Population length variability and nucleoid numbers in *Escherichia coli*

Chaitanya A. Athale* and Hemangi Chaudhari

[1]Division of Biology, IISER Pune, Central Tower, Sai Trinity, Pashan, Pune 411021, India

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Cell sizes and shapes are a fundamental defining characteristic of all cellular life. In bacteria like *Escherichia coli*, the machinery that determines cell length is complex and interconnected, spanning extracellular cues, biosynthesis and cell division. Few tools exist to study cell lengths in a population. We have developed and tested three automated image analysis routines on growing *E.coli* cultures to simultaneously measure cell lengths and nucleoid numbers in populations of bacteria. We find population profiles changing with culture density—higher density of culture leads to fewer long cells. Additionally, lab strains mutant for recA show a correlation between the number of nucleoids and cell length.

Contact: cathale@iiserpune.ac.in; chaitanya.athale@gmail.com

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Cell shape, size and its regulation are a fundamental property of cellular organisms. *Escherichia coli* is a rod-shaped bacterium and multiple genetic factors have been isolated over the years that affect its shape and size (reviewed in Osborn and Rothfield, 2007). Cell size in *E.coli* can be characterized as cell length and width. While width has been observed to be more or less constant, length of the bacteria has been shown to vary even in cells in the same environment and of the same age (Cullum and Vicente, 1978). For instance, the proportion of long cells has been shown to decrease with increasing cell density (Maclean and Munson, 1961). Decreasing temperature caused increased cell volume (Shehata and Marr, 1975) while more recently cell lengths were observed to decrease at 22°C compared with 37°C (Trueba *et al.*, 1982). The filamentous phenotype corresponding to elongated *E.coli* cells has been generated by inhibition, mutation or deletion of penicillin binding proteins, cell scaffolding proteins Mre and Mbl, cell division proteins FtsZ and MinCDE and nucleoid occlusion (reviewed in Goehring and Beckwith, 2005; Lutkenhaus, 2007). However, the population variability in such length measurements has been lacking.

Thus, reproducible quantification of cell sizes and subcellular structures is important to address combinations of effects of genetic and environmental factors on cell phenotype. Recent studies examining bacterial cell lengths have used cell sorters and particle analyzers to measure bacterial cell size (Julià *et al.*, 2010) and cell

cycle stage (Allman *et al.*, 1991). However, these methods lack spatial information about the cells or their subcellular structures. Microscopy and image analysis appear to be the optimal solution. Guberman *et al.* (2008) have estimated the accuracy of cell division in *E.coli* to be 2.9% by measuring the SD in the pinch position as a fraction of cell length in phase contrast images. Using green fluorescent protein (GFP) expressing cells, the SD of cell width was estimated as 6.6% (Männik *et al.*, 2009). More recently, a high-content image analysis tool has been applied to measuring the *E.coli* cell length variability and MinD oscillations (Sliusarenko *et al.*, 2011). The methods used works well on phase contrast and fluorescence images, but not on differential interference contrast (DIC) images.

Bacterial genomic DNA that is compacted in the cell is referred to as the 'nucleoid', distinguishing it from the membrane-bound nucleus found in eukaryotes. Nucleoids form an important part of the cell division machinery, as has been seen recently seen in their role in determining the positioning of the FtsZ ring necessary for equational division in *E.coli* (Bernhardt *et al.*, 2005). In *Bacillus*, incomplete DNA replication causes filamentation of cells (Bernard *et al.*, 2010) by 'nucleoid occlusion', i.e. the steric hindrance of the cell division machinery by nucleoids.

We have developed an algorithm to detect cell lengths in DIC microscopy images of populations of *E.coli* laboratory strains grown in liquid batch culture. We combined this with automated detection of nucleoids in fluorescence images. We confirmed with previous findings that the proportion of long cells decreases with increasing culture density. Additionally, the lab strain DH5α with attenuated recombinase activity shows a linear correlation between cell length and nucleoid numbers. We hypothesize that a higher rate of replication fork stalling in such strains results in the inhibition of cell division, while cell elongation continues leading to longer cells. The algorithm presented here demonstrates our ability to detect and quantify cell length variability in *E.coli* and relate it to the variability in nucleoid numbers. Changing properties like nutrient availability per cell appears to affect this population variability, in a manner consistent across strains.

2 METHODS

Bacterial strain: the strains used in the study were *Escherichia coli* K12 (MG1655), HB101 and DH5α grown in Luria Bertani (LB) rich liquid broth at 37°C with shaking.

Sample preparation: an inoculum of a single colony of *E.coli* in 10 ml LB was grown at 37°C, with shaking at 170 r.p.m., overnight. A 1:100 dilution of this saturated culture was grown in LB and incubated at 170 r.p.m. and 37°C. For staining the nucleoids, 0.1 µg/ml of 4′, 6-Diamidino-2-phenylindole

---

*To whom correspondence should be addressed.

dihydrochloride (DAPI, Sigma-Aldrich) was added to an aliquot and grown further for 15 min. The cells were resuspended in 1:4 (vol/vol) of methanol and the sample was placed on a glass microscopy slide (Medicos Appliances Centre, Pune, India) and allowed to dry. The sample was mounted using 2 µl of mountant [20 mM Tris, pH 8, 0.5% *N*-propyl gallate (Sigma-Aldrich), 90% glycerol] and used for microscopy.

Microscopy: an epifluorescence Zeiss Axio Imager Z1 (Carl Zeiss, Germany) upright microscope with EC Plan-Neofluar 40× lens of numerical aperture of 0.75 was used. Images were acquired using the Zeiss Image acquisition program (Axiovision version 4.8) in the DIC and fluorescence modes. The filter cube sets for DAPI excitation and emission were used for fluorescence measurements based on a mercury short arc lamp light source (X-Cite Series 120, Lumen Dynamics Inc., Canada).

Image processing: three strategies were used for image analysis of *E.coli* DIC images, based on either (i) previously described methods of automated edge detection (Gonzalez *et al.*, 2004) or (ii) in-house developed shadow and (iii) gradient detection methods as described in Section 3. All algorithms were implemented in MATLAB version 7, R14 (Mathworks Inc., MA, USA) in combination with the Image Processing Toolbox. Algorithms were compared to manual (averaged over three people) detection (a comparison with membrane labeled FM4-64 labeling demonstrates it to be a good measure of accuracy, Supplementary Fig. S1).

3 RESULTS

3.1 Algorithm

3.1.1 Edge detection Based on MATLAB ImageProcessing Toolbox (Mathworks Inc., MA, USA) recommended procedures (Gonzalez *et al.*, 2004), we detected cells using the Canny Edge Detector (Canny, 1986), followed by dilation, hole-filling and border clearing (Fig. 1A, i–iv). The segmented cell regions were then subjected to a skeletonization routine, which thinned the elongated structures to a 1-pixel wide contour. This was used for quantification of lengths (Fig. 1A, v).

3.1.2 Shadow detection The topographic appearance of the DIC image shows a prominent dark shadow as seen in the 2D images. The shadow was detected by histogram equalization to enhance contrast, median filtering, followed by a 2D wiener filter and an optimized

Fig. 1. The detection of cell length by the methods tested is applied to a representative DIC image of bacteria (i). (**A**) Edge detection generates binary edge information (ii), which by dilation (iii) and hole filling (iv) and thinning results in skeletons and cell lengths (v). (**B**) Shadow detection works by contrast enhancement (ii), smoothing (iii), thresholding to generate a binary image (iv) and object detection (v) for length measurement. (**C**) Gradient detection works by contrast enhancement (ii), smoothing by filtering (iii), combined XY-gradient detection (iv) and thinning to produce a skeleton (v) and cell length. The scale bar corresponds to 4 μm.

intensity threshold (Fig. 1B i–iv). Inbuilt object detection (Fig. 1B,v) was used to extract perimeter (P) and minor axis (l) lengths for closed objects in the binary image. From this, the length (L) of the bacterium was determined by $L = P/2 - \pi \cdot l/2$, assuming a spherocylindrical geometry (Fig. 1B). This method is used to account for the deviations from a rigid rod.

3.1.3 Gradient detection In order to improve the detection of the 3D perspective shadow of DIC images, after contrast enhancement and smoothing (Fig. 1C i–iii), a pixel-wise gradient in X ($\delta x_{i,j}$) and Y ($\delta y_{i,j}$) is calculated from the difference between a pixel and the values of neighbours. The neighbouring pixel value is an average of n neighbours and the gradients are:

$$\delta x_{i,j} = \left| I_{i,j} - \left\langle I_{(i:i+n,(j+1):(j+n+1))} \right\rangle \right|$$

and

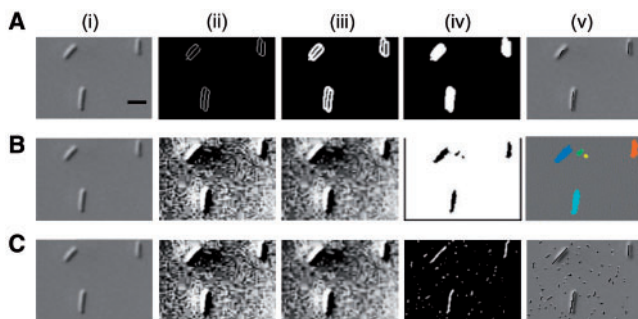$$\delta y_{i,j} = \left| I_{i,j} - \left\langle I_{((i+1):(i+n+1),j:j+n)} \right\rangle \right|$$

The indices *i* and *j* are the coordinates of a pixel, and n is the size of the averaging mask. The value of n was optimally set to $n = 3$ (Supplementary Fig. S2). The two gradient images were made binary by a threshold (τ) to obtain two binary-images D_x and D_y where

$$D_x \begin{cases} 0, \delta x_{i,j} > \tau \\ 1, \delta x_{i,j} \leq \tau \end{cases} \quad \text{and} \quad D_y \begin{cases} 0, \delta y_{i,j} > \tau \\ 1, \delta y_{i,j} \leq \tau \end{cases}$$
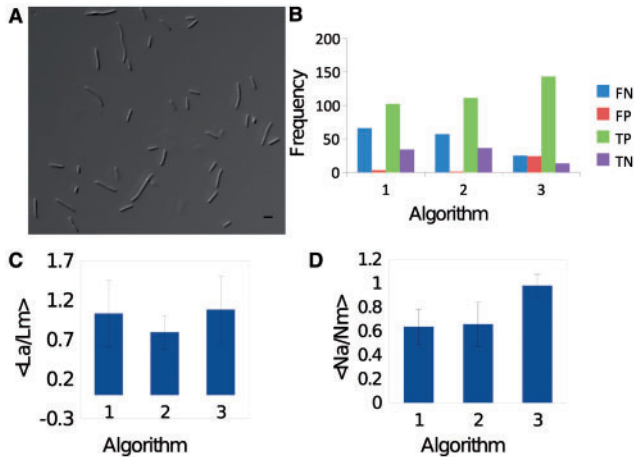
These were combined to obtain a binary image D_{xy}, where $|D_{xy}| = |D_x| \times |D_y|$ (Fig. 1C, iv). These binary contours were thinned to obtain skeletons, which correspond to length of the cell (Fig. 1C, v). Objects with a length below a threshold of 1.5 μm were removed.

3.1.4 Cell crossing-over detection Occasional crossing-over of bacteria was observed in the images (Supplementary Fig. S3A). To avoid overestimating lengths, a pixel-wise branch detection method was developed. Using the skeletonized image as an input, neighbouring pixels and their location were detected using a deconvolution filter. End points of the skeleton are those pixels with only one local neighbour in a neighbourhood of 8 (Moore's Neighbourhood). Branched structures were defined as skeletons with more than two end points. Using an end point as the initial location, the algorithm steps pixel-wise by detecting successive neighbours and calculating a running average of the angle of the skeleton ($\langle\theta_S\rangle$). At branch points, the angle of each neighbour is calculated (θ_N) where $N = 2 - 7$. That neighbour is chosen that minimizes the difference $\delta\theta_N = |\theta_N - \langle\theta_S\rangle|$. In case multiple neighbours have the same value for $\delta\theta_N$, a random choice is made. The process results in as many skeletons as end points (Supplementary Fig. S3B). The skeleton with the lowest SD of the mean angle is finally selected (Supplementary Fig. S3C), while the others are discarded.

3.1.5 Length frequency and phenotypic noise The data obtained from image analysis were plotted as frequency distribution of cell length. The prominent feature of the data is its long tail, which was fit to an exponential function $f(L) = A \cdot e^{\lambda}$. Although other functions such as lognormal could fit the entire distribution, the optimized length threshold of our detection algorithm (1.5 μm) could lead to artefacts with respect to very short cells. The exponential function provides us a measure of the population length distribution of long cells. We also calculate the intrinsic phenotypic noise (Ozbudak

Fig. 2. (**A**) Images of *E.coli* cells from DH5α strains showing a population of cells both long and short fixed on a slide. (**B**) The false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN) were quantified for the three algorithms in comparison with manual detection. The ratios of the algorithm:manual detection of (**C**) numbers of cells (Na/Nm) and (**D**) lengths (La/Lm) are plotted.

et al., 2002) as $\eta_L = \frac{\sigma_L^2}{\langle L \rangle}$, where $\langle L \rangle$ is the mean length and σ_L^2 is the variance.

3.1.6 Nucleoid detection Bacterial nucleoids were stained with DAPI, a DNA binding fluorescent dye, and fixed (see Section 2, sample preparation). Images of cells in the same field of view were acquired in DIC and fluorescence channels. The gradient detection algorithm was run to produce a skeleton. Since nucleoids occupy the entire diameter of the cell, a 5-pixel wide dilated skeleton mask was obtained from the DIC images. These coordinates were used to generate mean intensity profiles of the DAPI image along the cell length. This 1D intensity profile was then processed for peak detection using a simple valley threshold. If a point is at the edge of the valley deeper than the threshold, it is considered a peak. These peaks correspond well with the centroids of nucleoids.

3.2 Testing the algorithms

The evaluation of the algorithms was performed on images of bacteria obtained from a liquid broth and mounted on slides (Fig. 2A). While all three algorithms worked comparably well on isolated cells, using populations of cells differentiated one method over all others.

3.2.1 Number and length comparison The number of bacteria detected manually was considered to be the standard. The first two methods, edge detection and shadow detection, showed higher false negatives than the gradient method (Fig. 2B). The increase in false positives in gradient detection is related to its extreme sensitivity to out of focus cells that manual detection ignored. The true positives and true negatives in turn are, respectively, highest and lowest in gradient detection. The ratio of the number of bacteria automatically detected in each image (Na) by the manual value (Nm) (Fig. 2C) and the length ratios (Fig. 2D) were averaged for four images. Here too, we noticed that gradient detection performed better than the other two and is our method of choice. Cell counts obtained are
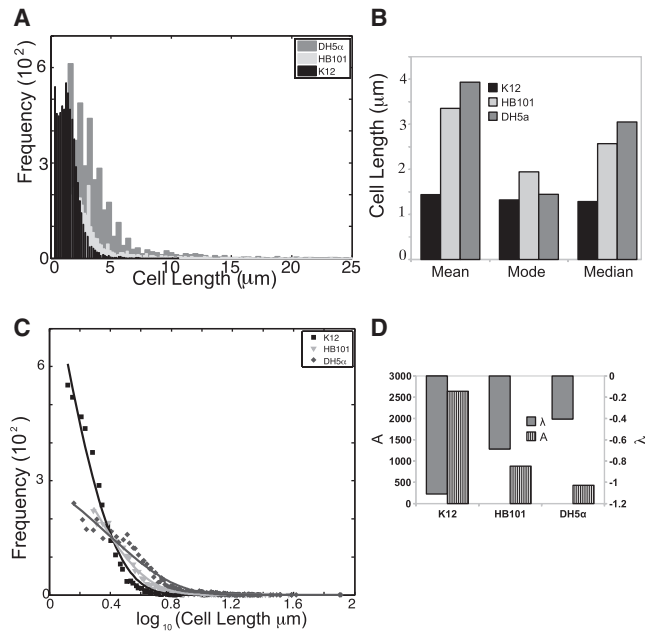

Fig. 3. (**A**) The frequency distribution of cell lengths for *E.coli* cells of strain K12 (*n* = 6549), HB101 (*n* = 2942) and DH5α (*n* = 3699). (**B**) The mean, mode and median of the lengths of three strains are compared. (**C**) The measured frequency of log cell lengths is plotted for K12 (black squares), HB101 (light grey triangles) and DH5α (grey diamonds) strains and the continuous lines are the exponential fits, (**D**) with the two parameters A (grey) and λ (stripes) for the three strains.

the best among the three methods tested, while cell lengths are still comparable between edge detection and gradient detection. The method of gradient detection is sensitive to the angle of the prism. In all our experiments, it is tuned to produce sufficient contrast, which generates the shadow along the length of the bacterial cells covering half its width.

3.3 E. coli length variability within lab strains

3.3.1 Length comparison *E.coli* K12 (MG1655) cells showed a frequency maximum of ∼2 μm with a sharp exponential drop for frequencies of longer lengths. In contrast, the E.coli strains HB101 and DH5α appeared to show a greater proportion of long cells as demonstrated by the longer tails in their distributions (Fig. 3A). We observed a difference in the mean and median lengths for the three strains while the mode length value was comparable (Fig. 3B). The fit to the length distributions was a scaled exponential with two parameters A and λ (Fig. 3C). The values for A and λ compared across the profiles show a definitive fold-change for DH5α over K12 (Fig. 3D). In addition, the percentage of long cells (length >8 μm) were low in K12 as expected, while DH5α and HB101 showed higher frequencies. The values of phenotypic noise (η_L) measured for the strains K12, HB101 and DH5α were 0.54, 2.17 and 3.43, respectively.

3.3.2 Effect of culture density To determine if this was dependent on cell culture density, we compared the density with the proportion of long cells and found that this value decreases with increasing culture density in growing cultures (Fig. 4A) of all three strains
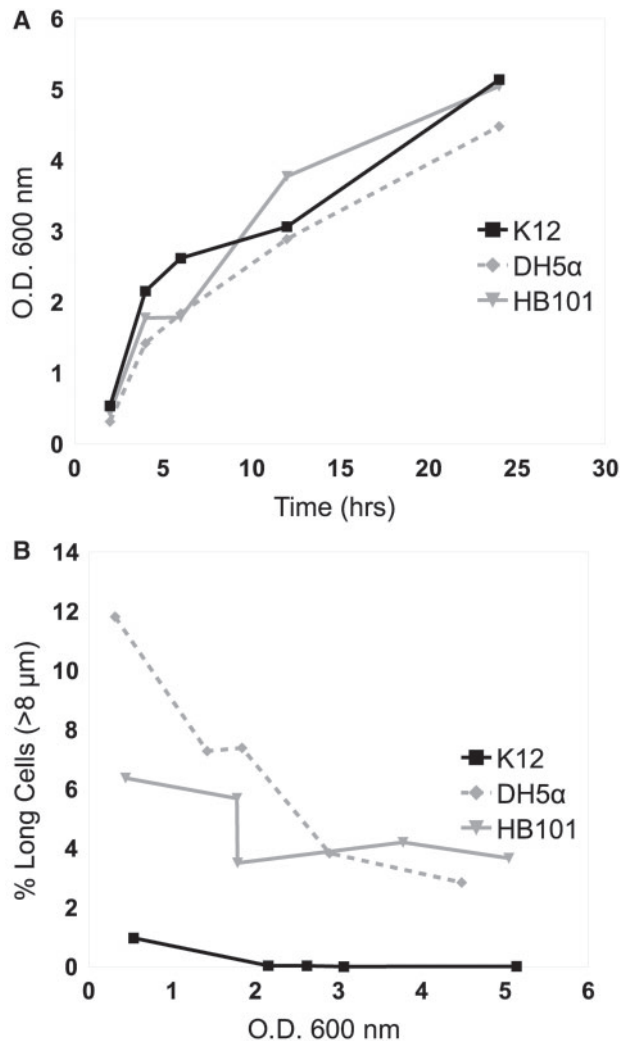
Fig. 4. (**A**) Growing cultures of three strains of *E.coli* K12, HB101 and DH5α in rich medium with shaking at 37˚C were sampled for cell length. (**B**) The increasing density of the culture corresponds with a decrease in the percentage of long cells (length >8 μm) for all three strains.

(Fig. 4B). This is consistent with the previous reports for *E.coli* B/r (MacLean and Munson, 1961). Thus, the difference between the strains K12, HB101 and DH5α lies in the quantitative change in the exponential fit to the length distributions, not in a novel distribution type.

3.3.3 Nucleoids   Using the cell length and nucleoid information from K12 (Fig. 5A) and DH5α (Fig. 5B) strains, we correlate length with nucleoid numbers (Fig. 5C). The wild-type strain K12 shows little or no correlation between the two, while the distribution of DH5α cells shows a linear correlation with a slope of 2.2 ($R^2 = 0.8$). This is the first time to our knowledge that such a correlation has been quantified in a mutant recA1 containing *E.coli* strain.

4   DISCUSSION

The image analysis algorithm presented here detects cell lengths and nucleoids in rod-shaped and filamentous bacterial populations.
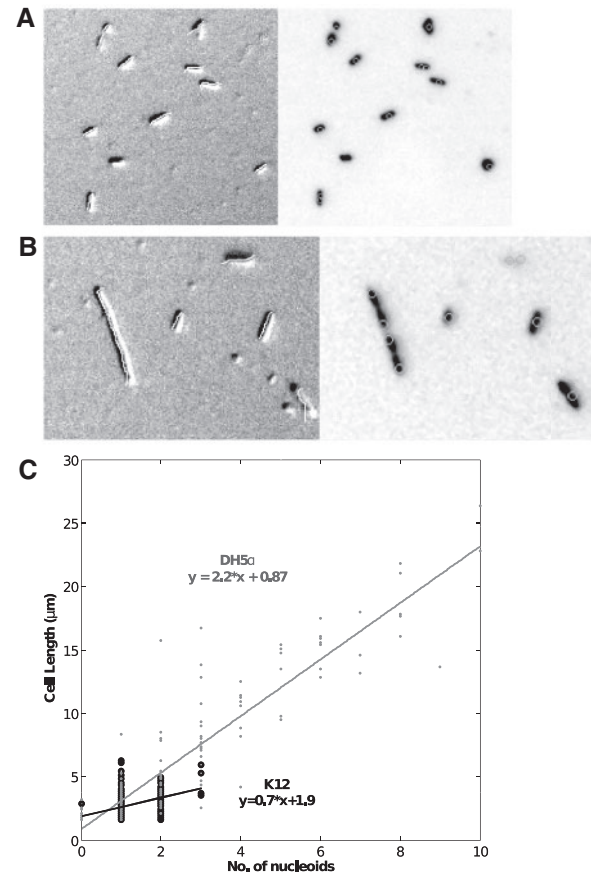
Fig. 5. Representative inverted images show detected lengths (white contours) in DIC images and nucleoids detected (outlined) in DAPI-stained cells of *E.coli* (**A**) K12 and (**B**) DH5α. (**C**) The plot of nucleoid number and corresponding cell length for K12 (black circles) ($n = 278$) shows little correlation (linear fit with $R^2 = 0.1$), while DH5α (grey dots) ($n = 196$) shows a linear trend (linear fit $R^2 = 0.8$) with slope 2.2 (please refer online for colour figures).

Automated image analysis is making it possible to relate the microscopic subcellular details of cells to phenotypic- and population-level properties (Sliusarenko *et al.*, 2011). Developing and optimizing algorithms as these therefore assumes more importance than before. A correlative approach between the cell morphology as observed in DIC and the DNA content has been performed before to address the role of nucleoids in bacterial cell division (Bernard *et al.*, 2010; Bernhardt *et al.*, 2005), but these approaches use interactive methods of quantification. DIC optics generates contrast based on gradients in the optical path length simplifying the detection of elongated objects such as rod-shaped bacteria. The method described here can be used on fixed cells, live cells in liquid or on agar pads.

Some recent studies have addressed variability in cell division at a population level, but these work best with phase contrast images, and ignore the role of the nucleoid (Guberman *et al.*, 2008; Sliusarenko *et al.*, 2011). Here we find *E.coli* K12 (MG1655) mean cell lengths show a SD of 40–60% of cell length, while HB101 and DH5α show 65–80% and 70–90%, respectively. The exponential fit to the length

frequency distribution allows us to quantify both the steepness of the distribution and the length of the tail. The long tail and shallow decay in length frequencies in DH5α and HB101 might be related to their common genetic background of defective recombinase A (recA). We speculate that since the replication fork stalling is stochastic, it will lead to a non-Gaussian distribution of the cell lengths by increasing the proportion of the longer cells. Thus, a RecA1 mutant strain like DH5α that has higher probabilities of stalling has a longer tailed distribution.

Our observation in growing liquid cultures of a decrease in percent long cells with increasing cell density in all three strains appears to suggest a common growth rate limited effect, independent of mutational background. This is consistent with the previous findings from continuous cultures (Maclean and Munson, 1961). The mutant strains, however, appear to have a quantitative amplification of this effect. Since cell elongation depends on nutrient availability, we can hypothesize that higher cell densities lead to lower frequency of long cells.

In all, it would appear to suggest such population studies in mutant strains might shed light on the robustness of *E.coli* K12 length regulation. Previously, bacterial strains lacking RecA were identified as 'noisy' at a gene expression level (Elowitz *et al.*, 2002). Here, we find phenotypic noise in terms of cell length to also increase in the presence of a recA1 mutation. This would appear to suggest the noise at the level of cell morphology and the gene expression variability might be related. Additionally, this study demonstrates how environmental factors like nutrient availability might modulate the variability. In future, it would be interesting to develop a theoretical model to study the consequences of replication fork stalling, cell elongation and nucleoid occlusion on population length variability. Additionally, experiments examining the single cell behaviour with tracking of individual cell lineages might yield further insights. This study allows us to begin to connect environmental effects with single cell division and replication.

ACKNOWLEDGEMENTS

REFERENCES

Allman,R. *et al.* (1991) Cell cycle parameters of Escherichia coli K-12. *J. Bacteriol.*, **173**, 7970–7974.

Bernard,R. *et al.* (2010) Nucleoid occlusion prevents cell division during replication fork arrest in Bacillus subtilis. *Mol. Microbiol.*, **78**, 866–882.

Bernhardt,T.G. *et al.* (2005) SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over chromosomes in *E. coli*. *Mol. Cell*, **18**, 555–564.

Canny,F.J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, **8**, 679–698.

Cullum,J., and Vicente,M. (1978) Cell growth and length distribution in Escherichia coli. *J. Bacteriol.*, **134**, 330–337.

Elowitz,M.B. *et al.* (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.

Goehring,N.W. and Beckwith,J. (2005) Diverse paths to midcell: assembly of the bacterial cell division machinery. *Curr. Biol.*, **15**, R514–R526.

Gonzalez,R.C. *et al.* (2004) *Digital Image Processing Using MATLAB*. 1st edn. (Pearson), Prentice Hall, NJ, USA.

Guberman,J.M. *et al.* (2008) PSICIC: noise and asymmetry in bacterial division revealed by computational image analysis at sub-pixel resolution. *PLoS Comput. Biol.*, **4**, e1000233.

Julià,O. *et al.* (2010) Skew-laplace and cell-size distribution in microbial axenic cultures: statistical assessment and biological interpretation. *Int. J. Microbiol.*, **2010**, 191585.

Lutkenhaus,J. (2007) Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. *Annu. Rev. Biochem.*, **76**, 539–562.

Maclean,F.I. and Munson,R.J. (1961) Some environmental factors affecting the length of *Escherichia coli* organisms in continuous cultures. *J. Gen. Microbiol.*, **25**, 17–27.

Männik,J. *et al.* (2009) Bacterial growth and motility in sub-micron constrictions. *Proc. Natl Acad. Sci. USA*, **106**, 14861–14866.

Osborn,M.J. and Rothfield,L. (2007) Cell shape determination in Escherichia coli. *Curr. Opin. Microbiol.*, **10**, 606–610.

Ozbudak,E.M. *et al.* (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.*, **31**, 69–73.

Shehata,T.E. and Marr,A.G. (1975) Effect of temperature on the size of Escherichia coli cells. *J. Bacteriol.*, **124**, 857–862.

Sliusarenko,O. *et al.* (2011) High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol. Microbiol.*, **80**, 612–627.

Trueba,F.J. *et al.* (1982) Effects of temperature on the size and shape of *Escherichia coli* cells. *Arch. Microbiol.*, **131**, 235–240.