. This type of data can be obtained through various super-resolution techniques such as Point Accumulation for Imaging in Nanoscale Topography (PAINT) [27, 76–78], Stochastic Optical Reconstruction Microscopy (STORM) [79, 80] or PhotoActivated Localization Microscopy (PALM) [81, 82] imaging.

The optimization process and a possible application thereof is illustrated in Fig 5. Panel (A) shows a STORM super-resolution reconstruction of the membrane protein marker LacY-eYFP. In the top panel, the coordinate system based on initial guesses (derived from binary image; grey) is shown. The bottom panel shows the result upon optimization based on the STORM-localizations data. Here, the radial distance r_c for all localizations is calculated and by comparison to the value for r the χ^2 (squared differences) is calculated. Minimization of the χ^2 gives the parameters for the coordinate system that best describes the bacterial membrane. Interestingly, the localizations along the membrane show a periodic fine structure. To further characterize this, the position along the perimeter of the cell was plotted in panel (B). The zero-point position is defined as the beginning of the 'top' part of the membrane ($l_c = 0$, $\phi = 0$, $r_c = r$, panel (A)) and the positive directions runs clockwise along the cell outline.

To extract the size of the periodic structures observed, at first a spatial auto-correlation [83] function was calculated and its low-frequency components were subtracted by means of a

sliding window, to reveal clear oscillations (panel (C)). By Fourier transforming this signal (panel (D)) we found a periodicity of the signal of 56 nm, which we attribute to an artefact resulting from the YFP fusion, as previously reported for MreB-YFP constructs [84].

In conclusion, optimization based on a SMLM super-resolution membrane marker is expected to provide the most accurate approach for creating a coordinate system within the cell (vide infra, section (D)). This allows for aligning and combination of multiplexed localization data from many cells with high accuracy.

2.3 Batch processing

To demonstrate the ability of ColiCoords to analyse a dataset consisting of many cells, we applied a typical analysis procedure to two sets of fluorescence microscopy images. The first set of *E. coli* cells were incubated with the reactive dye Cy3B-NHS to label the outer membrane, while the second set of cells expressed eGFP in the cytosol.

For each condition, a subset of 100 images were manually annotated through the use of a custom GUI element. These annotated binary images were then combined and used to train a convolutional neural network. An implementation of the U-Net architecture [73] neural network was used, based on the Keras [65] API using Tensorflow [64] as backend.

The brightfield images were first scaled down to 256x256 to preserve graphics card memory. A custom implementation of Keras' `Sequence` was used for preprocessing and augmentation. All input images were normalized using tanh-estimators [85] and augmented 8-fold through permutations of horizontal mirroring, vertical mirroring and transposing the images. This augmented data was randomly split into validation data and training data.

After training the neural network, the network was applied to the whole set of images to generate binary images. The images were then scaled back to 512x512 pixels. Next, binary shapes were filtered from the segmentation mask by their shape and size. The resulting binary images were used as input for ColiCoords, together with the corresponding brightfield and fluorescence images. The cell objects were filtered based on the results of optimizations by binary images and brightfield images as well as a measurement of the cell's radius from the brightfield image. This yielded a total of 2341 and 1691 cells for the Cy3B and eGFP datasets, respectively.

The analysis output is shown in Fig 6. In panel A, an individual cell for each condition are shown. Their radial distribution profile is shown in panel (B) (blue line), together with the individual datapoints from each pixel as red points. In panel (C), all cells from each dataset were aligned by transforming all pixel coordinates in the fluorescence images from each cell to cellular coordinates. These cellular coordinates are then transformed back onto a model cell with user-specified dimensions. The resulting point cloud is then used to generate an aligned fluorescence image by convolution with a 2D gaussian. The final image is displayed with spline interpolation. In panel (D), the average radial distribution for all cells is shown. The curves are individually normalized in the x-direction by setting the brightfield-measured radius to 1 whereas the y-direction is normalized by the maximum of each curve. The shaded region (standard deviation) therefore only reports on variations in the shape of the radial distribution profile.

2.4 Synthetic benchmarks

In order to evaluate ColiCoords' performance a synthetic dataset was generated and subsequently analyzed. By comparing the results for different conditions and optimization methods the accuracy can be benchmarked.
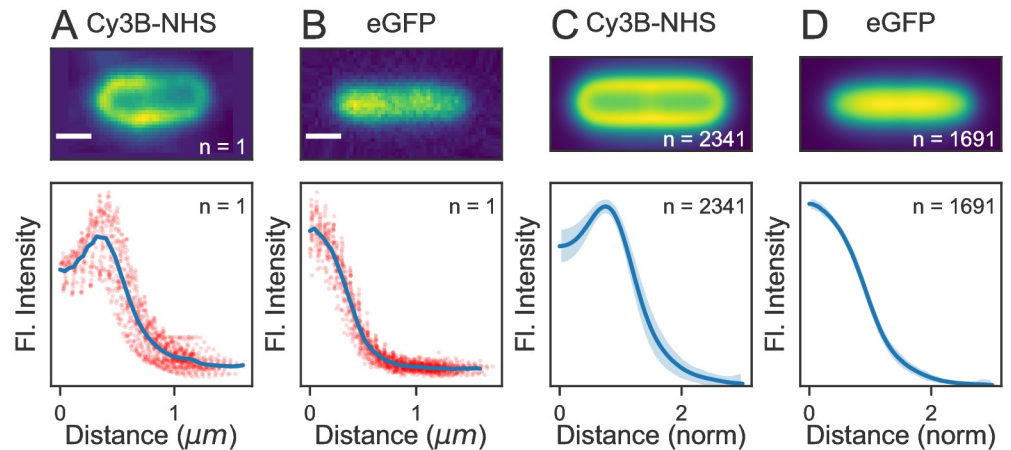
Fig 6. Batch processing of several thousand cell objects. Two datasets were analyzed, one where cells are labelled on the outer membrane with Cy3B-NHS and one with cell expressing eGFP in the cytosol. (A) Fluorescence images of single *E. coli* cells, scale bar 750 nm. (B) Radial distribution profiles for the cells in (A). Individual datapoints from every pixel in the image are shown as red points, together with the resulting radial distribution (blue line). (C) Aligned and averaged fluorescence images for n = 2341 and n = 1691 cells, respectively. (D) Average radial distribution profiles. The standard deviation is shown as a shaded region.

First, a set of 10000 synthetic cells were generated using ColiCoords' `synthetic_data` module. The geometric parameters describing these synthetic cells were chosen based on a set of measured *E. coli* cells. These synthetic cells have a binary image data element by default. A brightfield data element was added based on normalized brightfield radial distributions measured from *E. coli*. The ratio between the background intensity (no cells) of the brightfield and the maximum of the scattered light from the bacterial membrane was measured and kept constant to replicate realistic signal-to-noise ratios. The background light was set to one and then multiplied by the chosen photon count number (500, 1000 and 10000), after which the resulting values where drawn from a Poisson distribution to simulate shot noise. Finally, normally distributed noise was added with a standard deviation of 20 photon counts.

To each cell two homogeneously distributed membrane-localized STORM data elements were added. The first data element describes localizations at the inner membrane. The mean radial distance of the localizations was set according to where the inner membrane should be with respect to the radius of the cell measured from the brightfield image based on measurements with the construct LacY-eYFP. The standard deviation from the mean radius was set to 0.25 pixels (20 nm). A second STORM data element was added 100 nm further out with the same standard deviation.

From this set of synthetic cells, images were reconstructed. An average of 10 ± 3 cells were taken from the set and randomly rotated and placed in each image of 512x512 pixels (40 μm x 40 μm), ensuring at least 5 pixels distance between the cells' binary images. The coordinates of the cell's STORM data element were combined into one big STORM localizations table.

These processes yielded 1000 ground-truth binary images together with 3 sets of 1000 brightfield images with different photon numbers (500, 1000, 10000 per pixel) and therefore different signal-to-noise ratios (Example images in Fig 7A). The whole ColiCoords' data analysis pipeline was performed on this dataset to benchmark its performance.

First, the brightfield images were segmented. For all 3 signal-to-noise conditions, the first 400 images were used to train the neural network. After training, the network was applied to the whole set of 1000 images to generate binary images.
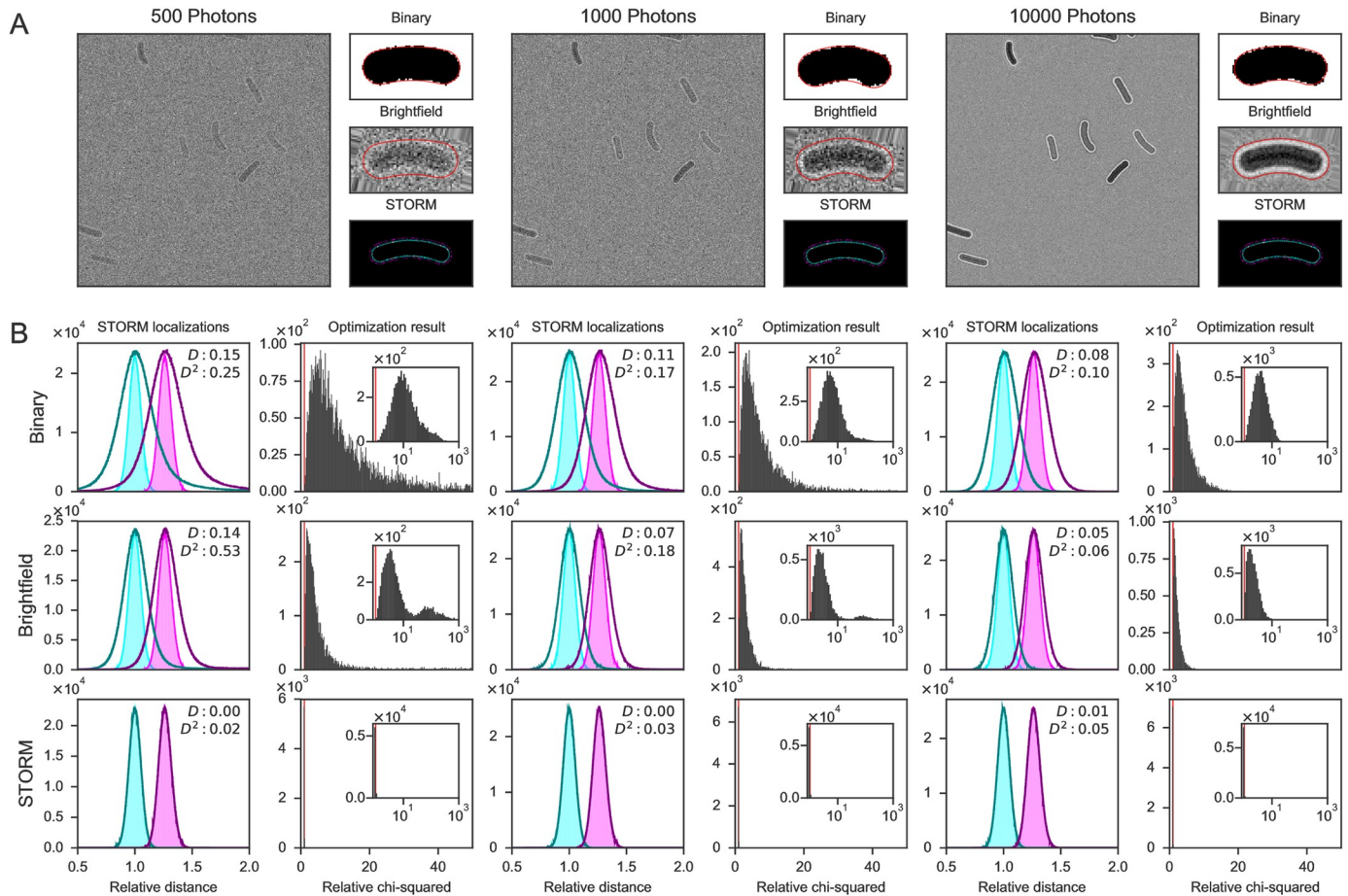
Fig 7. Benchmarking of ColiCoords software with a synthetic dataset. (A) Examples of generated input data brightfield images for different photon counts (500, 1000 and 10000 average per pixel) and example of one processed single cell. Different data elements (Binary, Brightfield, STORM) are shown together with the outline of the coordinate system (red/white line) where the coordinate system is optimized based on that data element. (B) Evaluation of the coordinate system of n = 6569, 7127, 7245 cells out of 7245. Left: relative radial distance of all inner (cyan) and outer (magenta) membrane STORM localizations. The radial position is normalized to the ground-truth inner membrane position. The ground-truth radial distance (light colours, filled histogram) is compared with radial distances calculated with the coordinate system as calculated for the different data elements (binary, brightfield, STORM) for different brightfield image photon counts (dark colours, line only). The absolute mean deviation (D) and the root mean squared deviation (D^2) are shown in the graph. Right: Relative minimization objective function χ^2. The χ^2 is calculated for the obtained coordinate system for each cell for the STORM data element for each condition. The obtained value is divided by the ground-truth χ^2 value, therefore a value of 1 indicates a perfectly fitted coordinate system (red line).

Next, the binary images were filtered. Bordering cells were removed and the remaining binary objects were selected based on their size and ellipticity where the selection criteria were obtained from the ground-truth binary images. The filtered and labelled binary images were then used by ColiCoords together with the brightfield and STORM data to generate a set of 'measured' cell objects.

The 'measured' cell objects obtained were correlated to the original ground-truth cell objects by first performing a cluster analysis on the input STORM table [86]. The ground-truth cells were then identified by encoding the combined STORM intensity and comparing to the ground-truth intensity of the STORM clusters. The corresponding measured cell was then found by comparing the centre-of-mass positions of the STORM clusters and the filtered binary images used as input. This yielded two sets of cells, one with ground-truth cells and one with corresponding 'measured' cells for comparison.

Cells which had too few STORM localizations compared to their ground-truth counterparts were discarded due to them being too poorly segmented. Several cells had too many STORM localizations due to localizations from neighbouring cells. These additional localizations were also removed. This yielded a total of 6569, 7129 and 7245 cells out of maximum 7245 cells for brightfield images with 500, 1000 and 10000 photons, respectively. This number is lower for input images with lower signal-to-noise because some cells are segmented poorly by the neural network.

Finally, the coordinate system of the cells was optimized using different approaches to compare their accuracy. The cells in each condition were initialized with guesses derived from segmented binary images and subsequently optimized based on the data elements binary, brightfield and storm (inner membrane). This yielded a total of 9 sets of cells which are summarized in Fig 7B. In the left column of each condition, the relative radial distances of all STORM localizations in the dataset are plotted as calculated by its coordinate system. The radial distances were normalized to the ground truth inner membrane distance to eliminate cell-to-cell radius variations. The ground truth radial distances of the inner (cyan) and outer (magenta) are shown in light colours and filled histogram. The calculated radial distances are shown on top with dark colours where the outer line of the histogram is shown only.

From these graphs the effect of the accuracy of the coordinate system can be shown on the resulting radial distribution graph. As can be seen, in the condition with the lowest number of photons (500) and a coordinate system optimized on the binary data element, the resulting radial distribution histogram is significantly wider than the ground-truth radial distribution. Compared to binary images derived from higher signal-to-noise images, the result more accurately describes the underlying ground-truth. This is further reflected by comparing the obtained χ^2 values from the optimization process. In the histogram on the right (second column) the relative χ^2 of each cell derived from its optimized coordinate system is plotted. The values are obtained by calculating the χ^2 value for STORM optimization (mean squared difference in distance between coordinate system outline and STORM localizations) and dividing this value by the ground-truth χ^2 value to get a relative χ^2. This value is an indication for the goodness of the fit, where a value of one indicates a perfect fit (red line in histograms).

As can be seen in the histogram, for 500 photons and optimization based on the binary image (Panel (B), top left), the median χ^2 value of optimization is roughly 10-fold higher compared to the ground-truth values. The inset (log scale x-axis and logarithmically increasing bin size) shows a long tail in the distribution to higher relative χ^2 values up to 500.

When comparing the binary optimization result for 1000 and 10000 photons, it can be seen that the radial distribution histograms become narrower and more accurately match the ground truth radial distribution. The distribution of relative χ^2 values also reflect this with median values of 5.8 an 3.8, respectively.

Moving one row down, the results are plotted for optimization based on the brightfield image. The initial guesses for optimization derived from the binary images and therefore the quality of the segmentation influences the final result. As can be seen from the radial distribution histograms, the coordinate system obtained more accurately describes the cell compared to binary optimization. For 10000 photons the radial distribution graph very closely matches the ground truth. In the relative χ^2 distributions a second population can be seen which is mostly present for lower photon numbers. This is a consequence of poor initial guesses and the optimization therefore returns a different local minimum. The median relative χ^2 values are 4.1, 2.6 and 1.9, respectively.

The third row shows the result from optimization based on the inner membrane STORM localization dataset. Since the STORM localizations are identical in all cells and are independent of the brightfield image, any difference in the optimization result is due to the initial

guesses. As expected, the obtained radial distribution histograms almost perfectly overlap with the ground truth histograms. The obtained median relative χ^2 values are all 0.99, however some small outliers are found, where the population sizes of values >50 are 2, 3 and 9 respectively.

In Fig 7 for 10000 photons and brightfield optimization it can be seen that the ground-truth result is matched closely but not completely. This is unexpected since the cells were generated from a perfect second-degree polynomial shape and therefore it was expected that the optimization result would more accurately describe the cell's shape. The χ^2 value of the obtained result was lower compared to the ground-truth, thus the problem is not that the algorithm returned a local minimum instead of a global minimum. Therefore, switching to a global optimization algorithm such as differential evolution [87] will not yield the desired improved result.

Instead, it was found that in the optimizer's implementation a choice of speed over accuracy is the origin of the discrepancy. Specifically, in the bootstrapping process to calculate a simulated brightfield image to compare with the measured image, the radial distribution of the brightfield image is calculated by normal binning (histogram) compared to convolution with a Gaussian kernel. This choice was made to keep the calculation times manageable, and optimizing of this procedure, possibly through a two step (course and fine) optimization process is expected to yield improved results.

The STORM optimization results (lower row, Fig 7) do closely match the ground-truth result. Any deviation originates from a small number of cells which do not converge to the correct solution. Since this population is larger for lower photon brightfield images, which has a less accurate segmentation result, this is be attributed to poor initial guesses for the coordinate system.

Finally, using an optimizer that is restricted by bounds on the optimization parameters is likely to reduce the subpopulations in both brightfield and STORM optimization. These features and the issues mentioned above will be addressed in a future performance update of ColiCoords.

3 Conclusions

ColiCoords is an open-source software package to analyse fluorescence microscopy data of rod-shaped cells. It allows for the transformation of Cartesian coordinates from any data source to cellular coordinates. This transformation can then be used to obtain distributions of fluorescence along the cell long or short axes or its perimeter and align the whole cell to allow for the combination of data from many cells.

The open-source nature and Jupyter-notebook based workflow together with an open and compact file format promotes open and reproducible analysis of microscopy data. As illustrated in this paper, ColiCoords facilitates the publication in the form of a 'reproducible article' [88], as recently featured in eLife, where data and code to generate the figures are bundled together with the publication. The code and data to generate the figures in this article is available online (see Methods).

ColiCoords' code is well documented and the main features are illustrated by example notebooks, which can be tested live on Binder. The software itself can be installed directly from both the Anaconda and PyPi package managers.

Analysis of synthetic data has shown that we can determine the cellular coordinates of SMLM super-resolution datapoints based on only the brightfield image of the cell with high accuracy (5%). If the coordinate system is derived from super-resolution measurements of a membrane marker, the accuracy is further increased (1%).

Although the current state of ColiCoords is a mature project, future updates are planned to be released, either by us or in collaboration with the community. Three major updates are planned; first a performance update to increase speed and accuracy of the optimization process, second a metadata update allowing the storage of image-associated metadata, and finally an update to allow the organization of cell objects into a hierarchical structure to process lineage and temporal information.

4 Methods

4.1 Cy3B-NHS staining

E. coli MG1655 cells were grown overnight in LB medium and subsequently diluted 1000 times in EZ rich medium (Teknova) with 0.2% glucose. The cells were grown until the culture reached an OD600 value of 0.3 after which the cells were centrifuged at 3000 rcf for 5 minutes and resuspended in PBS buffer. The amino-reactive dye Cy3B-NHS (GE Healthcare) was added to a final concentration of 1 mM and incubated while protected from light at room temperature for 1 hour [89]. Next, the cells were centrifuged at 3000 rcf for 10 minutes and the cell pellet was resuspended in PBS.

4.2 eGFP staining

E. coli BL21(DE3) cells were grown overnight in LB medium and subsequently diluted 1000 times in EZ rich medium (Teknova) with 0.2% glucose. The growth media was supplemented with the appropriate antibiotics. When the cell culture reached an OD600 of 0.3, eGFP expression was induced with IPTG at a final concentration of 200 μM for 30 minutes at 37 ˚C.

For imaging, 3 μL of the bacterial culture was transferred to a microscope coverslip and covered by an agarose pad [90, 91].

Imaging was done on an Olympus IX83 inverted microscope with an Olympus UAPON 100x NA 1.49 TIRF oil immersion objective. The excitation light (514 nm, Coherent) was coupled into the objective via the ET—442/514/561 Laser Triple band set (69904, Chroma) and the fluorescence was collected on a electron multiplying charge-coupled CCD camera (512x512 pixel, C9100-13, Hamamatsu). The focal position was held constant by the Olympus ZDC2 and images on multiple positions were collected automatically using Olympus' CellSens software.

4.3 Super-resolution imaging

LacY-eYFP expressing *E. coli* C41 cells were grown overnight in LB medium with appropriate antibiotics and subsequently diluted 1000 times in EZ rich medium (Teknova) with 0.4% glycerol. When the cell culture reached an OD600 of 0.3, Lacy-eYFP expression was induced with 0.01%(v) L-arabinose for 30 minutes at 37 ˚C. The cells were centrifuged at 3000 rcf for 5 minutes and resuspended in EZ rich medium with 0.2% glucose. 3 μL of the bacterial culture was transferred onto a microscope coverslide and covered by an agarose gel pad.

Images were captured with an exposure time of 50 ms using the open source software μManager [92, 93].

To obtain the final STORM reconstruction, 2000 frames were collected and processed with the ImageJ plugin ThunderSTORM [68].

All data and code used to generate the figures in this article can be found at Zenodo (DOI: 10.5281/zenodo.2637790) and GitHub, respectively.

Acknowledgments

Author Contributions

Conceptualization: Jochem H. Smit, Andreas Herrmann, Thorben Cordes.

Data curation: Jochem H. Smit, Yichen Li, Eliza M. Warszawik.

Formal analysis: Jochem H. Smit.

Funding acquisition: Andreas Herrmann, Thorben Cordes.

Investigation: Yichen Li, Eliza M. Warszawik.

Methodology: Jochem H. Smit.

Project administration: Andreas Herrmann, Thorben Cordes.

Resources: Eliza M. Warszawik, Andreas Herrmann, Thorben Cordes.

Software: Jochem H. Smit.

Supervision: Andreas Herrmann, Thorben Cordes.

Validation: Jochem H. Smit, Yichen Li, Eliza M. Warszawik.

Visualization: Jochem H. Smit.

Writing – original draft: Jochem H. Smit.

Writing – review & editing: Jochem H. Smit, Andreas Herrmann, Thorben Cordes.

References

1. Xie XS, Choi PJ, Li GW, Lee NK, Lia G. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. Annual Review of Biophysics. 2008; 37(1):417–444. https://doi.org/10.1146/annurev.biophys.37.092607.174640 PMID: 18573089

2. Gahlmann A, Moerner WE. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. Nature Reviews Microbiology. 2014; 12(1):9–22. https://doi.org/10.1038/nrmicro3154 PMID: 24336182

3. Yao Z, Carballido-López R. Fluorescence Imaging for Bacterial Cell Biology: From Localization to Dynamics, From Ensembles to Single Molecules. Annual Review of Microbiology. 2014; 68(1):459–476. https://doi.org/10.1146/annurev-micro-091213-113034 PMID: 25002084

4. Schneider JP, Basler M. Shedding light on biology of bacterial cells. Philosophical Transactions of the Royal Society B: Biological Sciences. 2016; 371(1707):20150499. https://doi.org/10.1098/rstb.2015.0499

5. Gijtenbeek V, A L, Kok J. Illuminating Messengers: An Update and Outlook on RNA Visualization in Bacteria. Frontiers in Microbiology. 2017; 8. PMID: 28690601

6. Elf J, Barkefors I. Single-Molecule Kinetics in Living Cells. Annual Review of Biochemistry. 2019; 88(1): null. https://doi.org/10.1146/annurev-biochem-013118-110801

7. Kapanidis AN, Lepore A, El Karoui M. Rediscovering Bacteria through Single-Molecule Imaging in Living Cells. Biophysical Journal. 2018; 115(2):190–202. https://doi.org/10.1016/j.bpj.2018.03.028 PMID: 29680157

8. Kapanidis AN, Uphoff S, Stracy M. Understanding Protein Mobility in Bacteria by Tracking Single Molecules. Journal of Molecular Biology. 2018; 430(22):4443–4455. https://doi.org/10.1016/j.jmb.2018.05.002 PMID: 29753778

9. Leake MC, Chandler JH, Wadhams GH, Bai F, Berry RM, Armitage JP. Stoichiometry and turnover in single, functioning membrane protein complexes. Nature. 2006; 443(7109):355–358. https://doi.org/10.1038/nature05135 PMID: 16971952

10. Badrinarayanan A, Reyes-Lamothe R, Uphoff S, Leake MC, Sherratt DJ. In Vivo Architecture and Action of Bacterial Structural Maintenance of Chromosome Proteins. Science. 2012; 338(6106):528–531. https://doi.org/10.1126/science.1227126 PMID: 23112333

11. Holden SJ, Pengo T, Meibom KL, Fernandez CF, Collier J, Manley S. High throughput 3D super-resolution microscopy reveals Caulobacter crescentus in vivo Z-ring organization. Proceedings of the National Academy of Sciences. 2014; 111(12):4566–4571. https://doi.org/10.1073/pnas.1313368111

12. Palayret M, Armes H, Basu S, Watson AT, Herbert A, Lando D, et al. Virtual-'Light-Sheet' Single-Molecule Localisation Microscopy Enables Quantitative Optical Sectioning for Super-Resolution Imaging. PLOS ONE. 2015; 10(4):e0125438. https://doi.org/10.1371/journal.pone.0125438 PMID: 25884495

13. Liao Y, Schroeder JW, Gao B, Simmons LA, Biteen JS. Single-molecule motions and interactions in live cells reveal target search dynamics in mismatch repair. Proceedings of the National Academy of Sciences. 2015; 112(50):E6898–E6906. https://doi.org/10.1073/pnas.1507386112

14. Yang F, Chen TY, Krzemiński u, Santiago AG, Jung W, Chen P. Single-Molecule Dynamics of the Molecular Chaperone Trigger Factor in Living Cells. Molecular microbiology. 2016; 102(6):992–1003. https://doi.org/10.1111/mmi.13529 PMID: 27626893

15. Barlag B, Beutel O, Janning D, Czarniak F, Richter CP, Kommnick C, et al. Single molecule super-resolution imaging of proteins in living Salmonella enterica using self-labelling enzymes. Scientific Reports. 2016; 6:31601. https://doi.org/10.1038/srep31601 PMID: 27534893

16. Beattie TR, Kapadia N, Nicolas E, Uphoff S, Wollman AJ, Leake MC, et al. Frequent exchange of the DNA polymerase during bacterial chromosome replication. eLife. 2017; 6:e21763. https://doi.org/10.7554/eLife.21763 PMID: 28362256

17. Diepold A, Sezgin E, Huseyin M, Mortimer T, Eggeling C, Armitage JP. A dynamic and adaptive network of cytosolic interactions governs protein export by the T3SS injectisome. Nature Communications. 2017; 8:15940. https://doi.org/10.1038/ncomms15940 PMID: 28653671

18. Lewis JS, Spenkelink LM, Jergic S, Wood EA, Monachino E, Horan NP, et al. Single-molecule visualization of fast polymerase turnover in the bacterial replisome. eLife. 2017; 6:e23932. https://doi.org/10.7554/eLife.23932 PMID: 28432790

19. Stracy M, Kapanidis AN. Single-molecule and super-resolution imaging of transcription in living bacteria. Methods. 2017; 120:103–114. https://doi.org/10.1016/j.ymeth.2017.04.001 PMID: 28414097

20. Garza de Leon F, Sellars L, Stracy M, Busby SJW, Kapanidis AN. Tracking Low-Copy Transcription Factors in Living Bacteria: The Case of the lac Repressor. Biophysical Journal. 2017; 112(7):1316–1327. https://doi.org/10.1016/j.bpj.2017.02.028 PMID: 28402875

21. Bisson-Filho AW, Hsu YP, Squyres GR, Kuru E, Wu F, Jukes C, et al. Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. Science. 2017; 355(6326):739–743. https://doi.org/10.1126/science.aak9973 PMID: 28209898

22. Lamberte LE, Baniulyte G, Singh SS, Stringer AM, Bonocora RP, Stracy M, et al. Horizontally acquired AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. Nature Microbiology. 2017; 2(3):16249. https://doi.org/10.1038/nmicrobiol.2016.249 PMID: 28067866

23. Lund VA, Wacnik K, Turner RD, Cotterell BE, Walther CG, Fenn SJ, et al. Molecular coordination of Staphylococcus aureus cell division. eLife. 2018; 7. https://doi.org/10.7554/eLife.32057

24. Ghodke H, Paudel BP, Lewis JS, Jergic S, Gopal K, Romero ZJ, et al. Spatial and temporal organization of RecA in the Escherichia coli DNA-damage response. eLife. 2019; 8. https://doi.org/10.7554/eLife.42761 PMID: 30717823

25. Crawford R, Torella J, Aigrain L, Plochowietz A, Gryte K, Uphoff S, et al. Long-Lived Intracellular Single-Molecule Fluorescence Using Electroporated Molecules. Biophysical Journal. 2013; 105(11):2439–2450. https://doi.org/10.1016/j.bpj.2013.09.057 PMID: 24314075

26. Plochowietz A, Farrell I, Smilansky Z, Cooperman BS, Kapanidis AN. In vivo single-RNA tracking shows that most tRNA diffuses freely in live bacteria. Nucleic Acids Research. 2017; 45(2):926–937. https://doi.org/10.1093/nar/gkw787 PMID: 27625389

27. Spahn CK, Glaesmann M, Grimm JB, Ayala AX, Lavis LD, Heilemann M. A toolbox for multiplexed super-resolution imaging of the E. coli nucleoid and membrane using novel PAINT labels. Scientific Reports. 2018; 8(1):14768. https://doi.org/10.1038/s41598-018-33052-3 PMID: 30282984

28. Volkov IL, Lindén M, Aguirre Rivera J, Ieong KW, Metelev M, Elf J, et al. tRNA tracking for direct measurements of protein synthesis kinetics in live cells. Nature Chemical Biology. 2018; 14(6):618–626. https://doi.org/10.1038/s41589-018-0063-y PMID: 29769736

29. Lukyanov KA, Chudakov DM, Lukyanov S, Verkhusha VV. Photoactivatable fluorescent proteins. Nature Reviews Molecular Cell Biology. 2005; 6(11):885–890. https://doi.org/10.1038/nrm1741 PMID: 16167053

30. Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nature Biotechnology. 2004; 22(12):1567–1572. https://doi.org/10.1038/nbt1037 PMID: 15558047

31. Shcherbo D, Murphy CS, Ermakova GV, Solovieva EA, Chepurnykh TV, Shcheglov AS, et al. Far-red fluorescent tags for protein imaging in living tissues. Biochemical Journal. 2009; 418(3):567–574. https://doi.org/10.1042/BJ20081949 PMID: 19143658

32. Nguyen AW, Daugherty PS. Evolutionary optimization of fluorescent proteins for intracellular FRET. Nature Biotechnology. 2005; 23(3):355–360. https://doi.org/10.1038/nbt1066 PMID: 15696158

33. Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal Structure of the Aequorea victoria Green Fluorescent Protein. Science. 1996; 273(5280):1392–1395. https://doi.org/10.1126/science.273.5280.1392 PMID: 8703075

34. Patterson GH, Lippincott-Schwartz J. A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells. Science. 2002; 297(5588):1873–1877. https://doi.org/10.1126/science.1074952 PMID: 12228718

35. Turkowyd B, Balinovic A, Virant D, Carnero HGG, Caldana F, Endesfelder M, et al. A General Mechanism of Photoconversion of Green-to-Red Fluorescent Proteins Based on Blue and Infrared Light Reduces Phototoxicity in Live-Cell Single-Molecule Imaging. Angewandte Chemie International Edition. 2017; 56(38):11634–11639. https://doi.org/10.1002/anie.201702870 PMID: 28574633

36. Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Röcker C, Salih A, et al. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. Proceedings of the National Academy of Sciences. 2004; 101(45):15905–15910. https://doi.org/10.1073/pnas.0403668101

37. Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, Revyakin A, et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. Nature Methods. 2015; 12(3):244–250. https://doi.org/10.1038/nmeth.3256 PMID: 25599551

38. Grimm JB, Brown TA, English BP, Lionnet T, Lavis LD. Synthesis of Janelia Fluor HaloTag and SNAP-Tag Ligands and Their Use in Cellular Imaging Experiments. In: Erfle H, editor. Super-Resolution Microscopy: Methods and Protocols. Methods in Molecular Biology. New York, NY: Springer New York; 2017. p. 179–188. Available from: https://doi.org/10.1007/978-1-4939-7265-4_15.

39. Grimm JB, Brown TA, Tkachuk AN, Lavis LD. General Synthetic Method for Si-Fluoresceins and Si-Rhodamines. ACS Central Science. 2017; 3(9):975–985. https://doi.org/10.1021/acscentsci.7b00247 PMID: 28979939

40. Butkevich AN, Belov VN, Kolmakov K, Sokolov VV, Shojaei H, Sidenstein SC, et al. Hydroxylated Fluorescent Dyes for Live-Cell Labeling: Synthesis, Spectra and Super-Resolution STED. Chemistry—A European Journal. 2017; 23(50):12114–12119. https://doi.org/10.1002/chem.201701216

41. Kolmakov K, Hebisch E, Wolfram T, Nordwig LA, Wurm CA, Ta H, et al. Far-Red Emitting Fluorescent Dyes for Optical Nanoscopy: Fluorinated Silicon-Rhodamines (SiRF Dyes) and Phosphorylated Oxazines. Chemistry. 2015; 21(38):13344–13356. https://doi.org/10.1002/chem.201501394 PMID: 26272226

42. Keppler A, Gendreizig S, Gronemeyer T, Pick H, Vogel H, Johnsson K. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. Nature Biotechnology. 2003; 21(1):86–89. https://doi.org/10.1038/nbt765 PMID: 12469133

43. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, et al. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. ACS Chemical Biology. 2008; 3 (6):373–382. https://doi.org/10.1021/cb800025k PMID: 18533659

44. Plass T, Milles S, Koehler C, Szymański J, Mueller R, Wießler M, et al. Amino Acids for Diels–Alder Reactions in Living Cells. Angewandte Chemie International Edition. 2012; 51(17):4166–4170. https://doi.org/10.1002/anie.201108231 PMID: 22473599

45. Kapanidis AN, Ebright YW, Ebright RH. Site-Specific Incorporation of Fluorescent Probes into Protein: Hexahistidine-Tag-Mediated Fluorescent Labeling with (Ni $^{2+}$: Nitrilotriacetic Acid) $_n$ −Fluorochrome Conjugates. Journal of the American Chemical Society. 2001; 123(48):12123–12125. https://doi.org/10.1021/ja017074a PMID: 11724636

46. Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, Yao Y, et al. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. Journal of the American Chemical Society. 2002; 124(21):6063–6076. https://doi.org/10.1021/ja017687n PMID: 12022841

47. Conrad C, Gerlich DW. Automated microscopy for high-content RNAi screening. The Journal of Cell Biology. 2010; 188(4):453–461. https://doi.org/10.1083/jcb.200910105 PMID: 20176920

48. Zeder M, Kohler E, Pernthaler J. Automated quality assessment of autonomously acquired microscopic images of fluorescently stained bacteria. Cytometry Part A. 2010; 77A(1):76–85.

49. Rivenson Y, Ozcan A. Toward a Thinking Microscope: Deep Learning in Optical Microscopy and Image Reconstruction. Optics and Photonics News. 2018; 29(7):34.

50. Pepperkok R, Ellenberg J. High-throughput fluorescence microscopy for systems biology. Nature Reviews Molecular Cell Biology. 2006; 7(9):690–696. https://doi.org/10.1038/nrm1979 PMID: 16850035

51. Kam Z, Zamir E, Geiger B. Probing molecular processes in live cells by quantitative multidimensional microscopy. Trends in Cell Biology. 2001; 11(8):329–334. https://doi.org/10.1016/S0962-8924(01)02067-0 PMID: 11489638

52. Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Böckelmann R, Malykh Y, et al. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. Nature Biotechnology. 2006; 24(10):1270–1278. https://doi.org/10.1038/nbt1250 PMID: 17013374

53. Cass JA, Stylianidou S, Kuwada NJ, Traxler B, Wiggins PA. Probing bacterial cell biology using image cytometry. Molecular Microbiology. 2017; 103(5):818–828. https://doi.org/10.1111/mmi.13591 PMID: 27935200

54. Stylianidou S, Brennan C, Nissen SB, Kuwada NJ, Wiggins PA. SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells. Molecular Microbiology. 2016; 102(4):690–700. https://doi.org/10.1111/mmi.13486 PMID: 27569113

55. Paintdakhi A, Parry B, Campos M, Irnov I, Elf J, Surovtsev I, et al. Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis: Oufti: image analysis software. Molecular Microbiology. 2016; 99(4):767–777.

56. Ducret A, Quardokus EM, Brun YV. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nature Microbiology. 2016; 1(7):16077. https://doi.org/10.1038/nmicrobiol.2016.77 PMID: 27572972

57. Jones E, Oliphant T, Peterson P, et al. SciPy: Open source scientific tools for Python; 2001–. Available from: http://www.scipy.org/.

58. Shen H. Interactive notebooks: Sharing the code. Nature News. 2014; 515(7525):151. https://doi.org/10.1038/515151a

59. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows; 2016.

60. Perkel JM. Why Jupyter is data scientists' computational notebook of choice. Nature. 2018; 563:145. https://doi.org/10.1038/d41586-018-07196-1 PMID: 30375502

61. Ingargiola A, Lerner E, Chung S, Weiss S, Michalet X. FRETBursts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET. PLOS ONE. 2016; 11(8):e0160716. https://doi.org/10.1371/journal.pone.0160716 PMID: 27532626

62. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biology. 2006; 7(10):R100. https://doi.org/10.1186/gb-2006-7-10-r100 PMID: 17076895

63. Sommer C, Straehle C, Köthe U, Hamprecht FA. Ilastik: Interactive learning and segmentation toolkit. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2011. p. 230–233.

64. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems; 2015. Available from: https://www.tensorflow.org/.

65. Chollet F, et al. Keras; 2015. https://keras.io.

66. Holden SJ, Uphoff S, Kapanidis AN. DAOSTORM: an algorithm for high- density super-resolution microscopy. Nature Methods. 2011; 8(4):279–280. https://doi.org/10.1038/nmeth0411-279 PMID: 21451515

67. Babcock H, Sigal YM, Zhuang X. A high-density 3D localization algorithm for stochastic optical reconstruction microscopy. Optical Nanoscopy. 2012; 1(1):6. https://doi.org/10.1186/2192-2853-1-6

68. Ovesný M, Křížek P, Borkovec J, Švindrych Z, Hagen GM. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. Bioinformatics. 2014; 30(16):2389–2390. https://doi.org/10.1093/bioinformatics/btu202 PMID: 24771516

69. Henriques R, Lelek M, Fornasiero EF, Valtorta F, Zimmer C, Mhlanga MM. QuickPALM: 3D real-time photoactivation nanoscopy image processing in ImageJ. Nature Methods. 2010; 7(5):339–340. https://doi.org/10.1038/nmeth0510-339 PMID: 20431545

70. Kamentsky L, Jones TR, Fraser A, Bray MA, Logan DJ, Madden KL, et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinformatics. 2011; 27(8):1179–1180. https://doi.org/10.1093/bioinformatics/btr095 PMID: 21349861

71. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012; 9:671–675. https://doi.org/10.1038/nmeth.2089 PMID: 22930834

72. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics. 2017; 18(1):529. https://doi.org/10.1186/s12859-017-1934-z PMID: 29187165

73. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation; 2015.

74. Coelho LP. Mahotas: Open source software for scriptable computer vision. Journal of Open Research Software. 2013; 1(1):e3. https://doi.org/10.5334/jors.ac

75. Roelfs M, Kroon PC. tBuLi/symfit: symfit 0.4.6; 2018. Available from: https://zenodo.org/record/1470553.

76. Sharonov A, Hochstrasser RM. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. Proceedings of the National Academy of Sciences. 2006; 103(50):18911–18916. https://doi.org/10.1073/pnas.0609643104

77. Spahn C, Cella-Zannacchi F, Endesfelder U, Heilemann M. Correlative super-resolution imaging of RNA polymerase distribution and dynamics, bacterial membrane and chromosomal structure in Escherichia coli. Methods and Applications in Fluorescence. 2015; 3(1):014005. PMID: 29148479

78. Foo YH, Spahn C, Zhang H, Heilemann M, Kenney LJ. Single cell super-resolution imaging of E. coli OmpR during environmental stress. Integrative Biology. 2015; 7(10):1297–1308. https://doi.org/10.1039/c5ib00077g PMID: 26156621

79. Rust MJ, Bates M, Zhuang X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nature Methods. 2006; 3(10):793–796. https://doi.org/10.1038/nmeth929 PMID: 16896339

80. Robinson A, McDonald JP, Caldas VEA, Patel M, Wood EA, Punter CM, et al. Regulation of Mutagenic DNA Polymerase V Activation in Space and Time. PLOS Genetics. 2015; 11(8):e1005482. https://doi.org/10.1371/journal.pgen.1005482 PMID: 26317348

81. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science. 2006; 313(5793):1642–1645. https://doi.org/10.1126/science.1127344 PMID: 16902090

82. Biteen JS, Thompson MA, Tselentis NK, Bowman GR, Shapiro L, Moerner WE. Super-resolution imaging in live *Caulobacter crescentus* cells using photoswitchable EYFP. Nature Methods. 2008; 5(11):947–949. https://doi.org/10.1038/NMETH.1258 PMID: 18794860

83. Ingargiola A, Bordalo BDA. OpenSMFS/pycorrelate: Version 0.3.2; 2019. Available from: https://doi.org/10.5281/zenodo.2653506.

84. Swulius MT, Jensen GJ. The Helical MreB Cytoskeleton in Escherichia coli MC1000/pLE7 Is an Artifact of the N-Terminal Yellow Fluorescent Protein Tag. Journal of Bacteriology. 2012; 194(23):6382–6386. https://doi.org/10.1128/JB.00505-12 PMID: 22904287

85. Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions. vol. 196. John Wiley & Sons; 2011.

86. Müllner D. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. Journal of Statistical Software, Articles. 2013; 53(9):1–18.

87. Storn R, Price K. Differential Evolution—A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. Journal of Global Optimization. 1997; 11(4):341–359. https://doi.org/10.1023/A:1008202821328

88. Perkel JM. Data visualization tools drive interactivity and reproducibility in online publishing. Nature. 2018; 554:133. https://doi.org/10.1038/d41586-018-01322-9 PMID: 29388968

89. Gunsolus IL, Hu D, Mihai C, Lohse SE, Lee Cs, Torelli MD, et al. Facile method to stain the bacterial cell surface for super-resolution fluorescence microscopy. The Analyst. 2014; 139(12):3174–3178. https://doi.org/10.1039/c4an00574k PMID: 24816810

90. Lu HP, Xun L, Xie XS. Single-Molecule Enzymatic Dynamics. Science. 1998; 282(5395):1877–1882. https://doi.org/10.1126/science.282.5395.1877 PMID: 9836635

91. Skinner SO, Sepúlveda LA, Xu H, Golding I. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent *in situ* hybridization. Nature Protocols. 2013; 8(6):1100–1113. https://doi.org/10.1038/nprot.2013.066 PMID: 23680982

92. Edelstein AD, Tsuchida MA, Amodaj N, Pinkard H, Vale RD, Stuurman N. Advanced methods of microscope control using µManager software. Journal of Biological Methods. 2014; 1(2):e10. https://doi.org/10.14440/jbm.2014.36 PMID: 25606571

93. Edelstein A, Amodaj N, Hoover K, Vale R, Stuurman N. Computer Control of Microscopes Using µManager. Current Protocols in Molecular Biology. 2010; 92(1):14.20.1–14.20.17.