Harnessing cell-to-cell variations to probe bacterial structure and biophysics

Julie A. Cass

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Paul Wiggins, Chair

Beth Traxler

Marcel den Nijs

Program Authorized to Offer Degree:
Department of Physics

University of Washington

Abstract

Harnessing cell-to-cell variations to probe bacterial structure and biophysics

Julie A. Cass

Chair of the Supervisory Committee:
Professor Paul Wiggins
Physics Department

Advances in microscopy and biotechnology have given us novel insights into cellular biology and physics. While bacteria were long considered to be relatively unstructured, the development of fluorescence microscopy techniques, and spatially and temporally resolved high-throughput quantitative studies, have uncovered that the bacterial cell is highly organized, and its structure rigorously maintained. In this thesis I will describe our `gateTool` software, designed to harness cell-to-cell variations to probe bacterial structure, and discuss two exciting aspects of structure that we have employed `gateTool` to investigate: (i)chromosome organization and the cellular mechanisms for controlling DNA dynamics, and (ii) the study of cell wall synthesis, and how the genes in the synthesis pathway impact cellular shape. In the first project, we develop a spatial and temporal mapping of cell-cycle-dependent chromosomal organization, and use this quantitative map to discover that chromosomal loci segregate from midcell with universal dynamics. In the second project, I describe preliminary time-lapse and snapshot imaging analysis suggesting phentoypical coherence across peptidoglycan synthesis pathways.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

GLOSSARY

CHROMOSOME: The circular moelcule of DNA possessed by prokaryotic cells

NUCLEOID: complex of condensed nucleic acids (chromosome) and associated proteins

LOCUS (LOCI): specific genetic region (regions) of the chromosome

SEGREGATION: the process of moving sister loci to opposite cell halves prior to cell division

PHENOTYPE: the characteristic traits of a bacterial strain

MORPHOLOGY: the characteristics of cell shape

KL DIVERGENCE: the penalty for modeling one distribution, p(x), with a another distribution, q(x)

P-VALUE: the probability of measuring an effect equal to or 'more extreme' than the observed effect, if the null hypothesis is true

NULL HYPOTHESIS: the hypothesis of no significant effect or change

SEGMENTATION: identification of cell boundaries in an image

GATING: selecting subsets of data, by setting quantitative cutoff values

KERNEL DENSITY ESTIMATE (KDE): probability density smoothed via convolution with a Gaussian kernel; prevent artifacts from binning

PG: peptidoglycan; sugar-amino acid polymer forming mesh-like cell wall around baceria

L-FORMS: bacteria lacking cell walls, resembling early evolutionary precursors to modern bacteria

ACKNOWLEDGMENTS

My research advisor Paul Wiggins, whose enthusiasm for biophysics sparked my own passion for the field, and set the path for my career. I came to the University of Washington never having studied biology; through his extensive knowledge and deep passion for research, Paul taught me to see each challenge as 'a feature, not a bug.'

The doctoral committee, including Beth Traxler, Deep Gupta, Jens Gundlach, Marcel den Nijs and Lutz Maibaum, for their guidance in my research. I am especially grateful to my fantastic mentors and collaborators Beth Traxler and Nathan Kuwada, who taught me much of what I know about biology and supported me in many ways.

My collaborators Jeannie Bailey and Colin Manoil for teaching me so much about new bacteria and genetic methods, and for being such a pleasure to work with.

My labmates Stella, Colin, Sarah and Jackie, for learning and working with me, as well as teaching me, throughout our time in the Wiggins lab.

My graduate advisor Catherine Provost who was there for me through exams, fashion crises, and life.

My close friends from my cohort, who have helped me through through graduate school since day one of this program. I'm also thankful for my 'Capitol Hill friends,' for making me laugh, and for keeping me sane.

My partner Carey, who has provided so much patience, care, and happiness throughout my doctoral work.

Most importantly, my family. I'm grateful to my extended family and to Maggie for their love and support. My whole life, my parents Chris and Brenda encouraged me to believe that any future was possible, with hard work and passion. They made countless sacrifices to

support me through my education in every way, and I cannot express what a gift this has been. My sister Laura showed me how to become a scientist, and provided so much care and guidance. Together my family made it possible for me to get here, and I am so thankful.

DEDICATION

to my parents, Brenda and Chris Cass,

who have supported my education and love of science in every way.

PUBLICATIONS

This thesis is based on the following articles:

(i) Cass, J.A., Kuwada, N.J., Traxler, B., and Wiggins, P.A. (2016) *Escherichia coli* chromosomal loci segregate from midcell with universal dynamics. Biophys J 110: 2597 2609.

(ii) Cass, J.A., Stylianidou, S., Kuwada, N.J., Traxler, B., and Wiggins, P.A. (2017) Probing bacterial biology using image cytometry. Molecular Microbiology 103(5): 818-828.

(iii) Bailey J., Cass, J.A., Wiggins, P.A. and Manoil C. (2017) Phenotypic coherence of mutants in the peptidoglycan synthesis pathway of *Acinetobacter baylyi*. *(In preparation.)*

(iv) Mangiameli S., Cass, J.A., Merrikh, H., Wiggins, P.A. *Bacterial replisomes show factory-like positioning. (In preparation for Current Genetics, 2017.)*

(v) Cass, J.A., Kuwada, N.J., Traxler, B., and Wiggins, P.A. (2017) Phase changes of the chromosomal structure. *(In preparation.)*

Chapter 1

INTRODUCTION

In the last quarter century, major advancements in multi-disciplinary approaches to physical science have enhanced our understanding of living systems. To improve our understanding of the world around us, we must see our surroundings through different scientific lenses; by supplementing our biological studies with the quantitative tools of physics and computation, we employ a more robust toolkit for probing living systems.

An appealing subject for these multi-disciplinary studies is the elucidation of mechanisms for maintaining structure and dynamics in bacteria. Contrary to the long-assumed model that these 'simple' cells are largely disordered, bacterial cell structure is highly organized and rigorously maintained both spatially and temporally. But while novel multidisciplinary techniques have uncovered this striking organization of bacterial cells, the mechanisms responsible for this structure remain unknown. Furthermore, bacteria are an order of magnitude smaller that human cells, existing on the micron scale where thermal fluctuations and statistical mechanics become relevant factors in the determination of dynamics. These factors attract physicists and quantitative scientists to the analysis of dynamics in the bacterial cell environment. In this thesis, I will discuss how wide cell-to-cell variations in cellular structure act both as a feature and a bug in our attempts to characterize the mechanisms guiding bacterial cell structure.

1.1 Outline

Cellular diversity arises for several reasons, including (i) evolution and (ii) scale. The survival of biological systems is achieved through small variations across populations; phenotypic variability can enhance fitness and act as a bet-hedging strategy in unpredictable environments

[5]. Those sub-populations with the most favorable features for their environment will continue to live and proliferate. In addition, we consider that the world of bacterial cells takes place on the micron scale, where thermal fluctuations play an appreciable role in dynamics [**?**]. Small fluctuations in environment or within the cell can lead to notable variations in sub-cellular localizations.

In this first chapter of this thesis, I outline the basic biological and physical principles and techniques needed for the research described in the following chapters. In Chapter 2, I introduce software developed for the high-throughput characterization of populations of varied bacterial cells, a method called 'image cytometry.' In Chapter 3, I present a study of the sub-cellular structure and dynamics of bacterial DNA, highlighting the challenges introduced by vast cell-to-cell differences in DNA dynamics. In the fourth chapter, I will describe an ongoing study of bacterial shapes, aimed at using timelapse microscopy and diverse cell and colony shapes to study the cell wall synthesis pathway. Finally, with concluding chapter, I will summarize the results of my doctoral work, and discuss directions in which these projects could be expanded upon.

1.2 The Bacterial Life Cycle

1.2.1 Central Dogma

The driving force in the survival of organisms is the ability to proliferate, and transmit information to future generations. To achieve this, all living things possess DNA: an instruction manual for making a copy of itself. Shaped as a double helix, the DNA consists of a sugar and phosphate backbone, linked by base pairs: 'A' and 'T', or 'G' and 'C'. Much as the English alphabet provides a set of letters, whose order gives rise to words and meaning, these bases act as an alphabet for genetic material. The unique ordering of these base pairs forms a set of instructions to create specific proteins. In the case of all bacteria in this thesis, the DNA is structured as a single circular loop, called the chromosome.

1.2.2 The Bacterial Cell Cycle

The bacterial life cycle is carried out through four major processes: i) cellular growth ii) DNA replication iii) DNA segregation and iv) cellular division (Fig. 1.1). During early cell growth, the genes that make up the chromosome are transcribed into RNA and translated by ribosomes into proteins needed for sub-cellular operations. During replication, the cell makes a copy of its chromosome, so that each of its future daughter cells will receive a copy of the DNA. The loci of the chromosome are continuously and sequentially 'segregate' following their replication. In this process, the two 'sister' loci are translocated towards opposite cell halves, in preparation for the division of the cell into two daughters. Finally, during division, a septum of cell wall forms, physically dividing the cell into two nearly identical copies of the original cell. This life cycle, referred to as the 'cell cycle', is then repeated for each of the two resulting cells (daughter cells); the process repeats continuously to build and maintain an entire colony of clonally related cells.

1.2.3 Sub-cellular structure

Appreciating the complexity of the cell cycle requires an understanding of the biology of the bacterial cell, and the length scales where these biological processes are occurring. For example, *Escherichia coli* cells are rod-shaped and approximately 2 μm by 0.5 μm in size. Unlike eukaryotic cells, they have no membrane-bound organelles; they consist mainly of proteins, cytoplasm, and the nucleoid (a condensed bundle of the chromosome and proteins). Its easy to imagine bacteria homogeneous, spacious environment, with the chromosome and proteins diffusing with little order.

Bacteria were long considered to be simple organisms, with the randomly folded chromosome and proteins diffusing throughout the intracellular space. It was only in the last century that advances in light microscopy clearly demonstrated that the chromosome is compacted into a discrete object [6]. This distinct shape convincingly suggested that the chromosome is in structured, suggesting well-regulated and predictable dynamics. The chromosomal

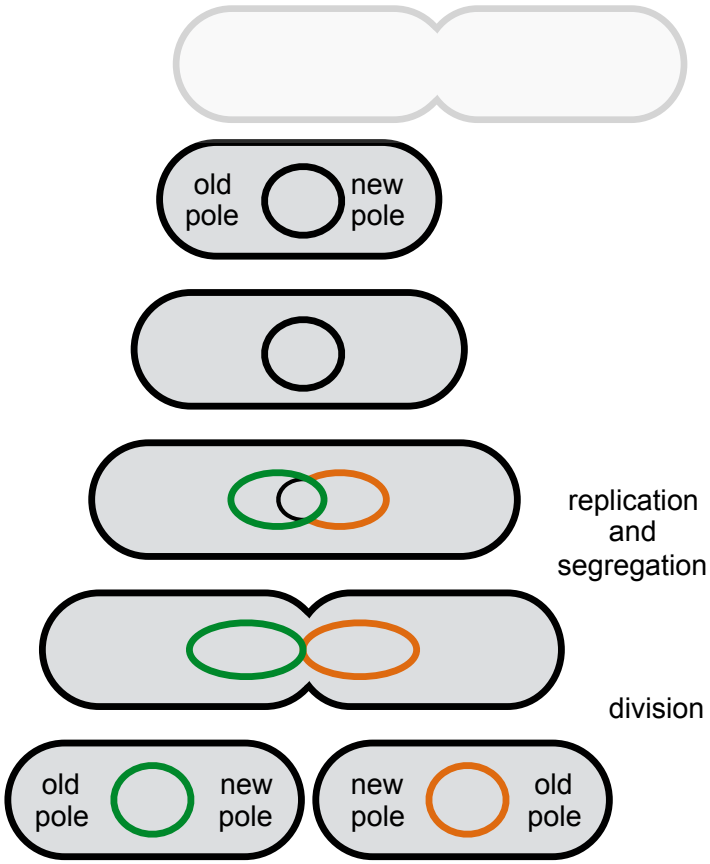

Figure 1.1: **Simple schematic of the bacterial cell-cycle.** In this simplified schematic, we see that each bacterial cell is born through the division of another cell ('mother'). The pole of the cell created through division is the 'new pole' and the pole inherited from the mother cell. Replication and segregation occur simultaneously. A more detailed schematic for replication/segregation in the context of the cell cycle is provided in the Chapter 3.

structure was further elucidated in the last quarter century, through the development of methods to visualize individual chromosomal loci in live cells using fluorescence microscopy [7, 8, 9, 10, 11] . This understanding of the robust spatial organization of the chromosome raised questions and driven recent models for how newly replicated DNA is segregated.

To better understand the biological necessity for well-managed chromosome structure, consider the size of the *E. coli* chromosome: this DNA loop consists of 4.7 million base pairs (Mbp). When stretched out straight, this DNA would cover a length of over $1mm$; the intracellular space it occupies, however, is only $\approx 1\mu$ long. Thus, to fit inside the cell, the chromosome must be linearly compacted more than 1,000-fold [12, 13]. The cell interior is accordingly densely packed (see Fig. 1.2), with the chromosome highly condensed just to fit inside. In order for the cell to copy this DNA and express proteins from it, its structure and dynamics must be highly managed.

This picture is further complicated by considering the life cycle of the cell. As a rod-shaped bacterial cell grows, it elongates along its long axis, and splits down its center to divide into two cells. The genes on the chromosome are continuously being expressed, to create more proteins, with the associated changed in DNA structure that accompany RNA synthesis. The single chromosome is being replicated, and the two copies (sister chromosomes) are being moved to opposite poles. The management of these processes within a densely packed space requires meticulous order of the cellular structure. Which begs the question to physicists and biologists: how does is structure maintained in these environments, and what mechanisms and forces guide this process?

1.3 Dynamics on the micron scale

1.3.1 Diffusion in the cell modeled by a random walk

We can model the diffusion of subcellular elements in the cytoplasm as Brownian particles in a fluid. For particles with number density $n(r,t)$ and velocity v(r,t), the current density will be $j(r,t) = n(r,t)v(r,t)$. According to Fick's Law,

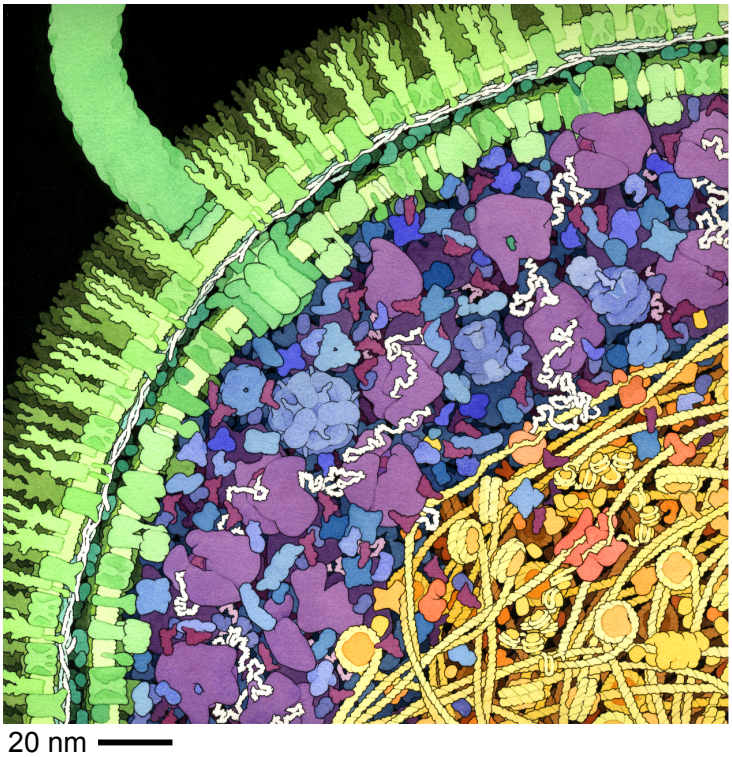

Figure 1.2: **The densely packed sub-cellular environment of *E. coli*.** This illustration shows a cross-section of a small portion of an *E. coli* cell. The cell wall, with two concentric membranes studded with transmembrane proteins, is shown in green. A large flagellar motor crosses the entire wall, turning the flagellum that extends upwards from the surface. The cytoplasmic area is colored blue and purple. The large purple molecules are ribosomes and the small, L-shaped maroon molecules are tRNA, and the white strands are mRNA. Enzymes are shown in blue. The nucleoid region is shown in yellow and orange, with the long DNA circle shown in yellow, wrapped around HU protein (bacterial nucleosomes). In the center of the nucleoid region shown here, you might find a replication fork, with DNA polymerase (in red-orange) replicating new DNA. Illustration by David S. Goodsell, the Scripps Research Institute

$$j(r,t) = -D\nabla n(r,t), \tag{1.1}$$

and the continuity equation,

$$\nabla \cdot j(r,t) + \frac{\partial n(r,t)}{\partial t} = 0, \tag{1.2}$$

we can write the particle number density as:

$$n(r,t) = \frac{N}{(4\pi Dt)^{3/2}} e^{-\frac{r^2}{4Dt}}. \tag{1.3}$$

Here D, commonly referred to as the 'diffusion coefficient,' characterizes the mobility. Calculating the mean squared displacement for diffusing particles in one dimension (i.e. $r = x$) using this number density, we find

$$\langle \Delta x^2 \rangle = \frac{1}{N} \int_0^\infty n(x,t) 4\pi x^4 dx = 2Dt. \tag{1.4}$$

The mean-squared displacement ('MSD') thus increases linearly with lag time.

1.3.2 The step-size distribution

In many cases of biological studies, a valuable measurement is the distribution of the sizes of each step in a random walk. For instance, in Chapter 3 we will consider a model for chromosome segregation by the exploring the distribution in sizes of steps taken by a fluorescently tagged chromosomal locus. In one dimension, the the probability of a certain step size x given the mean μ and the variance σ is

$$p(x|\mu,\sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{1.5}$$

The mean and the variance carry important information about the mobility and bias of the particle (or locus). We will proceed by exploring what each of these features of the step-size distribution tells us.

1.3.3 Variance and the scaling factor

For distributions possessing a zero-mean (i.e. $\mu = \langle \Delta x \rangle = 0$), the variance in the step size distribution $\sigma^2 = \langle \Delta x^2 \rangle - \langle \Delta x \rangle^2$ is simply the mean-squared displacement. In the previous section, we saw that the MSD scales linearly with lag time for diffusive motion ($\sigma^2 = 2Dt$). In contrast, the dynamics on the sub-micron scale inside the cell are often dominated by sub-diffusive motion. That is,

$$\sigma^2 \propto t^{\alpha}, \tag{1.6}$$

where $0 < \alpha < 1$.

1.3.4 Bias and the Langevin equation

The mean of the step size distribution is related to velocity by the time between steps.

$$\langle v(t) \rangle = \frac{\langle \Delta x \rangle}{\tau} = \frac{\mu}{\tau}. \tag{1.7}$$

In a microscopy experiment, where a continuous process is probed as regular intervals, the time between steps is the frame rate (τ). Mechanistically, this bias holds valuable information about the forces inside of the cell. Since we cannot directly probe force magnitudes inside of living cells, the bias is a convenient and valuable way for us to indirectly measure the force. To see how the velocity (or step size distribution bias) relates to the forces, we consider the net force on a locus. The net force is the sum of three types of forces: (i) random forces, defined as those whose ensemble average is zero (ii) viscous drag and (iii) average forces, or those whose ensemble average is non-zero. The thermal fluctuations, which are random forces, are non-negligible due to the small scale (sub-micron) of the dynamics we analyze. The viscous drag is proportional to the velocity of the locus by a factor γ, which is related to the diffusion coefficient. The sum of the forces on the locus can be denoted using the Langevin equation:

$$\sum F = \xi - v\gamma + F_{ave}, \tag{1.8}$$

Since all motion on this scale and in a dense environment is heavily damped, the forces sum to zero. To eliminate the affect of cell-to-cell variations on this force measurement, we must take the ensemble average of each term to determine the average force on a locus. Since the ensemble average of random forces is zero, this leave us with the simple relation:

$$\langle F_{ave} \rangle = \gamma \langle v \rangle. \tag{1.9}$$

Assuming an (approximately) constant diffusion coefficient, the average force is directly proportional to velocity of the locus the forces act on. Measuring the velocity of chromosomal loci provides us with indirect information about the magnitude of forces driving structure in bacterial cells.

1.4 Phenotypes of bacterial cells and cultures

1.4.1 Morphology

A common technique for studying the function of a protein is to delete the gene encoding it. Lacking the gene, the cell can no longer produce the target protein, and we can compare traits of parent and mutant cells. By contrasting these mutant cells to wild-type cells, we can draw conclusions about what functions the missing protein served. A description of a cell's characteristic behavior under certain conditions is called its phenotype; for example, the phenotype of elongation mutants (or cells lacking the genes that encode proteins for cell increasing cell length) is short, round cells. While the term phenotype embraces characteristics beyond physical appearance (*i.e.* behavior), our lab conducts imaging studies, and therefore focuses on the morphology: the aspects of phenotype characterizing the physical traits or shape of a bacterial strain.

1.4.2 Genetics

An advantage to our bacterial experimental systems is the ease with which we can manipulate these cells. Mutant strains of bacteria are constructed through a variety of strategies that

allow us to edit the set of genes carried by bacteria. Two genetic techniques are used for strain development in this thesis: (i) transformation and (ii) transduction.

Genetic transformation is the process of a host organism (bacteria) taking in extracellular (or 'exogenous') DNA. Ultimately, one hopes that a gene or genes encoded by that DNA will be expressed in the transformed cell. Exogenous DNA is received by the recipient cell from its surroundings through the cell membrane; cells which are able to take in DNA across their membranes are referred to as competent. Some bacteria are naturally competent, including *Acinetobacter baylyi*, making them attractive backgrounds for genetic studies (See Chapter 4). However, most bacteria must be made competent through artificial means such as heat, electric or chemical shock.

If the exogenous DNA is a linear fragment, it may be incorporated into the host chromosome by recombination (Fig. 1.3). We use this method in Chapter 4 to delete target genes by inserting antibiotic resistance in their place. If the exogenous DNA is a plasmid (or loop of DNA), the cell machinery replicates and segregates the plasmid so that it is passed on to cell progeny. In Chapters 2 and 3, we use transformations in *E. coli* to introduce plasmids expressing fluorescent proteins. When this fluorescence is localized, producing 'foci' or fluorescent light in the cell, it is used for the tracking of specific genetic loci in the cell (as in Chapter 3). When fluorescence is diffuse throughout the cell (i.e. cytoplasmic fluorescence), it can be used as an identification marker for the cell type (as in Chapter 2 of this thesis).

Transductions for the projects in this thesis were done using the P1 bacteriophage. This phage infects a host cell and injects its DNA through the membrane. Enzymes expressed by the injected DNA degrade the host chromosome into fragments; when progeny phages are assembled inside the infected cell, they sometimes mistakenly package DNA fragments from the host chromosome. The transducial phages (now carrying pieces the previously infected host DNA) inject that DNA into a recipient cell. Now, the exogenous chromosomal genes ejected by the phage will be incorporated into the recipient chromosome by recombination, and progeny of the cell will carry this altered chromsome.

We use this method to delete genes from the *E. coli* genome; bacteriophages are produced

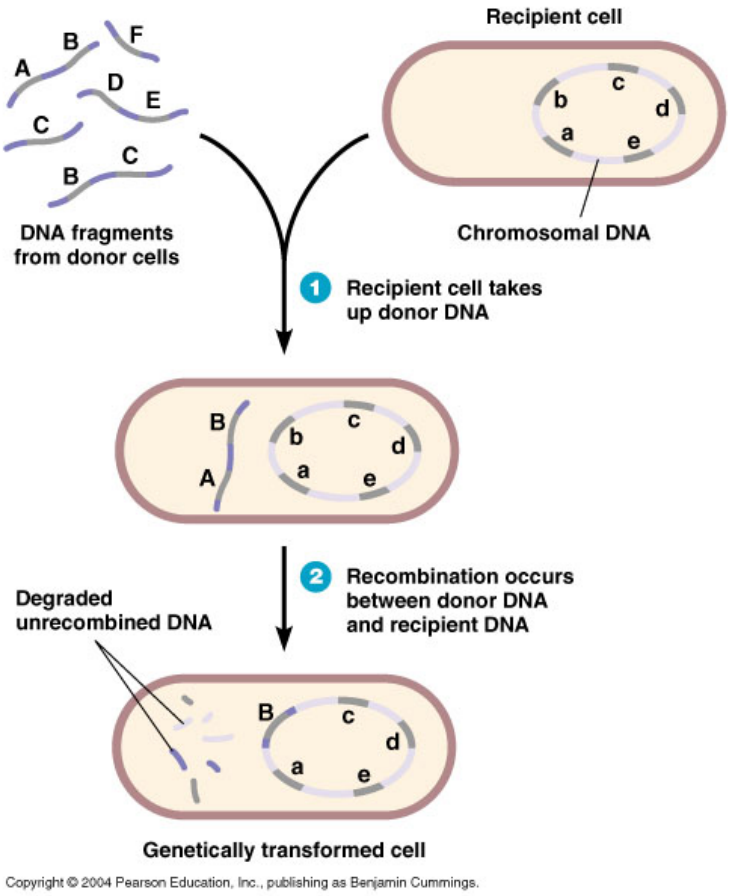

Figure 1.3: **DNA fragment transformation in bacteria.** (1) Competent recipient, or host, cells receive exogenous DNA fragment through their membrane. (2) Through recombination, DNA fragments are incorporated into the host chromosome. The exogenous genes may now be expressed by the cell to create new protein. This figure was produced by Pearson Education Inc.

to replace target recipient cell genes with antibiotic resistances. When the recipient cells are then grown on the antibiotic, only those cells which received the deletion will grow, resulting in an isolated strain possessing the desired deletion. This mutant can then be studied and its phenotype characterized.

1.5 Microscopy and cytometry

In this thesis, the characterization of wild-type and mutant cell phenotypes was carried out through microscopy and quantitative techniques. While imaging cells allows for qualitative characterization, wide cell-to-cell variations in morphology and subcellular dynamics make high-throughput techniques imperative for sifting the signal from the noise. Advancements in microscopy and quantitative tools have allowed us to develop methods to sort through thousands of cells, using the resulting range in cell phenotype as a feature in maximizing experiment efficiency, while applying statistical techniques to ensure that we model the 'typical' dynamics and shapes, rather than the varied features of individual cells.

1.5.1 Brightfield microscopy

The basic morphology of a cell can be observed through phase-contrast, or brightfield, microscopy. In this technique, cells are illuminated with white light; waves passing through the cells are refracted, causing all cell regions to appear in black [14, 15]. The result is a grayscale photo, where the empty regions of the slide are bright and the cell areas are seen as very dark. While this method does not provide any information about sub-cellular structure, it allows for the characterization of cell shapes. As we will discuss in Chapters 2 and 4, quantitation of cell morphologies can provide powerful information about subcellular components.

1.5.2 Fluorescence microscopy

For characterization of sub-cellular dynamics and structure, we use fluorescence microscopy. Through the transformation of plasmids encoding fluorescent proteins, cells are inoculated

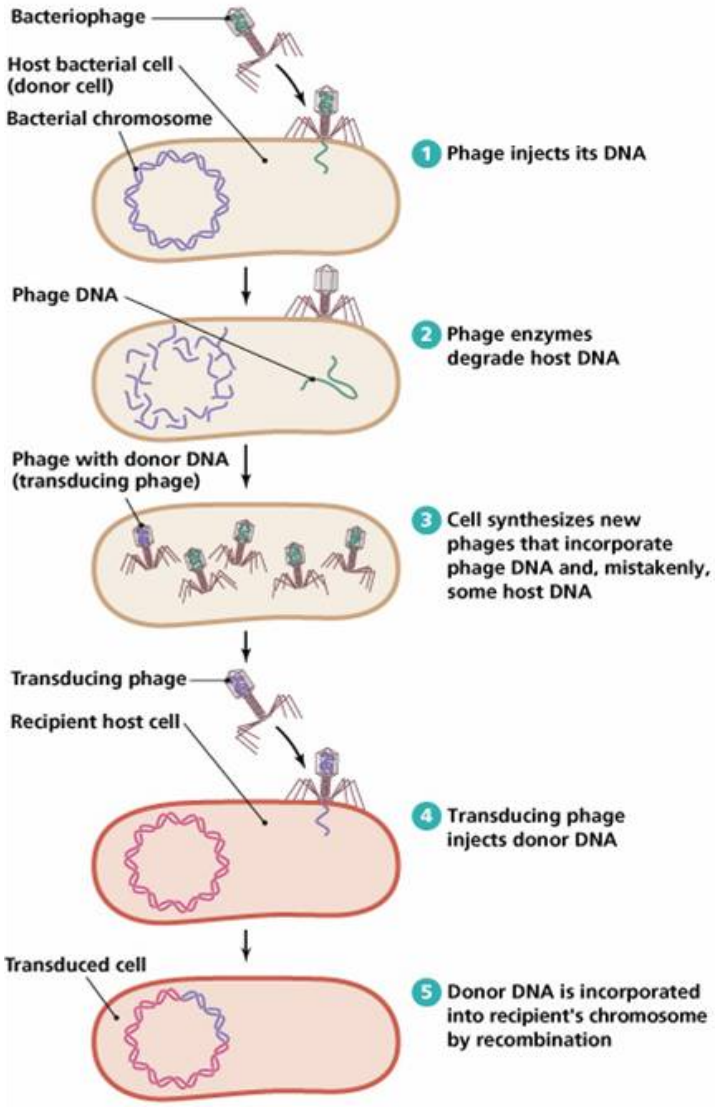

Figure 1.4: **Bacteriophage transduction in bacteria.** (1) A phage attaches to the outside of a host cell and injects its DNA through the membrane. (2) Enzymes are expressed from the phage DNA, degrading the host chromosome into fragments. (3) As new phages are synthesized, they package DNA fragments from their surrounding; in this way, host genes are packaged into the phage. (4) The phages now carrying genes from both the original phage and the host chromosome, inject DNA into a recipient cell. (5) The exogenous chromosomal genes carried by ejected by the phage are introduced into the recipient cell by recombination; progeny will possess this altered chromosome. This figure was produced by Pearson Education Inc.

with a special label which fluoresces when the cell is illuminated by a specific wavelength of light [15]. When the fluorescent label is bound in some way to a particular protein or locus in the cell, it produces distinct fluorescent spots, called 'foci.' In this thesis, I will provide examples how the analyses of the dynamics of foci can elucidate the underlying structure and dynamics in the cell.

1.5.3 Snapshot microscopy

Individual images of a microscope's field of view ('snapshots') can often to be used for efficient studies that do not depend on the time-dependent features of cells. In Chapter 2, I will provide examples of how the wide cell-to-cell variations existent in unsynchronized cells can allow ample information about cell identity to be extracted from only snap-shot images of a mixed cell culture.

1.5.4 Timelapse microscopy

For most studies in this thesis, we required time-dependent information about cell shapes (from brightfield microscopy, Chapters 2 and 4) and chromosomal locus dynamics (from fluorescence microscopy, Chapters 2 and 3). This is achieved through the acquisition of a series of images of the same field of view, taken at regular intervals, referred to here as 'timelapse microscopy'. By splicing together information from each successive frame, we can create movies of cell growth and locus motion, or extract time dependent localizations of chromosomal loci, allowing for the characterization of full cell-cycle DNA dynamics.

1.5.5 SuperSegger segmentation tool

Whether approaching wide cell to cell variations as a valuable statistical tool or a obfuscating challenge, it is important to gather information from a large sample size of cells. For the efficient detection of cells and extraction of quantitative information about their characteristics, we require a reliable, automated high-throughput image analysis tool. *Supersegger* is

an image processing and cell segmentation tool designed by our lab to achieve this.

Supersegger, a MATLAB-based software publicly available from the Wiggins Lab website, processes images through five main steps: (i) alignment, (ii) segmentation, (iii) linking, (iv) creation of cell files and (v) image cytometry [1] (Fig. 1.5). In the first step, images of the same field of view in successive frames are aligned to correct for slide drift. Next, a watershed method is used to identify cell boundaries. A neural net is implemented to train the software learn what a cell 'should' look like and correctly classify regions as cells. In the linking step, *SuperSegger* links information about each cell from frame to frame, creating a history of each individual cell over time. All information about cell lives and characteristics is recorded in cell files ('.mat' files carrying individual cell information) and a summarized list of all quantified cell characteristics called the *'Clist'* (short for *Cell list*). In the fifth step, the *Clist* is used for image cytometry and population analysis using the `gateTool` software, which is a powerful tool that takes advantages of variations in cell populations for the statistical analysis of cell behavior [16]. This process is described at length in Chapter 5 of this thesis.

1.6 Cell-to-cell variations as a 'feature and a bug' in quantitative cell biology

Variations within a living population are essential for their survival. In addition, due to their small size, bacteria dynamics are subject to stochastic fluctuations, where small environmental fluctuations can play a large role in sub-cellular dynamics. These variations in cell shape and sub-cellular dynamics can therefore provide either a wealth of information, or an obfuscating challenge. To use Paul Wiggins' prolific phrase, these cell-to-cell variations are both 'a feature and a bug.'

1.6.1 Variations as a bug

A common challenge in any kind of experiment is noise. Some features of the information collected from a study carry valuable information about the underlying mechanism ('signal'); other features reflect only stochastic fluctuations in a population ('noise'). One of the greatest

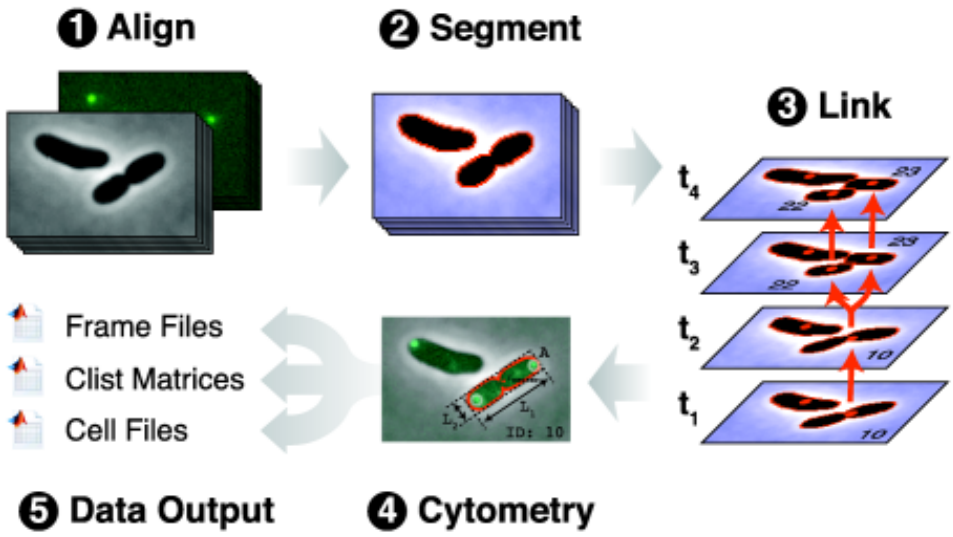

Figure 1.5: ***SuperSegger* work-flow.** The fluorescence and phase images are processed and aligned. During segmentation the cell regions are identified from the background. Then each cell region is linked to a corresponding cell region in the next frame and the cells receive unique ID numbers. Next, the properties and fluorescence characteristics of each cell are calculated. Finally, the program outputs three different types of outputs: Frame files, Clist matrices and Cell files. This figure is from *Stylianidou et. al. 2016* [1].

problems in designing an experiment is determining how to isolate the signal from the noise, to accurately portray the population-wide effects. One important way to achieve this is to to make many measurements (large N). The error on the mean of a measurement is calculated

$$\sigma_N = \frac{\sigma}{\sqrt{N}}. \tag{1.10}$$

To decrease the counting error in our biophysical studies, we require a large number of measurements- this is why optimizing high-throughput techniques has been an important piece of my graduate studies. The *SuperSegger* and `gateTool` software have therefore played an important role in making these quantitative biology studies possible.

While taking a large number of measurements can be an exciting challenge on its own, we also require a way to compare information from widely-varying cells. In the second chapter of this thesis, I will explore this problem by considering a study of the chromosomal organization in *E. coli* cells. Understanding the organization of the *E. coli* chromosome is an important stepping stone for modeling how structure is managed by the cell; we hope that understanding the dynamics of chromosomal motion will provide us with hints as to what must be moving it. Furthermore, the mechanism guiding the segregation of the two well-organized chromosomes in a pre-division parent into two daughter cells remains completely uncharacterized. By studying the organization and dynamics of the *E. coli* chromosome, we investigate an important question about the structure of bacterial DNA. Since chromosome dynamics vary widely from cell to cell, this required us to explore techniques for comparing bacterial chromosome dynamics across many widely-varying cells.

1.6.2 Variations as a feature

Although wide variations in the cellular shape and dynamics can act as a major challenge in probing bacterial structure, variations across a population also provide us with a large bank of information, which we can harness to improve our understanding of the population [17].

We learned in a Section 1.4 that measuring the responses of cells with differing phenotypes in different growth conditions is an important technique used by biologists. By studying cells

with slight phenotypical differences within the same population, we can quickly and easily achieve this technique in a single experiment.

We will explore this perspective on cell-to-cell variations in the second and fourth chapters of this thesis. First, in chapter two, we discuss the `gateTool` cytometry tool built for *SuperSegger*, which facilitates the use of high-throughput analysis and gating techniques for exactly these types of experiments. We then give examples of how the tool can be used to partition datasets into smaller sub-sets for analysis or how variations in a single population can be used to address prominent biological questions. Then, in chapter four, we introduce a specific study of bacterial structure focused on mutant cell and colony morphologies. The diversity of morphologies of bacteria is a fascinating aspect of cellular structure. But despite the myriad morphologies exhibited even by different cells of the same bacteria, the mechanisms driving morphogenesis and the evolution of bacterial morphology are not well understood. By studying the effects of gene deletions in the pathway of cell wall synthesis, we hope to see how different proteins contribute to varying cell morphologies.

Chapter 2

HIGH-THROUGHPUT AUTOMATED ANALYSIS METHODS

Advances in automated fluorescence microscopy have made snap-shot and time-lapse imaging of bacterial cells commonplace, yet fundamental challenges remain in analysis. The vast quantity of data collected in high-throughput experiments requires a fast and reliable automated method to analyze fluorescence intensity and localization, cell morphology and proliferation as well as other descriptors. Inspired by effective yet tractable methods of population-level analysis using flow cytometry, we have developed a framework and tools for facilitating analogous analyses in image cytometry. These tools can both visualize and gate (generate sub-populations) more than 70 cell descriptors, including cell size, age, fluorescence, *etc.* The method is well suited to multi-well imaging, analysis of bacterial cultures with high cell density (thousands of cells per frame), and complete cell cycle imaging. We give a brief description of the analysis of four distinct applications to emphasize the broad applicability of the tool.

2.1 Cell Sorting and Segmentation

Quantitative fluorescence imaging has become an important tool for the study of the bacterial cell biology [18]. Most cellular processes, from chromosome segregation to replication, are stochastic in nature and exhibit significant cell-to-cell variation [19, 20, 21]. This variation makes the quantitative statistical analysis of a significant number of cells essential [20]. For many biological problems, flow cytometry offers a tractable and rapid approach for the analysis of thousands of cells and offers powerful methods for sub-population analysis: Cells are *gated* (or selected) by fluorescence or scattering intensity, or a combination of intensities. Gating is used for both analytic applications (*i.e.* the determination of the

relative abundance of a sub-populations or the removal of cell debris from the analysis) and functional applications (*i.e.* Fluorescence Activated Cell Sorting, FACS).

Although flow cytometry is fast and efficient, many important cell biology questions demand an imaging approach where cellular ultrastructure can be characterized and the cell cycle dynamics captured for individual cells. In contrast to flow cytometry, the use of time-lapse imaging has the potential for complete cell cycle analysis and characterization of cells. While it is tractable to capture time-lapse images of tens-to hundreds of-thousands of cells with modern automated fluorescent microscopes, significant challenges remain in the analysis of these data sets.

Cell segmentation and analysis packages have been developed for microscopy (*e.g.* [22, 23]) and include some automated tools for analysis of these large data sets, but they are not as powerful and flexible as the tools commonly used in the analysis of flow cytometry data. For instance, although some existing packages can generate histograms of cell descriptors from segmented data, it is often necessary to define and analyze subpopulations of cells (*e.g.* removal of cell debris or non-proliferating cells, *etc*). In Chapter 2 we describe a powerful tool for sub-population analysis inspired by flow cytometry.

Our *Clist* (or *Cell list*) framework, and tool for data gating and visualization `gateTool`, facilitate efficient capture and manipulation of image cytometry data, including the generation of sub-populations from over 70 pre-programmed cell descriptors and provide a flexible framework for analyzing many cell biology problems. These descriptors include *birth time*, *division time*, *old pole age*, *long axis length*, *mean fluorescence intensity* and *cell area*, as well as the ability to add additional customized descriptors as required. We have recently described a cell segmentation and analysis package for bacterial cells (*SuperSegger*) [1]; *SuperSegger* and the `gateTool` are designed to be part of the same complete package, but can be used independently. That is, *SuperSegger* will automatically output segmented cell data as a *Clist* for seamless input to the `gateTool` for analysis, but a custom user-constructed *Clist* can also be used. In principle, the `gateTool` framework could be applied more broadly, to classify objects and facilitate analysis in a wide range of image analysis applications.

However, the *SuperSegger* software is designed specifically for the segmentation of bacteria cells. We will discuss the `gateTool` in the context of bacterial cell analysis.

We have already used this method, without detailed description, in a number of papers [24, 25, 19, 26, 4, 27, 20, 21, 1], and the software is available for download from the Wiggins Lab website (`http://mtshasta.phys.washington.edu/website/ssodownload.php`). The purpose of the current report is to describe the method and to demonstrate its potential. Here, we first give a brief description of the tools used for sub-population analysis, then we analyze a number of representative cell biology problems. In particular, we investigate a number of common assumptions (*e.g.* cell length is a good proxy for cell age) and interesting recent claims in the literature (*e.g.* aging in *Escherichia coli*), using the *Clist* tools to explore the robustness of these observed phenomena.

2.2 Results and Discussion

A matrix-based summary of time-courses. Our segmentation suite *SuperSegger* provides three partially redundant outputs: (i) *frames files* which contain all the data from a single time-point, (ii) *cell files* which contain all the data for a single cell for all time-points and (iii) the *Clist* (or *Cell list* matrix) which is a matrix-structured summary of all cells and all time-points [1]. This study focuses on analysis of the *Clist* matrix.

Due to the size of the typical processed data set, it is usually not practical to load the entire data set into memory. The purpose of the *Clist* matrix is to load only the data pertinent for population-level analyses. The schematic form of the matrix is shown in Table 2.1. Each row represents an individual cell tracked through the time-course and the columns represent a subset of the > 70 cell descriptors.

SuperSegger assigns each cell a unique cell ID which is preserved throughout the time course and is used as an identifier for subsequent analysis. The cell ID can been seen in the first column of the *Clist* matrix (Table 2.1). The ID of mother and daughter cells is also recorded (see columns 2-4 of the *Clist* matrix in Table 2.1).

Cell ID	*Mother ID*	*Daughter 1 ID*	*Daughter 2 ID*	*Birth Time*	*Division Time*	*Age (frames)*	*Old Pole Age*	*Long Axis*	*Short Axis*	*GFP Mean*	...
1	0	2	3	1	124	123	NaN	1.4	0.76	148.57	...
2	1	4	5	125	163	38	NaN	2.6	0.81	143.15	...
3	1	6	7	125	166	41	NaN	2.6	0.84	147.75	...
4	2	8	9	164	202	38	NaN	2.5	0.82	150.63	...
5	2	12	13	164	210	46	2	2.2	0.77	141.07	...
6	3	14	15	167	214	47	2	2.2	0.82	139.83	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Table 2.1: ***Clist* data structure.** *A schematic* picture of the *Clist* matrix. The matrix columns represent cellular descriptors (one value per cell) and the rows correspond to individual cells. At typical *Clist* generated from a single field of view can contain 5,000 cells, each with >70 cellular descriptors.

Cell cytometry and the *Clist*. *SuperSegger* performs a detailed characterization of the cell at each time-point. A summary of this information is stored in the *Clist* matrix. It is important to note that each element of the matrix is a single number (not time-dependent). Although this makes sense for some descriptors (Cell ID, duration of the cell cycle, *etc*) other cell descriptors change from frame-to-frame (*e.g.* cell length and fluorescence intensity). For most frame dependent descriptors, the value at cell birth (or in some cases cell death) is retained. For example, the length of a cell (Long axis) which grows throughout the experiment is stored as two *Clist* variables: `Long axis at birth` and `Long axis at death`. The *Clist* construction is easily customizable; addition to or removal of any cell characteristic from the *Clist* requires editing only a single line of the *SuperSegger* software. The name of each descriptor is also saved in the *Clist*. (See Section A.6 of the Appendix for details on the included cell descriptors, and how to add your own cell descriptors.)

Cell birth and death. Although the *Clist* is most powerful when used in the context of time-lapse complete cell cycle imaging, the matrix can still be computed for snapshot (or short time-lapse) analysis. If there is only a single frame, birth and death descriptors are identical. It is also important to note that the descriptions birth and death refer only to the appearance (and disappearance) of cells from the analysis. All cells present in the first frame are assigned frame one as their birth time, but new cells also appear due to cell division and cell tracking errors. The last frame in which the cell is observed is defined as its death time. Cell death could be the result of cell division (since new cells are assigned new IDs), the cell drifting from the field-of-view, a cell tracking error, the experiment ending, *etc*. Note that there are many equivalent strategies for ensuring that only new born cells that result from cell division (not tracking errors, *etc*) contribute to an analysis, as will be described below.

Time-dependent *Clist*. For some applications, the retention of only birth and death cell descriptors is too limiting. In addition to the standard *Clist*, we provide a second time-dependent *Clist*, called *Clist3D*, which contains time-dependent descriptors for each frame, each cell and a subset of the cell descriptors. The matrix contains Not-a-Number (NaNs)

in frames where the cell does not exist. In general, the direct manipulation of the *Clist3D* matrix is cumbersome although it is useful for some applications. One example is given in the Length as a Proxy For Cell Age section of this chapter.

Visualization and analysis of data Once the *Clist* is constructed, we have efficient methods for visualizing and analyzing population-level characteristics. We streamline the data analysis process using a function called `gateTool`. By passing keywords into this single function, the user can extract data, and implement a number of population visualization tools, including one and two dimensional histograms, kernel density estimates (KDEs; see Appendix A.9 for details) (Fig. 2.2, Fig. 2.3 and Fig. 2.4), and dot plots (Fig. 2.1A), *etc.* An easy to use GUI version of the `gateTool` is also available (See Appendix A.7).

To complement the population-level visualization tools, we also include tools for the analysis of fluorescence localization in single cells. *Single Cell Towers,* or mosaics, are time-lapse images of a single cell, stacked in chronological order, with consistent pole orientation (Fig. 2.3). In order to visualize fluorescence localization averaged over cell-to-cell variation, we have also implemented *Consensus Images*, the mean cell cycle dependent fluorescence localization pattern from sets of single cell towers, spatially scaled to the same cell length and width (See: Fig. 2.3). A detailed description for the implementation of *Single Cell Towers* and *Consensus Images* can be found in [20].

In addition, the `gateTool` will also compute basic statistics (mean, std, error, min and max, population size, *etc*) on any cellular descriptor and the raw data can be exported as matrices into MATLAB, as well as saved in csv and excel file formats.

Sub-population selection by gating. Tools for cell selection by image cytometry [28, 22] allow users to classify cells, however we implement a more flexible system inspired by flow cytometry to provide a robust tool for cell selection. In flow cytometry, a canonical method for the identification of sub-populations of cells is the definition of a *gate*: a subset of cell descriptors corresponding to a population of interest. Usually these gates are defined by selecting an interval of a single descriptor which is often visualized on the descriptor

histogram. For gating in two-dimensions, we implemented a polygon-based definition of the gated region in the two-dimensional descriptor space, shown for the long and short axes in Panels A and B of Fig. 2.1.

Although a single descriptor gate (selection) is straightforward to implement on a single population in MATLAB or R without the use of `gateTool`, it is often necessary to gate in two descriptor dimensions, or perform multiple sequential selections on two different populations. These more complicated but routine manipulations are easily performed in `gateTool`. For instance, the interactive tool for selecting gates is particularly powerful in two dimensional where custom shapes are often the most convenient method for defining a population but specifying the vertex coordinates numerically would be time consuming. *E.g.* see Fig. 2.1B. Another key feature of `gateTool` is the ability to nest an arbitrary number of gates. Importantly, data that are gated out are retained in the *Clist*, but ignored in subsequent analysis. After gating, we can choose to restore the complete un-gated dataset by simply empty the gate. This facilitates the ability to divide data from a single *Clist* into sub-populations in many different ways. This implementation also makes it straightforward to copy gates between *Clists*; this is useful when we wish apply consistent constraints to two separate *Clists* (i.e. Fig. 2.1)A and 2.1B.

Finally, the analysis can be applied simultaneously to any number of *Clists*, which can be treated as replicate experiments where cells are pooled, or from distinct conditions where the data are analyzed independently. For instance, we discuss an example below where the analysis is performed simultaneously on *Clists* generated for 48 strains.

Comparison to existing packages. Although other packages allow for image analysis, and some offer tools for sub-population selection, our `gateTool` method offers a more extensive and flexible framework for cell analysis and sorting. For example, Oufti [23] stores data as a "cell list" containing 18 cell descriptors for all cells in all frames, and the ability to "filter" cells by a single cell descriptor at a time, from a list of five cell descriptors. In comparison, the *SuperSegger*-generated *Clist* contains 70 cell descriptors and frame-independent structure

Descriptor	Mutual Information M (bits)
Fluorescence	0.99
Long and Short	0.9
Eccentricity	0.8
Long	0.5
Short	0.5
Area	0.4

Table 2.2: **Mutual Information by Cell Descriptor.** We computed the mutual information from morphological cell descriptors and mean cellular fluorescence. Fluorescence is the most informative descriptor for sorting bacterial species; however the combination of short and long axis information trails by less than 0.1 bit, serving as a suitable pair of cell descriptors for sorting bacterial species.

offers users a more extensive quantitation of user data, in a more compact structure. In addition, `gateTool`'s capabilities for two-dimensional gating from the set of 70 cell descriptors greatly enhances the ability to perform novel analyses. Another package, MicrobeJ [22], allows users to filter cells by using a cutoff value; although this does allow users to gate out cells, the use of a single cutoff value limits the flexibility in sub-population selection. Histograms can be generated, but there are no two-dimensional visualization (or gating) options. Neither Oufti nor MicrobeJ offers users the ability to add their own customized variables to the set of cell descriptors which can be analyzed. Our *Clist* and `gateTool` offer users a novel image analysis tool for highly customized analysis with an easy-to-use framework.

Removal of debris from analysis. In practice, it is difficult to prepare samples completely free from microscopic debris that can be mistakenly identified as cells during the cell segmentation process. Although *SuperSegger* generates few false positive events for the analysis of wild-type *E. coli* (False positive rate $\alpha < 0.01$ [1]), some mutants, like *minC*, can

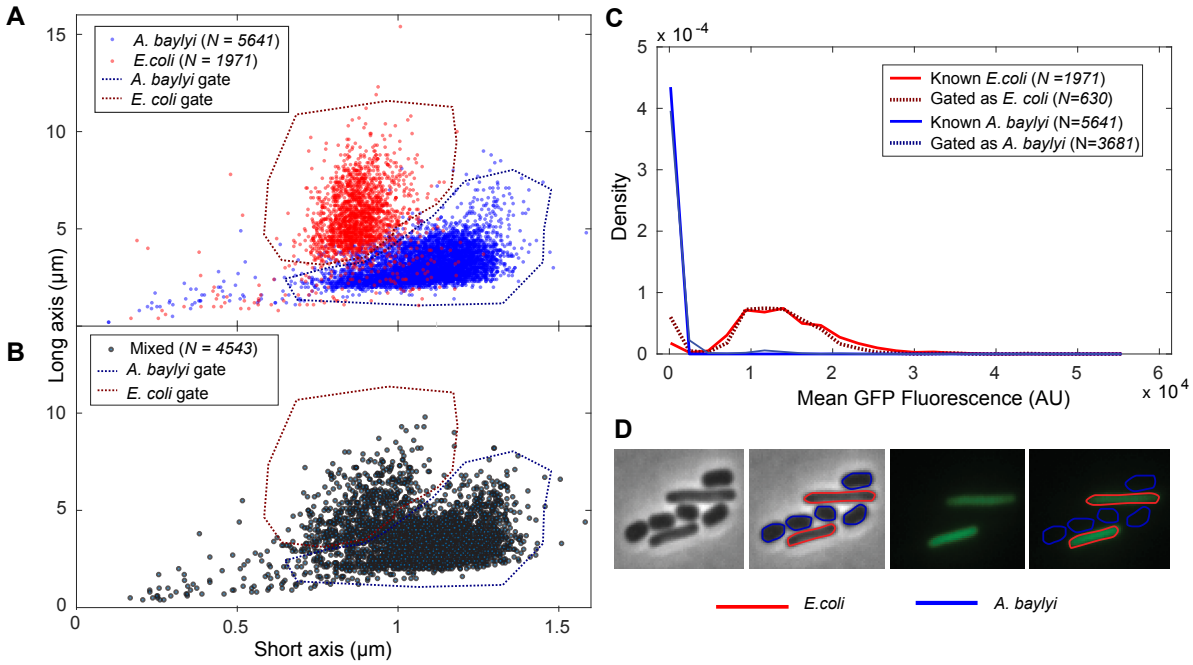

Figure 2.1: **Identification by morphology verses fluorescence. Panel A:** A dot plot of long and short axes for the *A. baylyi* (blue) and *E. coli* (red) cells, imaged separately. Based on this data, appropriate gates for *A. baylyi* and *E. coli* based on morphology are selected (dotted). **Panel B:** A dot plot of long and short axes for the mixed *A. baylyi* and *E. coli* cells, with the *A. baylyi* and *E. coli* gates from Panel A applied. **Panel C:** Histogram of mean GFP fluorescence for known *A. baylyi* (solid blue) and *E. coli* (solid red) cells, as well as mixed cells sorted as *A. baylyi* (dotted blue) and *E. coli* (dotted red). We find that the fluorescence data for the mixed cells sorted by morphology closely match the fluorescence data for the known populations. **Panel D:** Bright field and fluorescence images of a sample field of view, with and without cell outlines determined by morphology and fluorescence gates.

significantly increase the amount of cell-like debris (e.g. [29]). Gating can therefore provide an effective method for removing debris from the analysis.

Snapshot analysis.

In image-based cytometry of snap-shot images, we have used the following descriptors to remove debris: (i) major axis length, (ii) minor axis length, (iii) area, (iv) phase gray level, and (v) region score. The first three identifiers need little explanation and usually offer sufficient discrimination without additional descriptors. Phase gray level can be quite informative since isolated cells are phase dark. The region score is a proxy for the likelihood of a region being a cell, generated automatically by *SuperSegger* using a machine learning algorithm trained on test data sets. It combines information from many descriptors not included in the *Clist* and can be an effective tool, provided that the cells of interest are similar in morphology to the training set.

For time-lapse analysis. For the removal of debris from time-lapse analysis over complete cell cycles, we usually use of the `stat0` descriptor alone. `stat0` is the cell tracking status defined as follows:

`stat0`	Birth Div.	Death Div.	No Errors
0	✗		
1	✓	✗	
1	✓	✓	✗
2	✓	✓	✓

For many analyses, restricting the analysis to cells for which a complete cell cycle is observed (`stat0`= 2) makes additional gating unnecessary. ion time-lapse imaging of cells that do not complete a full cell cycle, we have used a number of other methods for removing debris and cells with segmentation errors in addition to the descriptors described for snap-shot analysis. The descriptor `error frame` contains the frame number of the first error (or a NaN if there is no error). Descriptors are also included for the maximum (`dL max`) and minimum (`dL min`) change in the cell length (major axis length) between frames over the course of the

time-lapse experiment. These descriptors enable the detection of segmentation errors by large changes in cell length. A useful descriptor for intermediate duration time-lapse is the estimated rate of growth:

$$k \equiv (\log L_{t+\tau} - \log L_t)/\tau, \tag{2.1}$$

where L_t is the cell length at time t and τ is the interval between frames.

Distinguishing cells in a mixed population. Although bacteria are single-celled organisms, they interact and exhibit many forms of cooperative behavior [30]. To study these phenomena, it is necessary to distinguish between the bacterial species. We have recently made a quantitative characterization of T6SS-based contact-mediated cell killing in *Pseudomonas aeruginosa* [25]. Like many previous studies, we introduced an exogenous fluorescent label to differentiate between cells and used the observed cellular fluorescence (descriptor `fluor mean`) to differentiate between T6SS$^{\pm}$ cells [25]. In our experience this is a good, although not perfect, descriptor for distinguishing cells.

To measure the quality of fluorescence as a proxy for cell identity, we cultured wild-type MG1655 *E. coli* and the same strain expressing GFP (expressed from plasmid pZA12-GFP), then considered the mutual information of the cell fluorescence and various morphological cell descriptors. The mutual information is a measure of how much two distributions depend on one another; given the set of values for one variable, how well can you predict the set of values for the second variable? For a mathematical explanation of the mutual information, please see Appendix A.11. In the context of inference on the cell identify given an equal number of GFP± cells, the average information content of the fluorescence intensity (*i.e.* the mutual information) is:

$$M(\text{GFP intensity}, \text{strain}) = 0.99 \text{ bits}, \tag{2.2}$$

compared to one bit, which is the information required to unambiguously identify the cell. This corresponds to an average information loss of just 0.01 per cell relative to perfect information about cellular identity and this loss results from the small overlap in the fluorescence intensities of GFP± cells.

Distinguishing cells by morphology.

Although the use of cytoplasmic fluorescence to differentiate between strains is a canonical method in cell biology, it is not always convenient. We decided to explore the possibility of differentiating between bacterial cells based on cell morphology alone, rather than using a fluorescence-based descriptor. To determine whether this approach would be feasible, we considered a co-culture of *Acinetobacter baylyi* and *Escherichia coli*. When grown in liquid minimal media, *E. coli* cells have a long, thin rod-like morphology (*bacillus*), while *A. baylyi* are quite round in comparison (*coccobacillus*). We would therefore expect the eccentricity ε, defined as the ratio of the major to minor axis length, to be a good proxy for cell identity. In the context of inference on the cell identify given an equal number of *A. baylyi* and *E. coli* cells, the average information content of the eccentricity (*i.e.* the mutual information) is:

$$M(\varepsilon, \text{species}) = 0.8 \text{ bits}, \tag{2.3}$$

compared to one bit, the information required to unambiguously identify the cell. This information loss corresponds to an average information loss of 0.2 bit per cell relative to perfect information about cellular identity. Again, this loss results from the small overlap in the eccentricity of *A. baylyi* and *E. coli* cells. For these cells with ambiguous eccentricity, the cell identify cannot be reliably determined.

The eccentricity is a convenient one dimensional measurement which combines information about major and minor axis length; however, inference on cell identify using both the major and minor axes length descriptors is more informative. The mutual information between morphology (using both descriptors) and species is:

$$M(\text{morphology}, \text{species}) = 0.9 \text{ bits}, \tag{2.4}$$

roughly a 10% increase in information content over the eccentricity. Although the information loss for cell morphology is greater than it is for fluorescence, the morphology may be sufficient. Remember that for many applications, ambiguous cells can simply be removed by a judicious choice of the gates. A table of the estimated mutual information for morphological and

fluorescence cell descriptors is provided in Table 2.2. An example of morphology based gating is shown in Panel B of Fig. 2.1.

Gating on fluorescence versus morphology. To demonstrate the process of gating with `gateTool`, we create *A. baylyi* and *E. coli* sub-populations using either the fluorescence or the morphology descriptor. We cultured (non-fluorescent) *A. baylyi* and *E. coli* (expressing GFP) both (i) separately and (ii) mixed together, then collected snap-shot images of all three cultures. Panel A of Fig. 2.1, shows the dot plot of the long and short axis lengths of the individually grown *E. coli* and *A. baylyi* cells. Using this information about the cells of known species, we created gates that can be used to sort the two species by morphologies.

In Panel B, those gates are applied to the dot plot of the mixed cells' lengths, to predict which cells are *A. baylyi* (blue gate) and which are *E. coli* (red gate). In Panel C, we provide the histogram of fluorescence for the known *A. baylyi* and *E. coli* cells, as well as the "sorted" mixed cells. We find that the predicted fluorescence distributions for both species closely match their known fluorescence distributions.

Panel D of Fig. 2.1 displays the phase and fluorescence images of a sample field of view containing *A. baylyi* and *E. coli* cells. Each of these images is accompanied by the same image, with cell outlines of blue (*A. baylyi*) and red (*E. coli*), based on the chosen cell sorting gates. Based on our training data and gates, we estimate the *A. baylyi* misclassification rate as 1.3% and the unclassified rate as 2.9%. The *E. coli* misclassification rate is 6.0% and unclassified rate is 5.3%. "Unclassifed" cells are those cells which were not sorted by any gate. In general, a lower misclassification rate can be achieved by setting stricter gates, at the cost of a higher unclassified rate.

Length as a proxy for cell age. In flow cytometry and snap-shot experiments (where cell age information is unavailable) relative cell lengths are commonly used as a proxy for cell age (*e.g.* [31, 32]). The natural cell-to-cell variation in cell length implies that it is an imperfect proxy for cell age. To place cell-length dependent data in context, it is interesting to determine how informative this data is with respect to true time-lapse imaging for deter-

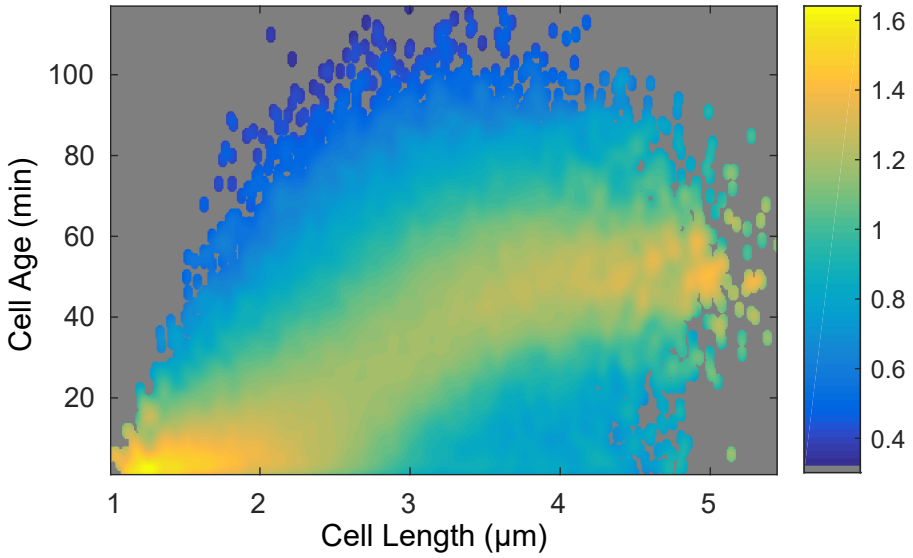

Figure 2.2: **Conditional probability of cell age given a cell length.** This bivariate histogram displays the cell length and age from over 56,000 measurements. The coloring of the points shows the resulting conditional probability of the cell age, given the cell's length.

Variable 1	Variable 2	Mutual information M (bits)	Effective resolution: $\mathcal{N}$
t	L	0.8	1.8
τ	L	0.5	1.5
τ	λ	2.8	7.1

Table 2.3: **Estimated Mutual information of cell length and age.** We estimated the mutual information of three combination of four variables: t is the age of the cell in minutes, τ is age of the cell relative to the length of the cell cycle, L is the length of the cell is microns, and $\lambda \equiv \log(L/L_0)/\log(L/L_1)$ is the log length, re-scaled by initial and final length. The effective resolution is effective number of frames per cell cycle: $\mathcal{N} \equiv 2^M$.

mining the timing of cellular events. Again, it is useful to formulate this problem in terms of mutual information, or the information content of cell length with respect to cell age. In this case, we will compute the number of codewords, *i.e.* $\mathcal{N} \approx 2^M$, which is comparable to effective temporal resolution, in frames per cell cycle achieved by snapshot imaging.

To analyze the information content of cell length, we characterized wild-type MG1655 *E. coli* proliferation over two hours with one minute time resolution. A kernel density estimate showing cell age, conditioned on cell length, is shown in Fig. 2.2. It is immediately clear from the plot that due to the significant variation in cell length, length will be an uninformative predictor of cell age.

We computed the mutual information for three pairs of descriptors and the results are shown in Table 2.3 (the additional conditional probability plots for these descriptor pairs are provided in the Appendix Fig. A.13). The most relevant mutual information is between cell length (L) which is observed and relative age τ, defined as the cell age relative to the cell cycle duration. This mutual information is the smallest, a fraction of a bit per cell. The effective temporal resolution is only ≈ 1.5 frames per cell cycle.

To demonstrate that the variation in both cell length and cell cycle duration are responsible for the low mutual information, we also computed mutual information for relative length

and relative age, which is almost 3 bits and corresponds to an effective frame rate of 7 frames per cell cycle. (The relative length cannot be determined in snapshot imaging.)

There are two important implications of this experiment: (i) cell-length based analysis of cell cycle dynamics is a poor substitute for true time-lapse imaging since the significant variation in both the duration of the cell cycle and the length of cells makes inference of cell age from cell length uninformative. (ii) Although the typical information content of this data is quite low, the sub-population of the shortest and the longest cells are still somewhat informative implying that cell-length based analysis, while inefficient, is sufficient for some problems.

Over-expression phenotypes. To demonstrate the analysis potential of the *Clist* and `gateTool` in the context of a high-throughput time-lapse analysis, we analyzed data from a near-complete library of fluorescent fusions. We had previously imaged the ASKA library [33], a genome-scale, plasmid-based collection of inducible fluorescence fusions to genes of *E. coli* [20] to characterize protein localization dynamics in a model bacterium. An interesting potential feature of the ASKA library is the ability to characterize the effect of protein expression on cell morphology, where protein expression is quantified via fluorescence. We harness the natural variation in gene expression to generate a range of protein expression levels [34, 35]. A natural test case are the genes *ftsAZ* whose gene products assemble to form the z-ring, responsible for septation, but whose over expression results in a loss-of-function phenotype: filamentation [36, 37].

For previous studies, our lab imaged all ASKA fusions displaying nondiffuse localization [20, 27]. We re-analyzed the data from these high-throughput imaging experiments, which included both the FtsZ and FtsA fusions. This is accomplished with a single command using the `gateTool`. For compactness, we show data only from the 48 strains imaged with the FtsZ and FstA strains, rather than from the roughly 800 strains imaged.

We computed the statistics on the `cell length at death` and ordered the 48 strains by average length. (See Fig. 2.3A.) The *ftsAZ* strains (bolded in Fig. 2.3A) were longest

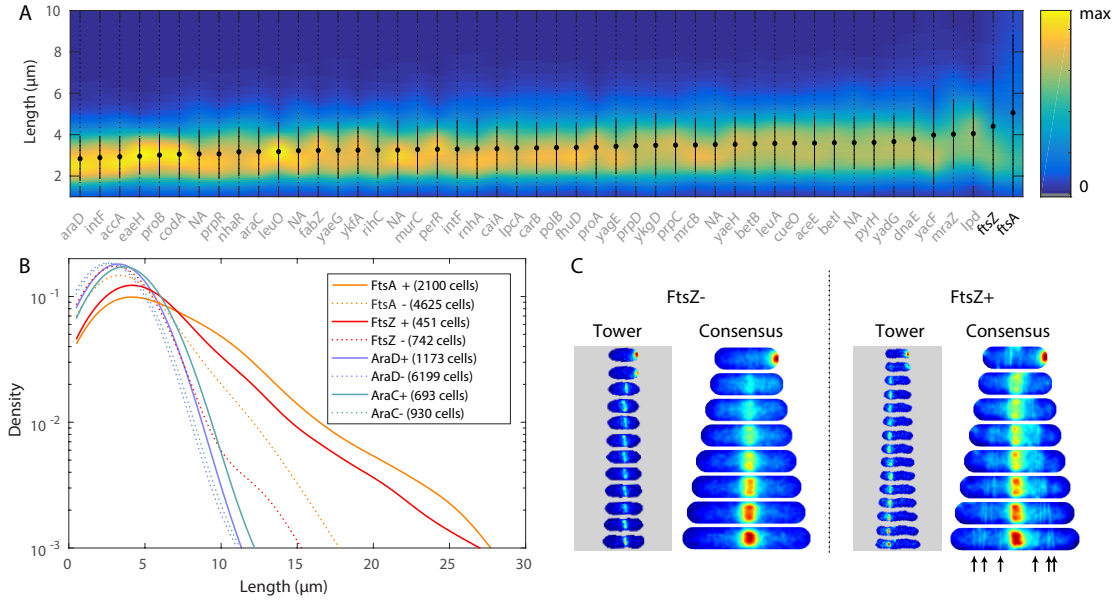

Figure 2.3: **Over-expression phenotypes. Panel A:** KDE of cell length (at death) for 48 strains (including *ftsZA* fusion strains) and statistics: minimum to maximum (dotted), mean (circle) and standard deviation (solid line). Genes labeled as NA are currently not annotated genes in MG1655. Strains are ordered by mean length. **Panel B:** Analysis of high (+) and low (−) expression sub-populations of four strains. High-expression sub-populations of FtsA and FtsZ have long length. **Panel C:** Single cell tower and consensus images for FtsZ high and low expression sub-populations. The consensus image for the + sub-population shows a cloud around midcell, reflecting the averaging of tower images of z-rings with aberrant localization. Arrows have been included to point out the various clouds of higher intensity, localized away from cell center, resulting from misplaced z-rings in many cells.

of the 48 strains. We then constructed two sub-populations: protein expression level above (+) and below (−) the mean expression level, as measured by mean fluorescence intensity. The distributions of cell lengths are shown for *ftsZ*, *ftsA*, *araC*, and *araD* fusion strains. (The strains *araCD* are included as a control.) Although all strains had longer length in the (+) populations, the *ftsAZ* strains show a pronounced increase in the number of long cells relative to the (−) population, exactly as predicted. (See Fig. 2.3B.) The mechanism for cell filamentation is clear from the consensus images: The (−) population shows a precisely localized z-ring forming at midcell, whereas the (+) population shows a cloud of intensity

from z-rings with aberrant localization. (See Fig. 2.3C.)

Determinants of growth rate. F. Taddei and coworkers have recently proposed a mechanism for aging in cells which divide symmetrically[38], based on the definition of a pole age. Pole age is defined as the number of divisions since the pole was created by septation [38]. For instance, a single progenitor cell is shown in Panel A, Fig. 2.4 with poles of unknown age. After two rounds of division, there are four progeny, including two *new* daughters (old pole age 2) and two old daughters (old pole age greater than 2). Taddei and coworkers tracked cell proliferation on an agarose pad by microscopy and, intriguingly, reported that old cells grow more slowly than new cells, consistent with existence of an aging mechanism in *E. coli*. However, in subsequent work using a microfluidic chamber that retained the cells with the oldest poles (mother machine), S. Jun, Taddei and coworkers reported a constant growth rate of cells as a function of pole age [39]. Although the authors were careful to note that these new results were not explicitly in conflict with the previous report, they suggest that *E. coli* growth rate and the aging phenomenology may be more subtle than initially proposed [40].

It is important to note that the reported reduction in growth rate is very small (approximately 7%) and therefore the analysis of the aging phenomena necessitates the characterization of a significant number of cells and is therefore a natural application of the `gateTool`. Due to the small size of the reported change in the growth rate, many subtle mechanisms could contribute. We reconstituted the proliferation experiment using wild-type MG1655 *E. coli*, without the exogenous YFP protein which is highly transcribed from a phage promoter and therefore has the potential to generate replication conflicts [41], and without the potential of DNA damage due to phototoxicity from fluorescence excitation [42]. Both changes were possible since our cell segmentation (*SuperSegger*) is generated from phase contrast images rather than from fluorescence images so YFP and fluorescence imaging are not required for our experiment and analysis.

To characterize the cell proliferation, we inoculated single log-phase cells from liquid

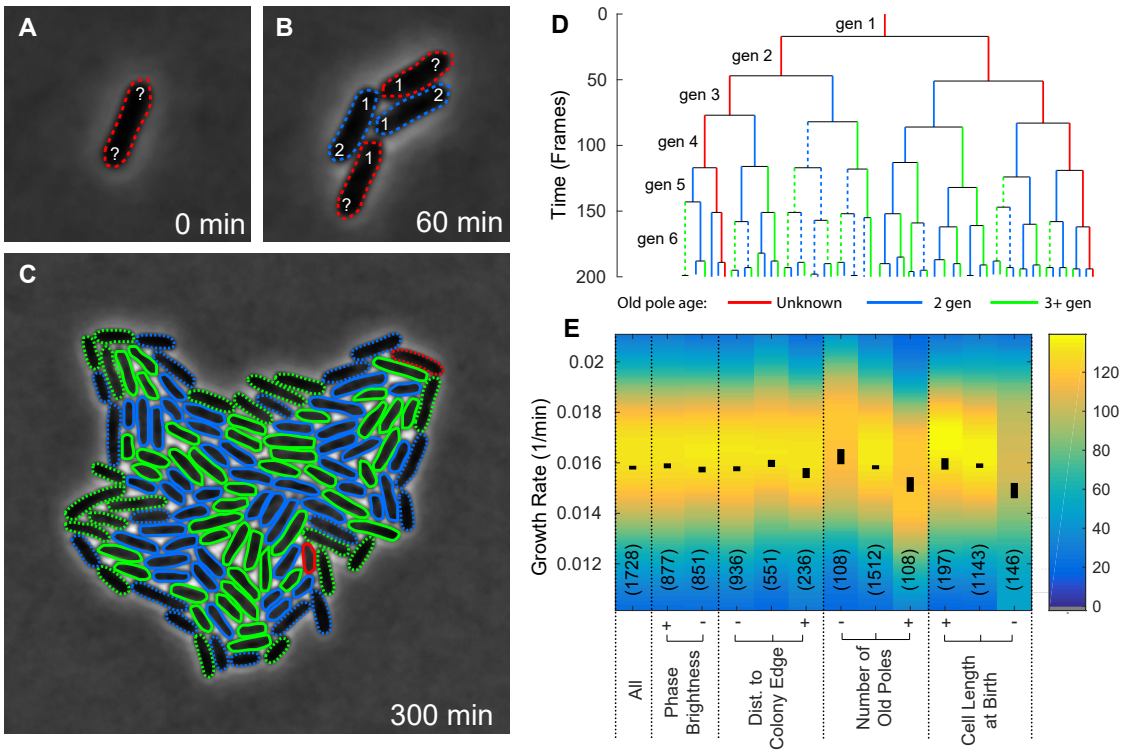

Figure 2.4: **Effect of cell growth characteristics on cell age. Panel A-C:** A sample field of view of a colony growing from a single progenitor cell, shown at t = 0, 60 and 300 minutes. The color of each cell outline indicates the old pole age: two (blue), greater than two (green), or inherited from progenitor/unknown (red). The lines are dashed (or solid) to represent cells at the edge (or inside) of the colony at birth. **Panel D:** A sample lineage tree from a single progenitor cell. The length of each line gives the age of the cell. The color and style of the lines again indicate the old pole age and position of the cell within the colony. **Panel E:** We compare the mean cell age for six different conditions of cell growth: (i) on the edge of a colony, (ii) on the inside of a colony (iii) inheriting the oldest poles (iv) inheriting the most new poles (v) longest at birth, and (vi) shortest at birth. Each cell descriptor is gated into three groups: low (−), medium, and high (+). It is clear from this figure that the mean cell cycle duration (or cell age) and standard error on the mean are consistent, independent of pole age.

media onto agarose pads and imaged individual cells with one minute time resolution for 300 minutes. The images were then segmented and a small number of segmentation errors were corrected by hand such that we had 50 cells tracked through six complete generations without errors [1]. Images from a typical single-cell proliferation experiment are show in Panels A-C in Fig. 2.4.

We investigated a number of hypothetical determinants of cell growth rate that could be measured by microscopy: cell length at birth, position in the colony, pole age and phase brightness. (Phase brightness is included as a control and it not expected to be a determinant of cell growth rate.) Since growth on agarose pads is not rigorously steady-state log growth, we analyzed only cells in the sixth generation. To analyze the role of cell length at birth we gated the cells into two sub-populations short and long, corresponding to cells shorter or longer at birth than the mean cell length at birth. To analyze the role of colony position at birth, we gated the cells into two sub-populations outside and inside, corresponding to cells that contact the boundary of the colony versus cell that do not contact the boundary respectively. (Inside (outside) cells are shown with solid (dashed) outlines in Panels A-C in Fig. 2.4.) Finally we considered old and new cells, as defined above. Old (green), new (blue) and cell with the progenitor pole (red) are shown in Panel A-C in Fig. 2.4. A lineage tree of the cells shown in Fig. 2.4A-C is provided in Panel D. In this tree, the length of each vertical line on the lineage tree is the length of a cell's complete cell cycle; once the cell divides, a vertical line is drawn for each of its two daughters. Again, the color and style of the vertical line is chosen according to old pole age and position within the colony.

The KDEs, the error on the mean and standard deviation for the growth rates of the gated sub-populations are shown in Panel E of Fig. 2.4. (The p-values for the $(+)$ and $(-)$ sub-populations of each cell descriptor are provided in Table A.8 in the Appendix.) This sub-population analysis revealed that cell length at birth and number of old poles are notably the best predictors of cell cycle duration of the four determinants we investigated, with the difference in colony position and phase brightness being statistically insignificant. In this context, the "oldest" cells are those cells with the most old poles (five or six generations

of old poles) in their lineage; the middle and smallest group contain cells with two to four (middle) and zero or one (-) old poles in their lineage. Note that although both cell length and number of old poles have a similar sized effect on growth rate (7%), the correlation between cell length and number of old poles is weak. Our results recapitulating the lower growth rate of cells with older poles are consistent with the original reports of F. Taddei and coworkers [38]. On the other hand, the statistical significance of other factors, like cell length at birth, suggest that there may be many determinants of cell growth rate with comparable statistical significance.

2.3 Materials and Methods

The analysis for this work was applied to data taken from a variety of new and previous studies. Data for the "Over-expression Phenotypes" and "Length as a proxy for cell age" experiments were initially collected for previous studies (respectively, [20, 27] and [21]. For a complete list of strains and imaging procedures used in this work, see the Appendix Section A.8 and Table A.7.

All imaging was conducted using a Nikon Ti-E inverted wide-field fluorescence microscope with a large format sCMOS camera (Andor NEO) and controlled by NIS-Elements. Image processing and analysis was completed using *SuperSegger* custom MATLAB (Natick, MA) software [1]. This software is available for download from the Wiggins Lab website (`http://mtshasta.phys.washington.edu/website/ssodownload.php`). For complete information about the download and manual for *SuperSegger*, please see Appendix Section A.4.

Chapter 3

DYNAMICS OF THE *E. COLI* CHROMOSOME

The structure of the *Escherichia coli* chromosome is inherently dynamic over the duration of the cell cycle. Genetic loci undergo both stochastic motion around their initial positions and directed motion to opposite poles of the rod-shaped cell during segregation. We developed a quantitative method to characterize cell-cycle dynamics of the *E. coli* chromosome to probe the chromosomal steady-state mobility and segregation process. By tracking fluorescently labeled chromosomal loci in thousands of cells throughout the entire cell cycle, our method allows for the statistical analysis of locus position and motion, the step-size distribution for movement during segregation, and the locus drift velocity. The robust statistics of our detailed analysis of the wild-type *E. coli* nucleoid allow us to observe loci moving toward midcell before segregation occurs, consistent with a replication factory model. Then, as segregation initiates, we perform a detailed characterization of the average segregation velocity of loci. Contrary to origin-centric models of segregation, which predict distinct dynamics for *oriC*-proximal versus *oriC*-distal loci, we find that the dynamics of loci were universal and independent of genetic position. This work was produced in collaboration with N.J. Kuwada, B. Traxler and P.A. Wiggins. Much of the contents of this chapter is published in Biophysical Journal ([21]).

3.1 Bacterial Chromosome Organization and Dynamics

Efficient and timely segregation of the chromosome is essential to bacterial cell proliferation. Although a segregation mechanism in *Bacillus subtilis* and *Caulobacter crescentus* has been extensively characterized, *Escherichia coli* is of particular interest due both to its historical role as a model bacterial system and the absence of homologues of the the well-established

partitioning genes *parAB*. The mechanisms responsible for DNA segregation in *E. coli* have not yet been identified.

E. coli possess a single, circular chromosome that is condensed into a nucleic acid and protein complex called the nucleoid [43]. Chromosome replication is initiated near midcell at the origin of replication (*oriC*) and proceeds bidirectionally around the chromosome [44, 45]. Before replicated sister loci can segregate, a brief "cohesion" period occurs, with intercatenation linkages holding the sister chromosomes together [46, 47, 48]. These linkages are protected by a sequestering protein, SeqA, which prevents Topoisomerase IV from removing linkages [48]. Following cohesion, sister loci are segregated sequentially [49, 50]. The underlying dynamics for the critical process of moving the replicated DNA remains poorly characterized and mechanisms responsible unknown [51, 19].

In order to understand the mechanisms for chromosome dynamics in this model organism, it is important to accurately characterize the spatial organization of its chromosome. However, there have been conflicting reports about the sub-cellular localization of loci in the *E. coli* nucleoid. Currently, one point of contention is the relative positioning of the left and right arms of the chromosome to *oriC*. In a *left-right filament* model of nucleoid structure, *oriC* begins at midcell, and each arm of the chromosome extends towards one of the cell poles (See Figure 3.1A). This model is supported by live-cell snap-shot imaging using either the Fluorescent Repressor-Operator System (FROS) or the ParB-*parS* system [52, 53, 54]. Our own live-cell quantitative characterization of the nucleoid in the model strain AB1157 showed that the left-right linear structure was maintained with a very high level of precision and that just 10% of chromosome (the *ter* region) was de-condensed and stretched between the left and the right poles of the nucleoid [24]. On-the-other-hand, analysis of MG1655 was consistent with a domain-based organization consisting of four structured macrodomains (*ori*, left, right and *ter*) and two unstructured regions [55]. Although these results were qualitatively consistent with the left-right filament structure, it was difficult to reconcile a *ter* macrodomain in MG1655 with the observation that it was decondensed in AB1157.

In another *ori-ter filament* model of nucleoid arrangement, *oriC* is positioned either at

the old pole or midcell, with the left and right arms of the chromosome folded together forming an *ori-ter filament* filament extending towards the new pole side of the cell (Figure 3.1B). This arrangement has been observed in a study characterizing *E. coli* in rapid growth [31]. Further complicating our current understanding of *E. coli* nucleoid arrangement, initial investigations of nucleoid structure using Fluorescence *In Situ* Hybridization (FISH) in fixed cells showed both the above-described arrangements successively. Early in the segregation process, *oriC* is positioned at the pole with the left and right arms of the chromosome folded together, forming an *ori-ter* filament [56, 57, 58] (as in the first frame of Figure 3.1B, but with *oriC* at the old pole). Then, later in the segregation process, *oriC* moves to midcell with the left and right arms on opposite sides of the origin, forming an left-right filament [56, 59] (as in the final frame of Figure 3.1A).

It is helpful to consider these inconsistencies in the larger context of bacterial chromosome organization by comparing the nucleoid structure *E. coli* to other model bacteria. As in *E. coli*, reports of *Bacillus subtilis* have suggested a variety of organizational patterns, including *ori-ter* [60, 61, 62, 63, 64, 11] or left-right arrangement [65, 66], as well as alternation between the two [67]. Reports in *Caulobacter crescentus* and *Vibrio cholerae* have been more straightforward, with both bacteria possessing *ori-ter* chromosome arrangement [49, 68, 69]. Furthermore, the segregation mechanisms present in these bacteria have been identified and extensively characterized; both *C. crescentus* and *V. cholerae* possess the well-known active partitioning *ParAB* proteins [70, 71, 72, 73, 74], while the homologous Spo0J-Soj system acts as the segregating mechanism in *B. subtilis* [75, 76]. With no homologous system yet discovered in *E. coli*, the mechanisms responsible for segregation in this model bacterium remain unknown.

It is possible that conflicting models in *E. coli* are the consequence of the failure to account for the cell-cycle dynamics of the nucleoid structure. In our study, we set out to address these unresolved issues by imaging cells over the entire cell cycle to capture the cell cycle dependent structure directly, instead of relying on cell length as a proxy for cell age. While many previous studies have made contributions to the effort of resolving nucleoid

arrangement inconsistencies, they are largely focused on short time scales (e.g. [77, 78, 79, 80]) and are therefore not suited to study the segregation processes which plays out on the time scale of the cell cycle. Those studies that have captured trajectories throughout the cell cycle analyze relatively few trajectories [81, 77, 82]. The capture and analysis of a large number of trajectories is of central importance. We have previously reported on a detailed study of the dynamics of *oriC* throughout the cell cycle in which we reported that locus segregation is achieved by loci undergoing *weakly-biased* stochastic motion [19]. To detect the weak bias in the motion, the motion must be averaged over many of trajectories in order to average out the stochastic motion.

In this study, we use quantitative methods and robust statistics to further probe the segregation dynamics of the entire chromosome. This effort to describe the cell-cycle dynamics of individual loci is achieved through four distinct steps: (i) to establish a detailed, quantitative and dynamic model for nucleoid structure throughout the cell cycle in the wildtype strain MG1655, (ii) to determine, analyze and compare the step-size distributions for *oriC* and other loci, (iii) to characterize the relationship between genomic position and the motional bias (or drift velocity) just after the initialization of segregation to probe the uniqueness of *oriC* dynamics, and (iv) to analyze these dynamics in a *seqA* deletion in order to perturb the segregation process at an early stage.

3.2 Materials and Methods

Strains and Growth Conditions Single chromosomal loci were fluorescently labeled using a ParB-*parS* system. These strains were constructed from an MG1655 background with *parS*-CmR cassettes inserted at different intergenic loci (Figure 3.2A) [77]. A full list of these insert locations is provided in Table A.1 of the Appendix. The IPTG-inducible ampicillin-resistant plasmid pALA2705 encoding the fusion GFP-ParB [50] was transformed into these strains for fluorescent imaging of loci. For full-nucleoid imaging, an IPTG-inducible ampicillin-resistant plasmid harboring a Fis-GFP fusion [83] was transformed into MG1655. For a 1

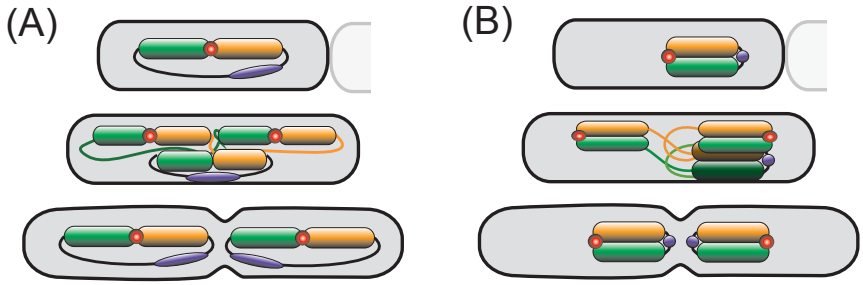

Figure 3.1: **Cellular-scale models for chromosome structure.**Schematic models of the nucleoid with a (A) *left-right* filament structure and (B) *ori-ter* filament structure. The left (right) arm of the chromosome is green (orange). *oriC* is shown in red and *ter* is in purple. From top to bottom, the cells in each stack represent chromosome structure at cell birth, during chromosome replication, and before division.

hr cell cycle, cells were grown in M9 minimal media supplemented with 0.2% glycerol as the carbon source and 0.2% casamino acids, 10 μg/mL Thiamine hydrochloride and 100 μg/mL Ampicillin. All experiments were performed in this media unless otherwise specified. For slow growth (2 hr cell cycle), casamino acids were replaced by 100 μg/ml of Arginine, Histidine, Leucine, Threonine and Proline. For locus imaging, overnight liquid cultures were diluted into fresh media and grown to $OD_{600} \sim 0.1 - 0.2$ at 30 °C. Cells were induced for 30 minutes at 30 degrees C, with 0.5 mM IPTG. For nucleoid imaging, overnight liquid cultures were diluted into fresh media and cultured for at least three doubling times before cells were induced. ParB-GFP was induced with 1 mM IPTG for 30 minutes at 30 degrees C, then harvested, and inoculated onto agarose pads for imaging.

Sample Preparation and Microscopy. Slides were prepared by pouring 1 mL growth media mixed with 0.2% LM-agarose (Cat. no. 16520-020) into 1.5 cm square wells. 2 μL of induced cells are spotted onto the pads and then covered with a coverslip and sealed using VaLP (1:1:1 vaseline, lanolin and paraffin). To minimize drift slides were allowed to rest 1 hour at 30 degrees C prior to imaging. Imaging was completed using a Nikon Ti-E inverted wide-field fluorescence microscope with a large format sCMOS camera (Andor NEO) and

controlled by NIS-Elements. Samples were kept at 30 degrees C throughout the imaging process using an environmental chamber. The total imaging time for each slide was 4 hours, with an imaging frame rate of 1 frame per min. For each frame, both brightfield and fluorescent image were captured. Image processing and analysis was completed using custom MATLAB (Natick, MA) software [19].

3.3 Results

Complete Cell-Cycle Imaging. To observe locus positioning within growing cells, we labeled loci with the widely used ParB-*parS* system [77]. Seven strains were constructed in total, each with an exogenous *parS* inserted at a unique intergenic chromosomal locus. These insertions were found to have no measurable effects on cell growth or morphology (Fig. 2.1), with cells doubling every 50 minutes. Chromosome structure was imaged using wide-field fluorescence microscopy at 1 minute intervals for 4 hours, allowing us to capture the dynamics of fluorescently labeled loci throughout the entire cell cycle in thousands of cells over multiple generations. Locus dynamics are visualized in kymographs (Fig. 3.2B) which display long-axis locus dynamics as a function of time. At each time point, the brightfield image of the cell is used to identify the cell boundaries, and the position of the fluorescently-labeled chromosomal loci along the long-axis of the cell is determined by fitting a Gaussian point spread function. The kymograph is thus the trajectory of the chromosomal locus through time, with long axis position of the locus along the y-axis of the kymograph and time from cell birth along the x-axis.

One significant concern regarding all live-cell DNA labeling methods is the introduction of exogenous cohesion between the sister chromosomes due to the labeling system [84]. Complete-cell-cycle imaging facilitates a detailed analysis of the duration of the cell cycle and the lengths of new born and dividing cells. Analysis of these distributions reveal that the induction of the labeling system does not significantly perturb the cell cycle, as would be predicted if there were significant cohesion induced by the labeling system. (See Fig. 2.1)

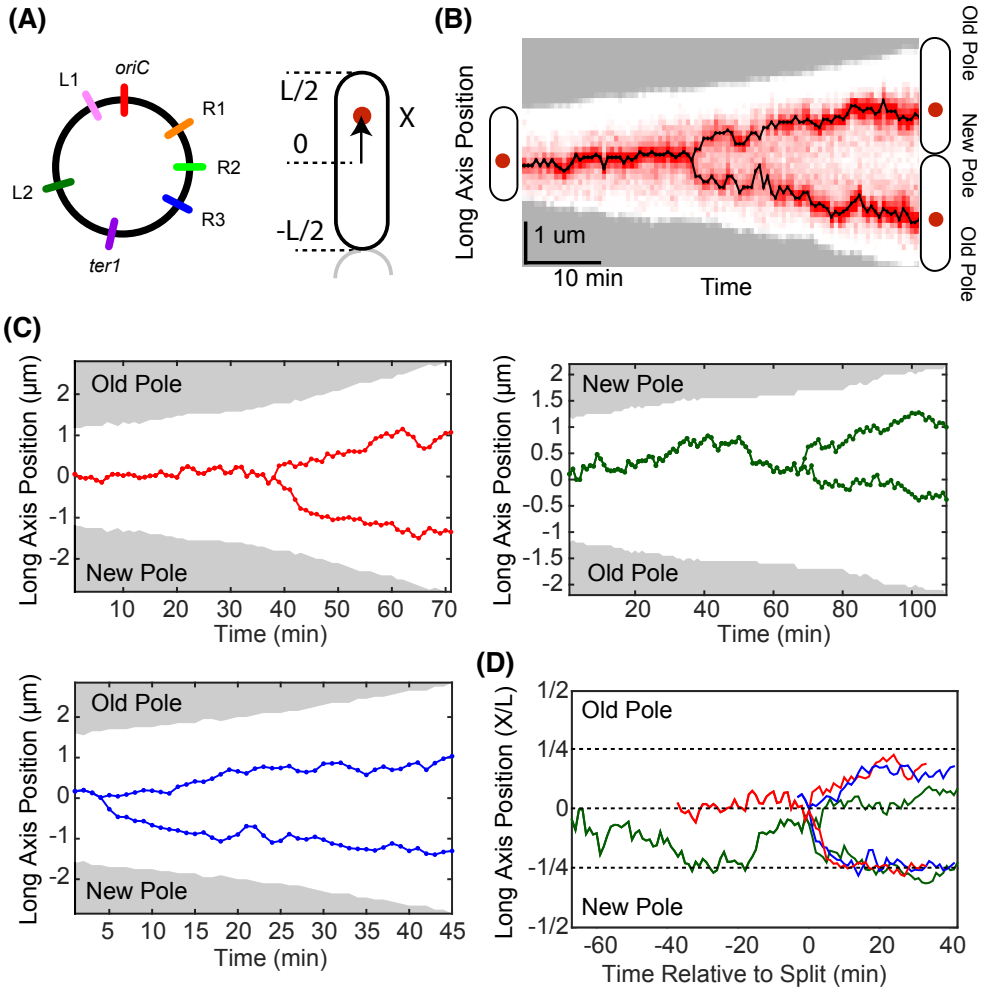

Figure 3.2: **Trajectory Alignment.** (A) Chromosome map of all labeled genetic loci and sample cell to visualize normalization of positions to cell length units L. $X = 0$ indicates midcell, and $X > 0$ ($X < 0$) represents the old (new) pole sides of the cell. (B) An example kymograph. To produce kymographs, cell images were projected along the long axis of the cell and aligned sequentially at midcell. Gray regions indicate cell boundaries determined by the bright field image. Red pixels indicate a fluorescent signal. (C) Three sample fluorescent tracks from *oriC* kymographs in physical coordinates display the high level of cell-to-cell variation in localization dynamics. (D) Tracks from (C) are normalized to cell length at each point in time, oriented by cell pole, synchronized by splitting times, and overlaid. This method, when repeated with thousands of tracks, can then be used in order to generate a 3D histogram of locus positioning throughout the cell cycle.

Locus trajectory alignment. Figure 3.2C demonstrates that there is significant cell-to-cell variation in the length of the cell cycle, cell morphology, and the timing of replication initiation. Therefore, in order to statistically analyze data from many independent cells, we required a consistent and tractable method to combine locus trajectories from thousands of cells. To achieve this goal, we determine the long axis displacement of each locus from the cell center X_i at each time point i as a fraction of cell length L_i to generate the relative long axis position X_i/L_i (Fig. 3.2A), and (ii) temporally align each trajectory by its split time, i.e. the time in which newly replicated chromosomal loci copies become visually distinct prior to segregation (Fig. 3.2D). As trajectories are calculated from projections along a single axis, we also have to choose a consistent coordinate system to orient the long-axis of the cell. Since we observe the entire cell cycle, we can orient the cells either by the orientations of the new pole (polar orientation) or by the average position of the locus at the end of the cell cycle (genomic orientation), as described in Section A.2 of the Appendix.

Locus occupancy dynamics. To visualize the typical cell-cycle dynamics of loci, we plot the *locus occupancy* as function of relative cellular position and time relative to the locus splitting time. The locus occupancy is a two-dimensional histogram of locus position with long-axis locus position on the y-axis and time on the x-axis. We have generated histograms both for polar and genomic orientation. The polar-oriented occupancy for four different loci is shown in Figure 3.3. The mean trajectory for each locus is shown as a solid black line.

Consistent with the canonical model of locus positioning in slow growing cells [85], the *oriC* locus begins the cell cycle at midcell and moves to the quarter cell positions over the period of roughly ten minutes. *oriC* then remains at the quarter cell positions for the rest of the cell cycle. Also consistent with the filament model [19, 24], a majority of the chromosome appears well positioned about a mean position, with the exception of *ter*-proximal loci, which display a wide distribution of locus positions. For example, the *ter*-proximal locus *ter1* shows a wide localization distribution throughout the new pole side of the cell prior to splitting as previously reported [24], although the locus appears to localize to midcell prior to splitting.

This behavior for *ter1* has been observed previously [86]. We, however, observe a more broad pattern of this behavior; consistent with the *ter1* and R3 locus occupancies, regardless of pre-split localization, all loci appear to split in close spatial proximity to midcell with high fidelity (See Fig. 3.4C). In the canonical model of *E. coli* replication, the fork moves with the chromosome, and therefore loci would be expected to initiate segregation from their home position. In contrast with this model, all loci move towards midcell before splitting.

But the *ter1* locus occupancy also shows a very significant quantitative differences with the quantitative map of nucleoid structure in AB1157 that we previously reported [24]. In MG1655, R3 is localized at roughly the new-pole quarter-cell position early in the cell cycle. The left-right filament model predicts that this histogram should show peaks at both the new and old-pole quarter-cell positions. Orienting the chromosome by genomic orientation is also consistent with R3 localizing to the new pole position.

The dynamics of mean locus position. To compare the dynamics of loci it is convenient to compute the mean locus position during the cell cycle. (Note that this perspective is less meaningful for *ter*-proximal loci since the mean is not representative of the typical locus position as can be seen from the locus occupancy of *ter1* shown in Figure 3.3.) We compute the mean using both genomic and polar-orientations as described in section A.2 in the Appendix. These trajectories are plotted in Figure 3.4 as a function of the cell-cycle age, relative to the *oriC* split time.

As described previously, *oriC* is initially positioned at midcell prior to splitting, and then moves to the quarter-cell positions following splitting. The data has excellent sampling, and the mean can be seen to transition smoothly between midcell and the quarter cell position with no obvious features.

But for non-*oriC*-proximal loci, the polar-orientation mean trajectories reveals a structure consistent with the structure implied by the positioning of R3: The *ori-ter* filament structure appears to be predominant in MG1655. Since all loci are localized closer to midcell on average than the origin after segregation initiates at $t = 0$ min, the *ori-ter* filament structure

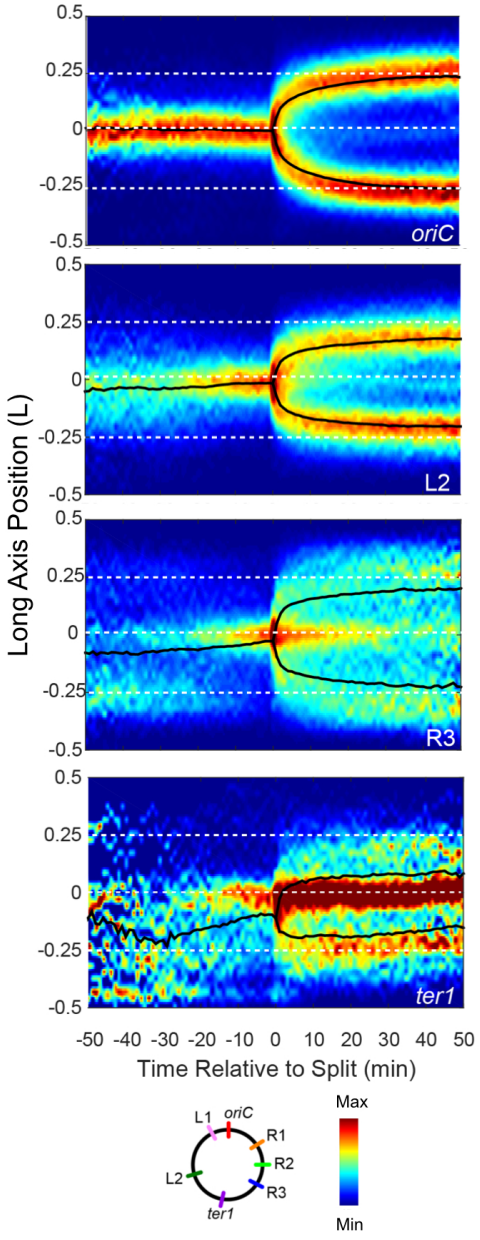

Figure 3.3: **Trajectory Histograms.** Histograms of synchronized trajectories (see Figure 1D) throughout the cell cycle for several chromosomal loci. Tracks are oriented with $X > 0$ indicating the old pole side of the cell and $X < 0$ indicating the new pole. The locus label and number of cells contributing to each histogram are as follows: (A) *oriC*, N = 3528 (B) L2, N = 2254 (C) R3, N = 1416 (D) ter1, $N = 406$. Histograms for all seven loci are included in Figure A.2 in the Appendix.

(*i.e.* Fig. 3.1B) is predominate over the left-right filament structure (*i.e.* Fig. 3.1A) . These observation are consistent with the structure that has been reported for rapidly proliferating cells [31].

To test whether the observed *ori-ter* filament structure was an artifact of the GFP-ParB labeling system conferred by sister cohesion, we repeated the experiment using the strain AB1157 under slow growth conditions. This strain has a *left-right* filament structure (e.g. [24]). As shown in Figure A.3, the right arm of the chromosome is not oriented with respect to the new pole in AB1157, even when labeled using the ParB-*parS* system under analogous imaging conditions. Therefore, ParB-*parS* does not appear to result in new pole localization in complete-cell-cycle imaging experiments.

These observations are not wholly unexpected since both Niki *et al.* [56] and Bates and Kleckner [59] reported similar patterns in fixed cells where the chromosome is oriented in *ori-ter* configuration for a significant fraction of the cell cycle. Furthermore, this appears to be the dominant structure in rapidly proliferating cells [31].

Dynamics of the nucleoid density. The mean locus position and the occupancy also imply that on average the loci are all much closer to midcell than observed in AB1157 [24]. To confirm that this more-condensed nucleoid was not an artifact of locus labeling or analysis, we directly imaged the nucleoid using a Fis-GFP fusion [83].

The Fis-GFP fusion non-specifically coats the nucleoid without affecting cell growth, allowing us to visualize *in vivo* DNA localization throughout the cell cycle. As in the locus tracking experiments, this cell-cycle imaging offers us the ability to orient cells by pole age. Example snapshots of labeled nucleoids in individual cells at different points in their life cycle are shown in Fig. 3.5A. These images show a clear asymmetry in DNA positioning, with a strong bias towards the single new pole at the beginning of the cell cycle and towards the cell center (which will become the two new poles) near the end of the cell cycle. In order to get an average nucleoid shape distribution throughout the cell cycle, we build a consensus localization pattern, as described in the Ref [20].

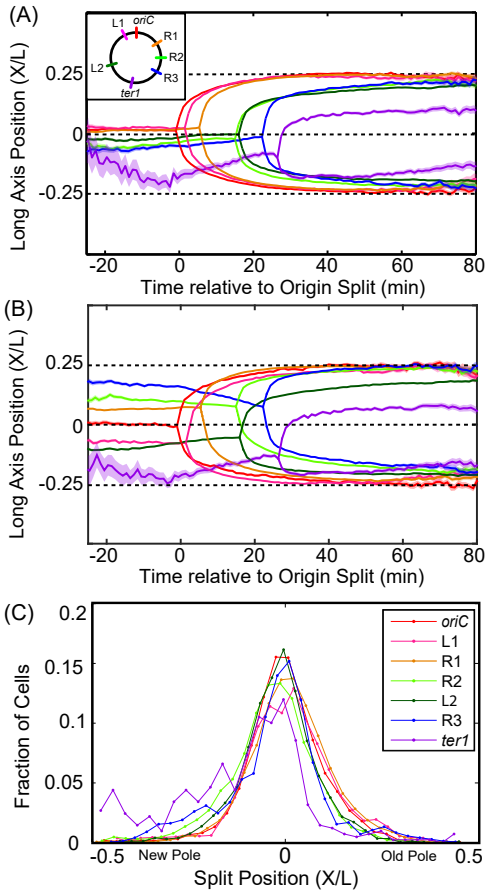

Figure 3.4: **Mean Locus Position Trajectories and Histogram of Split Positions.** Mean locus trajectories aligned by split time oriented using polar (A) and genomic orientation (B). The observation that all loci lie inside *oriC* demonstrates the predominance of the *ori-ter* nucleoid orientation. Loci are also observed to move toward midcell shortly before the initiation of segregation. The tight spacing of the mean trajectory curves demonstrates that the nucleoid configuration is much more compact than observed in AB1157. The noisiness of the mean trajectory at long and short times is due to the small number of cells with cells cycles significantly exceeding 1 hr. (Error regions show the error in the mean assuming all observations are uncorrelated.) (Note that the mean is less meaningful for *ter*-proximal loci, as their positioning before and after the split is less precise, as can be seen from the locus occupancy of *ter1* shown in Figure 3.3.) (C) Histogram of splitting locations for all loci in all cells when trajectories are aligned by pole orientation (as in (A)) . Regardless of genetic location or locus long-axis positioning at cell birth, all seven loci split near midcell with high fidelity.

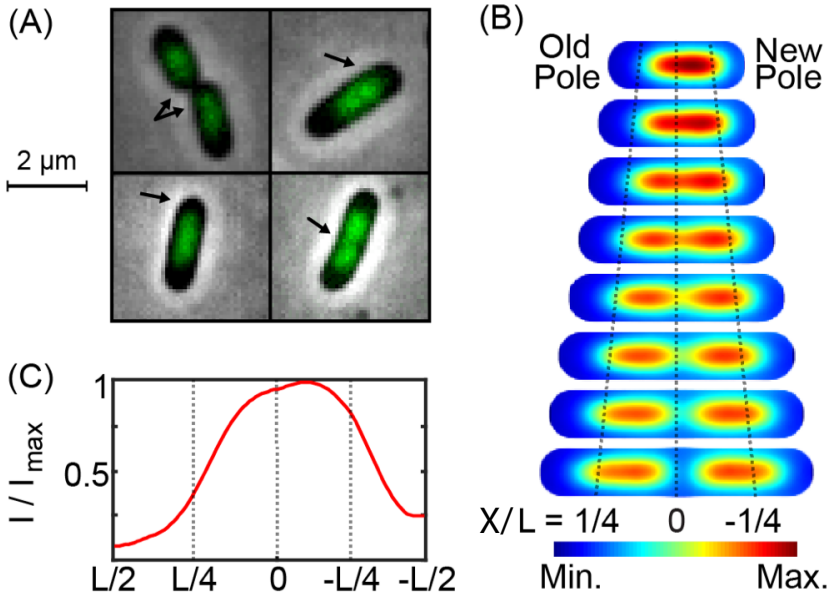

Figure 3.5: **Whole-nucleoid imaging.** (A) Fis-GFP labeled nucleoids in single cells. Arrows indicate the new pole (or future new poles) of cells. (B) Consensus localization image of labeled nucleoids from 230 complete cell cycles shows a compact nucleoid asymmetrically distributed towards the new cell pole. (C) Mean DNA density in the first frame of the cell cycle oriented by cell pole (right side, negative).

We observe that the consensus localization pattern is consistent with our model: the regions of highest chromosomal concentration fill no more than 50% of the cell length at all points in the cell cycle, and tend to be localized asymmetrically towards the new-pole at cell birth (Fig. 3.5C) and therefore we find the overall nucleoid conformation to be significantly more compact in MG1655 than we previously observed in AB1157 [24].

Step-size distributions. Both the locus occupancy distributions and the mean locus position trajectories provide insight into the global dynamics of segregation, but the step-size distribution provides that most direct probe of how the locus motion is perturbed on a local level during the segregation process. The step size is defined:

$$x_i = X_{i+1} - X_i, \tag{3.1}$$

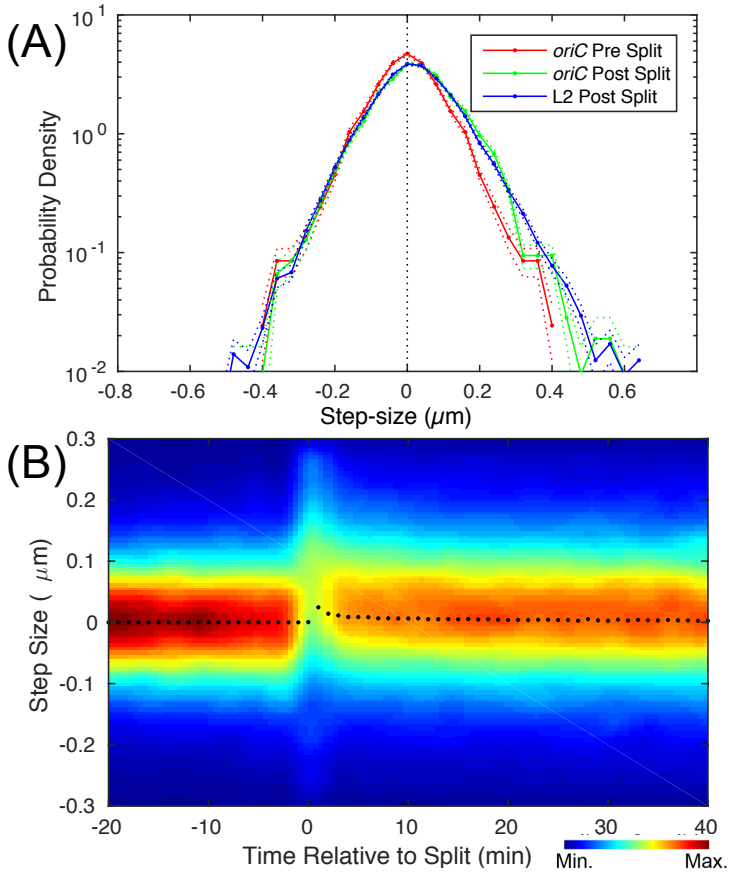

Figure 3.6: **Step-size distributions.** (A) The step-size distribution is the probability of observing a step size of a lag time of one minute. The width of the distribution characterizes the stochasticity of the motion whereas the mean step-size characterizes the directedness of the motion. The initiation of segregation results in a weak bias of the stochastic motion in the direct of average motion. The motion of *oriC* transitions between unbiased motion (red) to weakly biased motion (green), as seen by the shift in the distribution. The *oriC* (green) and L2 (blue) show essentially identical stepsize distributions during the first ten minutes of motion. (Error regions show expected counting error.) (B) Time-dependent histogram of *oriC* step-sizes. Aside from an increased spread in step sizes at the time of locus splitting (t=0), the distribution in step size is quite homogeneous for the times before and after the split, with a zero bias and consistent spread prior to the split and a similar spread but small positive bias after the split.

where X_j is the locus position in frame j and the time interval between frames is one minute.

The null hypothesis for the step-size distribution is a normal or Gaussian distribution, with probability distribution function:

$$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{3.2}$$

where μ is the mean (which we will refer to as the bias) and σ is the standard deviation, that is directly related to the mobility of the locus. The Gaussian distribution emerges in quite generic circumstances, even if the underlying dynamics is not diffusive and the cell (*i.e.* media) in which the loci (*i.e. particles*) are diffusing is disordered (*e.g.* [87]).

A number of the existing models of chromosome structure make predictions that are most naturally quantified via analysis of the step-size distribution. For instance, Kleckner and coworkers have proposed that large displacements, which are observed during the segregation process, correspond to a large-scale structural rearrangement of the chromosome due to the unsnapping of structural elements, referred to as *chromosome snaps* [88]. Our own analysis of the motion of *oriC* in the model strain AB1157 did identify large steps [19, 80], but we could not reject the null hypothesis that large steps were a generic feature of chromosome loci, irrespective of whether loci were undergoing segregation or stay-at-home motion [19, 89]. Since AB1157 has a significant number of mutations which could inactivate some segregation mechanisms, it was therefore interesting to repeat our analysis in wildtype cells (MG1655) [90].

To compare step-size distributions between strains and phases of the cell cycle, we use the KL Divergence, a natural measure of the difference between two distributions p and q [91, 92]:

$$K \equiv \int \mathrm{d}x\ p(x) \log \frac{p(x)}{q(x)}. \tag{3.3}$$

The divergence is interpreted as the information loss of modeling true distribution p with model q[1]. We will interpret the divergence as the *distance* between distributions. For identi-

[1]Note that the frequentist statistical test for whether two distributions are identical is the Kolmogorov–Smirnov test. This is *not* the question of interest. Our interest is measuring the difference between many distributions that are statistically distinct.

cal distributions, K is identically zero otherwise $K > 0$. The magnitude of K is interpreted as follows: The number of observations before information loss is significant is $N \approx K^{-1}$.

Our analysis of the step-size distributions in MG1655 were consistent with our previous observations of *oriC* in AB1157 [19]. As previously reported, we observe that the step-size distribution has fat-tails: The number of large magnitude steps is greater than predicted by diffusion (a Gaussian step-size distribution) [82, 19, 80]. Two lines of evidence argue against this phenomenon playing a generic and indispensable role in the segregation process.

(i) The existence of large steps alone does not necessarily imply that these steps are the consequence of an active and segregation-specific mechanism. Similar fat-tailed step-size distributions are generically observed for other large complexes tracked in the cell [93, 94, 4, 80]. To quantitate the degree of failure of the diffusion model, we compare the divergence between the observed and diffusion-modeled distributions for both *oriC* and a control data set of non-functional MS2-GFP-mRNA complexes [4] (Fig. A.5):

$$K_{oriC} = 0.026 \quad \text{and} \quad K_{\text{MS2}} = 0.13, \tag{3.4}$$

demonstrating that the observed *oriC* step-size distribution is in fact far more Gaussian than one might predict from the dynamics of a large exogenous complex, which is expected to have no specific interactions. Clearly the non-Gaussian distribution is not remarkable in itself.

Furthermore the absolute size of the divergence is quite small. For our experimental frame rate we are able to capture ten steps in the first ten minutes of segregation during the most rapid motion. The information loss from those ten steps cumulatively is $10 \times K_{oriC}$, which is still not significant (i.e. > 1). Therefore the typical cell *does not* experience steps that are significantly different from the predictions of a Gaussian step-size distribution, but one in four cells do.

(ii) Although the snap model predicts large steps in the direction of increased locus separation, we observe large steps in both directions with nearly equal probability. To demonstrate this, we generate two step-size distributions: $p(x)$ and $q(\delta x)$ (Fig. A.6). $p(x)$

is the step-size distributions during the first ten minutes of locus segregation, oriented in the direction of the position of locus at the end of the cell cycle (A.4A). $q(\delta x)$ is identical to p except it is the distribution of mean-subtracted steps $\delta x \equiv x - \overline{x}$ (A.4B). We then compute the divergence between $p(x)$ and $p(-x)$ and between $q(\delta x)$ and $q(-\delta x)$:

$$K_{pp} = 9.7 \times 10^{-2} \quad \text{and} \quad K_{qq} = 5.7 \times 10^{-3}, \tag{3.5}$$

respectively. K_{pp} measures the size of the bias in the direction of motion. If the motional bias is caused by a simple shift of the distribution (a drift velocity) then $K_{qq} = 0$. Since K_{qq} is just 6% of K_{pp}, it suggests that the bias is almost entirely created by a drift velocity since a shift removes all but 6% of the bias-induced divergence. In summary, although large steps are observed, they do not appear to be the dominant mechanism of movement of loci.

The analysis of the dependence of the step-size distributions on locus genomic position could potentially provide insight into proposed *oriC*-centric models of segregation if significant differences in the distributions were observed. But, in the first ten minutes of segregation, we observe the step-size distribution of loci at different genomic positions to be strikingly similar. The distributions for L2 and *oriC* are shown in Fig. 3.6. The divergence between the *oriC* and L2 distributions is just $K = 5.1 \times 10^{-3}$. We shall discuss the significance of this result in the "Drift Velocity" section of this chapter.

Sub-diffusion and the Gaussian step-size distribution. It has already been reported that loci undergo sub-diffusive motion [81, 77, 19, 80, 95, 96, 78, 97, 79]. It is important to note that the observation of a Gaussian step-size distribution, although consistent with diffusion, does not imply the motion is diffusive (*e.g.* [87]). If the motion were described by diffusion, the step-size variance (*i.e.* the MSD) would scale linearly with time $\sigma^2 = 2Dt$, but in fact $\sigma^2 \propto t^\alpha$ where the scaling exponent α has been reported as sub-diffusive: $0 < \alpha < 1$ [81, 77, 80, 95, 96, 78, 97, 79], consistent with our own observations [19]. But our immediate interest is bias in the motion, the component of the motion that does not average to zero during segregation and is therefore responsible for ensuring that one locus is localized to each daughter at the end of the cell cycle [19]. Since the step-size distributions are well

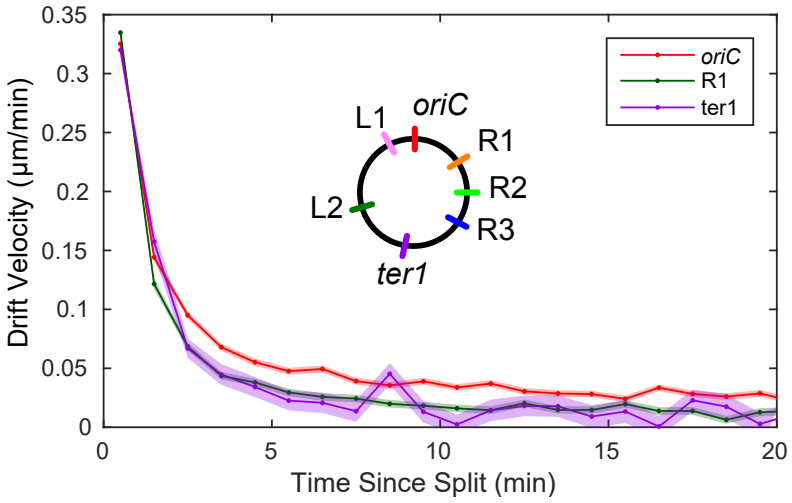

Figure 3.7: **Mean drift Velocities.** Profiles of the relative drift velocities of sister loci are provided for *oriC*, L2 and ter. At short time periods after the initial splitting of sister loci, all loci appear to have similar drift velocities, consistent with a model for segregation which treats all loci identically. At later time periods, sister *oriC* loci have larger relative drift velocities, allowing them a larger net separation along the length of the cell. (Error regions show the error in the mean.)

approximated by a Gaussian distribution with a time dependent bias, as seen for *oriC* in Figure 3.6B, it is convenient to focus on a discussion of the time dependence of the bias (drift velocity).

Drift velocity. The drift velocity is defined as the average bias in the locus motion:

$$\overline{v} \equiv \mu/\tau, \tag{3.6}$$

where $\mu \equiv \overline{x}$ is the averaged (oriented) step size and τ is the inverse frame rate. The drift velocity is the property of locus motion which underlies the segregation process [19]. As we previously described at *oriC*, loci transition from nearly unbiased motion ($\overline{v} = 0$) to biased motion ($\overline{v} \neq 0$) during segregation, before returning to unbiased motion around the locus home position in slowly growing cells [19]. A second motivation for the analysis of the drift velocity is that in models it is related to the average applied force [19, 95]. It is important to note that we define the drift velocity in terms of a finite difference in Eqn. 3.1 and therefore

the drift velocity should not be interpreted as an instantaneous velocity. (See Section A.3 of the Appendix for a more detailed discussion.)

To characterize the genomic-position and cell-cycle dependent variation in segregation dynamics, we measured the drift velocity of foci both as a function of genomic position and period of the cell cycle. Again, our expectation was that *oriC* would be distinct since replication initiates at this locus and loci are known to segregate sequentially beginning with *oriC* [50, 98]. It is generally believed that whatever the mechanism driving segregation, *oriC* is expected to undergo different dynamics (*e.g.* [99]). Our null hypothesis is that all loci move with identical dynamics.

As we previously reported for the origin [95], the drift velocities for all loci follow an approximate power with time (Appendix, Fig. A.8). Fig. 3.7 shows the drift velocity for three loci (*oriC*, L2 and *ter*) as a function of time since the locus-segregation initiated. The observed drift velocity clearly rejects the null hypothesis that all loci are identical, but strikingly this mismatch occurs late in the segregation process rather than early. Shortly after segregation initiates, all loci experience large, but nearly identical drift velocities:

$$\overline{v} = 0.33 \ \pm 0.01 \ \mu\text{m/min}. \tag{3.7}$$

The observation of a genomic-position-independent drift velocity is counter intuitive since large drift velocities correspond to large forces that are experienced early in the segregation processes and that we had speculated were the consequence of an active mechanism. But, with respect to these early drift velocities, the behavior at *oriC* is generic: All other loci characterized had similar drift velocities to *oriC*.

In order to test whether the universal drift velocity was an artifact of the ParB-*parS* labeling system, we compared the dynamics of loci labeled by LacI-mCherry and GFP-ParB in the same strain. Nearly identical drift velocities were observed in this strain as well (Fig. A.9).

***seqA* mutant.** To perturb the motion immediately after the initiation of segregation, we characterized the drift velocity in a *seqA* mutant. SeqA is a regulator of replication initiation

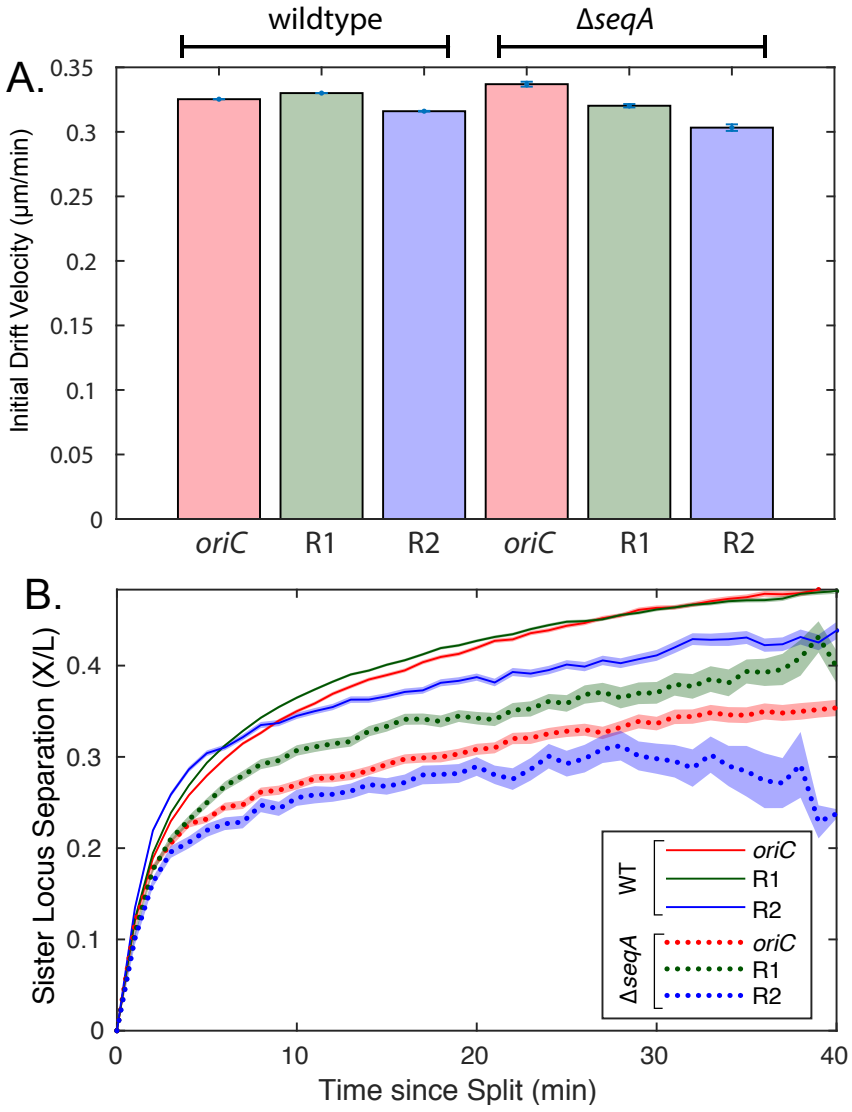

Figure 3.8: **Segregation phenotype of *seqA* mutant.** To perturb DNA structure shortly after replication, we constructed a *seqA* deletion. (A) Δ*seqA* has a nearly universal initial drift velocity. (B) The smaller separation for *seqA* loci implies that all loci move segregate with a lower drift velocity. This is observed as the integrated effect of a velocity which is only slightly smaller acting throughout the segregation process.

which acts by sequestering hemimethylated GATC sites during the replication process [100]. Dam methyltransferase competes for the SeqA binding sites and methylates them, reducing SeqA binding affinity. SeqA binds newly replicated DNA specifically and binds many sites in the vicinity of *oriC*. It has been hypothesized that SeqA might play an important structural role in physically sequestering SeqA sites in addition to its regulatory role [101, 102, 48, 103]. We therefore tested whether a *seqA* mutant perturbs segregation dynamics and in particular, segregation dynamics during the initial rapid translocation period.

We capture chromosome dynamics in MG1655 Δ*seqA*. As expected, we observed that a significant fraction of the cells had multiple copies of *oriC* at the beginning of the cell cycle. To characterize changes that were not the result of increased DNA copy number, we analyzed only cells with a single copy of *oriC* at the beginning of the cell cycle.

To investigate whether SeqA perturbs the initial locus segregation dynamics, we computed the drift velocity as a function of time after the initiation of segregation. Although SeqA does perturb the segregation dynamics, the initial segregation velocity is nearly identical to that observed in wildtype cells, as shown in Panel A of Figure 3.8. In Δ*seqA*, like wildtype cells, *ori*-proximal loci move more rapidly later in the segregation process. But, the fastest moving loci in the Δ*seqA* cells were still slower than all locus positions observed in the wildtype cells. These observations are consistent with SeqA playing either a direct or indirect role in facilitating the movement of loci to their home positions, but interestingly it has only a weak effect on the initial locus dynamics.

Growth phenotypes

While our original goal in this project was to study segregation dynamics, the chromosomal organization we observed drove our interest in unifying the dissenting models for *E. coli* chromosome structure in MG1655 and AB1157 studies.

One hypothesis that could link these seemingly disparate models is that the chromosome changes its spatial organization in different growth phases. It has been suggested that in stationary phase the chromosome becomes "crystalline" or takes on a rigid structure, which

may differ from the structure under log growth conditions. It is possible then, that the *left-right* structure could be restored in MG1655 for re-germinating cells (cells which have been in stationary-phase and are returning to log-phase growth).

To address this possibility, an upcoming has been designed to apply our cell-cycle dynamics mapping to cells that have been grown to stationary, and are imaged on pads that either contain nutrients to re-start growth (media pads) or do not (PBS buffer pads). The imaging of cells on media pads will allow us to observe whether the structure of the chromosome changes as cells transition back into growth, while the PBS imaging acts as a control. While most of this project is forthcoming, we have begun preliminary experiments by looking at just the WT *oriC* ParB-GFP labeled strain. Cells were grown to stationary in M9 Casamino media with IPTG for induction overnight at 30 degrees. The cells were then removed from the heater and left on the bench at room temperature for three days. This prep was repeated for the two experiments: (i) stationary - the stationary cells were then spun down and resuspended in PBS buffer for 24 hours, then imaged on PBS agarose pads, and (ii) re-entering log growth - the cells were were imaged on M9 Casamino pads (no wash step).

The cells imaged on PBS were intended to serve as a control, in which all cells possessed one or two copies of the chromosome with no notable dynamics (i.e. the chromosome simply diffuses around a home position). We were surprised, however, to find that 6% of these cells had fluorescence tracks exhibitng a single copy of the chromosome splitting, with sister copies segregating apart (as the tracks in Fig. 3.2). This result was surprising, and we hoped that we may gain insight into these unexpected segregation events by comparing those dynamics to the segregation dynamics in cells imaged on media pads.

Since the number of segregating tracks in the stationary dataset was so small, a localization heatmap approach (as in Fig. 3.3) was not possible. Instead, we chose to compare the mean-squared displacement (MSD, See "Dynamics on the Micron Scale" in Chapter 1) of the loci in the stationary growth experiments. By comparing the scaling factor (α), we could quantitatively compare the segregation dynamics between the two cases.

As shown in Chapter 1, the MSD for sub-diffusive loci obeys the relation:

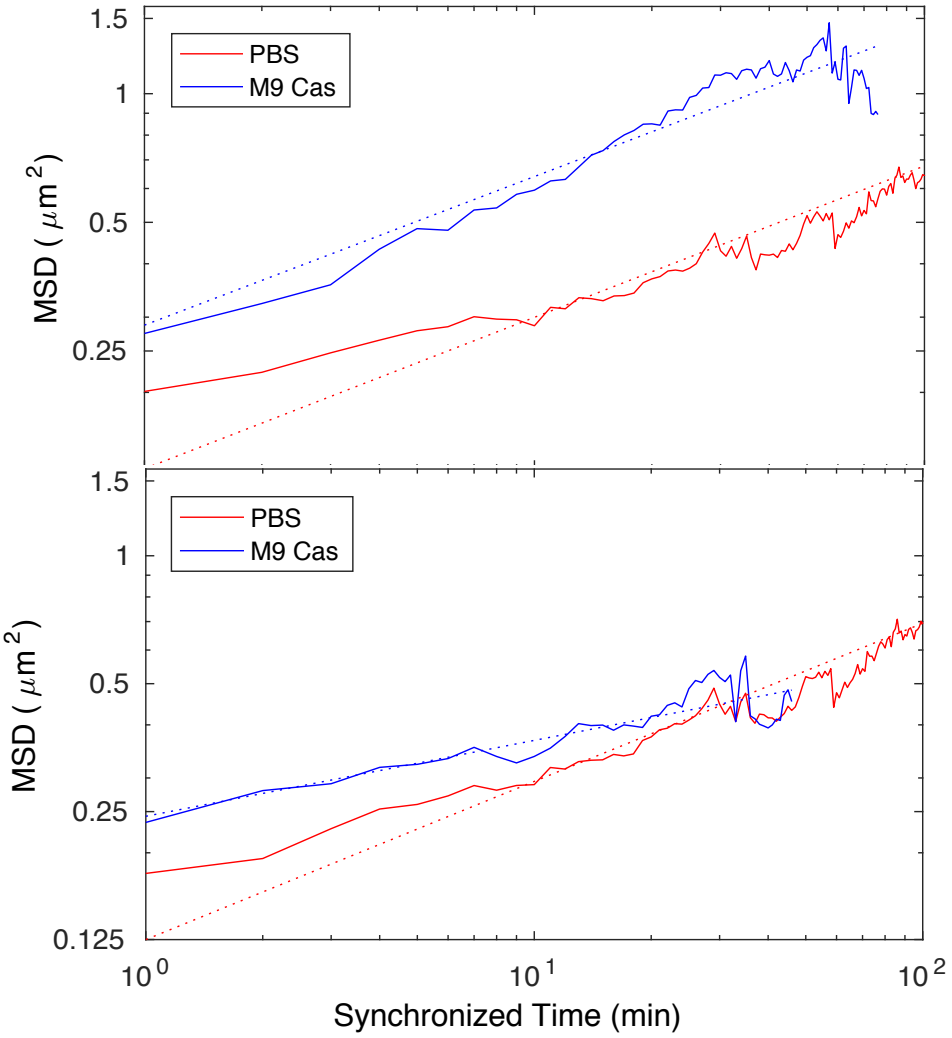

Figure 3.9: **MSD scaling for stationary cells imaged with and without nutrients (media and PBS pads, respectively). (Top) When all generations of cells were included in our analysis, the scaling parameter for both populations of cells was approximately the same (≈ 0.35). However, when we isolated only the first generation of cells in both imaging experiments, we observed a greater disparity between their dynamics. The scaling parameter of cells imaged on PBS and media pads, respectively, were 0.38 and 0.18.**

Pad	Generations	Scaling (α)
PBS	All	0.37
M9Cas	All	0.18
PBS	First	0.35
M9Cas	First	0.35

Table 3.1: **MSD Fits.** Fill this in later.

$$\langle \Delta x^2 \rangle = 2Dt^{\alpha}. \tag{3.8}$$

Thus, on a log-log plot of MSD vs time, we can calculate the scaling factor of the MSD as the slope of the line:

$$log(\langle \Delta x^2 \rangle) = log(2D) + \alpha log(t). \tag{3.9}$$

The resulting MSD's and fits are shown in Figure 3.9, with the scaling parameters for each fit provided in Tab. 3.1. When we included all generations of cells, we observed that the scaling factors for cells on both pad types were approximately the same. However, later generations of the cells grown on media pads are typically in log growth and therefore no longer indicitive of cells transitioning out of stationary phase. To isolate cells that are truly transitioning from stationary to growth phase, we repeated the analysis, including only the first generation of cells. In this analysis, we found that the stationary cells had a higher scaling factor than the cells re-entering lod-phase growth. This indicates that the motion of the segregating chromosomes in the stationary cells is more highly directed (less reminiscent of diffusion) than the motion of chromosomes in the cells re-entering log phase! While further studies are needed to further probe this result, these unexpected dynamics could be the result of unexpected metabolism of cell surroundings in stationary cells... or they could bare exciting implication about bet-hedging in a stationary population of cells.

3.4 Discussion

In this study, we have used complete cell cycle imaging to observe dynamic chromosome structure throughout the entire *E. coli* cell cycle. We aim both to map out the cell-cycle-dependent structure of the chromosome and to perform a detailed analysis of the segregation dynamics of loci.

Conflicting *ori-ter* (*i.e.* Fig. 3.1B) and *left-right* (*i.e.* Fig. 3.1A) filament organizations have been reported for the *E. coli* nucleoid. Under slow growth conditions, with a two hour cell cycle, we had previously performed a detailed quantitative analysis of the strain AB1157 and reported a precisely-structured *left-right* filament organization [24], consistent with other recent live-cell measurements of the chromosome structure in *E. coli* [53, 52]. In this study, we characterized the wildtype strain MG1655 with a one hour cell cycle. Here we report that the dominant structure under these conditions is the *ori-ter* filament organization, consistent with some previous reports using FISH [56, 59, 57, 58] and recent measurements characterizing the chromosome structure during multifork replication [31]. We have supported this report using histograms of locus positions (Fig. 3.3), mean locus positions (Fig. 3.4) and nucleoid imaging (Fig. 3.5); while each of these methods has its own limitations, we feel that the consistency between these results provides strong support for an *ori-ter* filament, with loci splitting at midcell.

Consistent with rapid growth, we also see a segregation pattern of locus movement more reminiscent of a replication factory model [65]: Loci were observed to localize to midcell, before segregation initiated, then split and move rapidly outward to their home positions. This pattern is most striking for loci on the left and rights arms of the chromosome far from *oriC*. These observations suggest that although the replication forks in *E. coli* has been observed to separate (*e.g.* [86]), conflicting with a literal interpretation of the replication factory model, they are likely in close proximity for most of the cell cycle, consistent with the spirit of the replication factory model [65].In light of our results, we provide a figure summarizing the current models for chromosome and repication forks in the replication

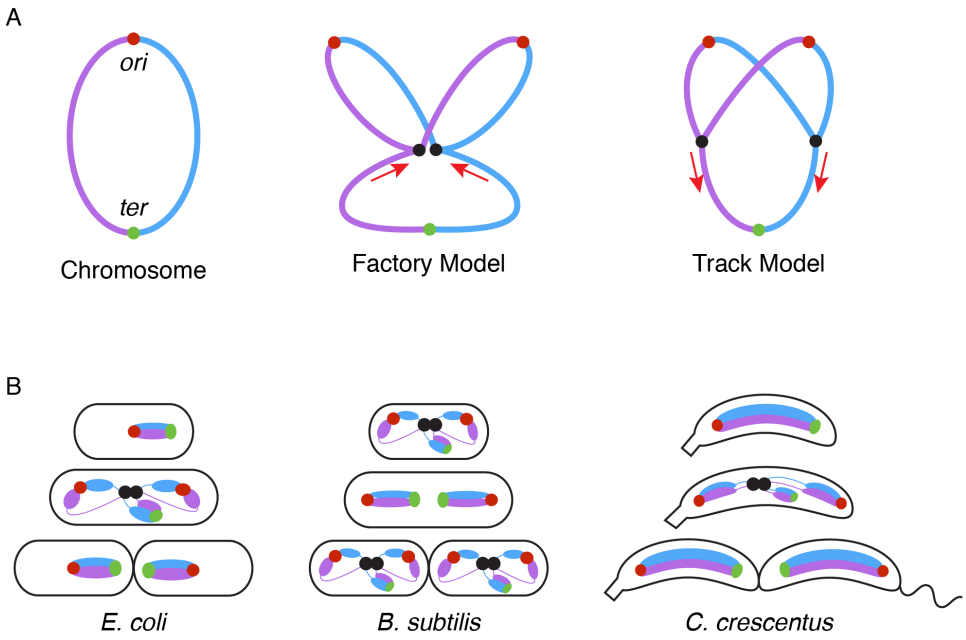

Figure 3.10: A) Schematic diagrams for the Factory and Track models. Most bacteria have a single circular chromosome (left), with a single origin (red dot), positioned roughly opposite the terminus (green dot). The left and right arms of the chromosome are colored pink and green, respectively. In the Factory Model (center), DNA is pulled through the replisomes (black dots) in the direction indicated by the red arrows. In the Track Model (right), replisomes translocate along the template DNA. Red arrows indicate the direction of replisome motion. B) The Factory model is shown for the chromosomal organizations of three model organisms. Decondensed DNA is represented by lines.

factory model for *E. coli* *E. coli*, *B. subtilis*, and *C. crescentus* (Fig. 3.10) [104]

We analyzed the chromosomal locus trajectories to determine whether the translocation during segregation was better modeled by the slow accretion of many small biased steps or the result of a small number of large steps, corresponding to large-scale structural transitions in the chromosome (*i.e.* snaps) that have been reported previously [82, 88]. We observe that all loci segregate rapidly over the first minute, but we have been unable to detect an unambiguous statistical signature of large steps in segregation after this time. As we described previously [19], although large steps can happen, they occur (i) throughout the

cell cycle, (ii) with roughly equal probability in and against the direction of bias and (iii) in only a fraction of cells during rapid translocation. Our observations and analysis of motion of different locus positions in MG1655 are consistent with our previous report: Most loci appear to move by the slow accumulation of biased motion, rather than by large steps. We emphasize that the absence of an enrichment of large steps during segregation does not invalidate the snap-model.

A second unexpected feature of the chromosome dynamics was the universality of the dynamics of all loci at the start of the segregation process. Our previous detailed characterization of the dynamics of *oriC* had revealed that loci feel a directional bias immediately upon the initiation of segregation, an observation that is most consistent with an active mechanism (e.g. [99]). If an active mechanism exists which drives the segregation of the origin, it is natural to expect other loci that are segregated later by a condensation driven mechanism to exhibit different dynamics. Furthermore, one might naïvely expect the forces to be most distinct early in the segregation process when the drift velocities and the forces themselves are strongest [19, 95]. Even if the loci were segregated by an entropic mechanism [31], simulations suggest that the initial segregation of *oriC* should be slower than later loci [99]. Therefore whatever the mechanism, it was our expectation that the initial segregation dynamics of *oriC*-proximal loci should be qualitatively distinct from other loci.

Given these general arguments in favor of qualitatively distinct dynamics after segregation for different regions of the chromosome, we were therefore quite surprised to observed that both the drift velocity, and the step-size distribution were both nearly identical for all loci characterized. Of course, this may be a coincidence: Despite the distinct ParA-dependent mechanism that segregates the origin in *C. crescentus* [105], the average speed only varies by a factor of two [49]. Still the similarity between loci dynamics in *E. coli* is much greater than observed in *C. crescentus*, consistent with the hypothesis that there may be some universal mechanism at work in the segregation of all loci in *E. coli* (*e.g.* [106]). (The mechanism is *universal* in the sense that it action on the loci leads to nearly identical dynamics during the initial steps of segregation.) Kleckner and coworkers have long speculated that there is

an accumulation of stress in the chromosome [107] that is released in an un-snapping event [88, 82]. Since this process could be driven by MukBEF, super-coiling [108, 109] and other processes that lead to longitudinal condensation [110], these mechanisms could explain the universality in the initial segregation dynamics.

Since the greatest mismatch in the drift velocity occurs five or more minutes after segregation initiates, it is interesting to speculate about the possible mechanism for this increased drift velocity for loci proximal to *oriC*. The deletion of *seqA* was observed to have the strongest effect on the bias fives minutes after the initiation of segregation, consistent with a model where late-stage segregation is facilitated by cellular machinery. There is precedence for additional mechanisms that act late in the segregation process in *C. crescentus* [74]. On-the-other-hand, it is important to realize that due to the overall chromosome topography in *E. coli* under the experimental conditions, *oriC* moves the furthest of all loci and therefore it must have a higher drift velocity to reflect this greater travel distance. The mechanism for higher drift velocity long after the initiation of segregation might be nucleoid-structure-induced and quite prosaic from a mechanistic perspective [31].

Chapter 4

PHENOTYPIC COHERENCE OF MUTANTS IN THE PEPTIDOGLYCAN SYNTHESIS PATHWAY

Diversity in cell morphologies is another fascinating aspect of bacterial structure. Although the shapes and growth patterns of bacteria vary widely, the mechanisms driving morphogenesis and the evolution of cellular shape are not well understood. In addition, the morphologies of mutant cells can provide novel insight into the functions of knocked-out genes.

Our broad goal in the quantitative study of cell morphologies is to develop a framework for characterizing the phenotypes of mutant cells. If we observe phenotypic coherence across cellular pathways, we can create a robust bank mapping deleted gene function to cell morphology. Ideally, we could than use this bank of information to work backwards- that is, to predict gene function by comparing the morphologies of mutants lacking genes of *unknown* function to the shapes of mutants lacking genes of *known* function.

To maximize the impact and robustness of this framework, we have designed this project to include essential genes (or genes expressing proteins which are requisite for the cell to survive). Unsurprisingly, these studies are easier with non-essential genes. Since these knock-out mutants are viable, they can be imaged at any time following the transformation for the study and characterization of their morphologies. For the imaging of essential gene knock-outs, however, the cells must be imaged during a window of time between when the transformation is made and the cell shape begins to change, and cell death. Depending on the gene, this window of time can be wide or small, ranging from a couple hours to a day. In section 4.2, we will explore how this challenge also introduces a value opportunity: the ability to observe the process of depletion of an essential gene, and the morphology changes

that occur prior to, or even leading to, cell death.

This ambitious project, with novel strain development and imaging techniques, is broad and challenging. As a result, this study is still in its preliminary stages. In this thesis, I will introduce the details of this project by describing our efforts to develop a bank of phenotypically coherent mutants from a pathway of particular interest to cellular morphology: peptidogylcan (PG) synthesis (for a schematic of PG structure and synthesis see: Fig. 4.1).

The peptidoglycan cell wall provides structure for cellular shape. Bacteria can grow, divide, and change shape, by synthesizing PG in specific regions of the cell. The two main processes of PG synthesis are (i) elongation and (ii) division (Fig. 4.2). During elongation, PG is inserted along the length of the cell to allow the cell to grow along its long axis. During division, PG is synthesized at the center of the cell to form a septum to physically separate the mother cell, dividing it into two daughter cells.

Our goal in this study is to examine the phenotypic pathways associated with peptidoglycan synthesis, in hopes of better understanding the mechanisms driving morphogenesis. By characterizing the morphological effects of knocking out genes in the PG synthesis pathways, we can determine whether there is morphological coherence, or consistent patterns in mutant cell shapes, across pathways. This project is still in the preliminary phase, so this chapter will describe the broad strokes of the project, and an outline of our hopes for where it will lead.

4.1 Peptidoglycan synthesis

4.1.1 PG synthesis pathways in A. baylyi

The production of PG in *A. baylyi* begins with a precursor pathway, common to all cell wall growth. Following the precursor pathway, the production of PG branches into two independent pathways: the elongasome and the divisome.

The set of proteins in the 'elongasome' are responsible for building peptidoglycan all long the cell length, to increase growth along the long axis. A similar study, conducted

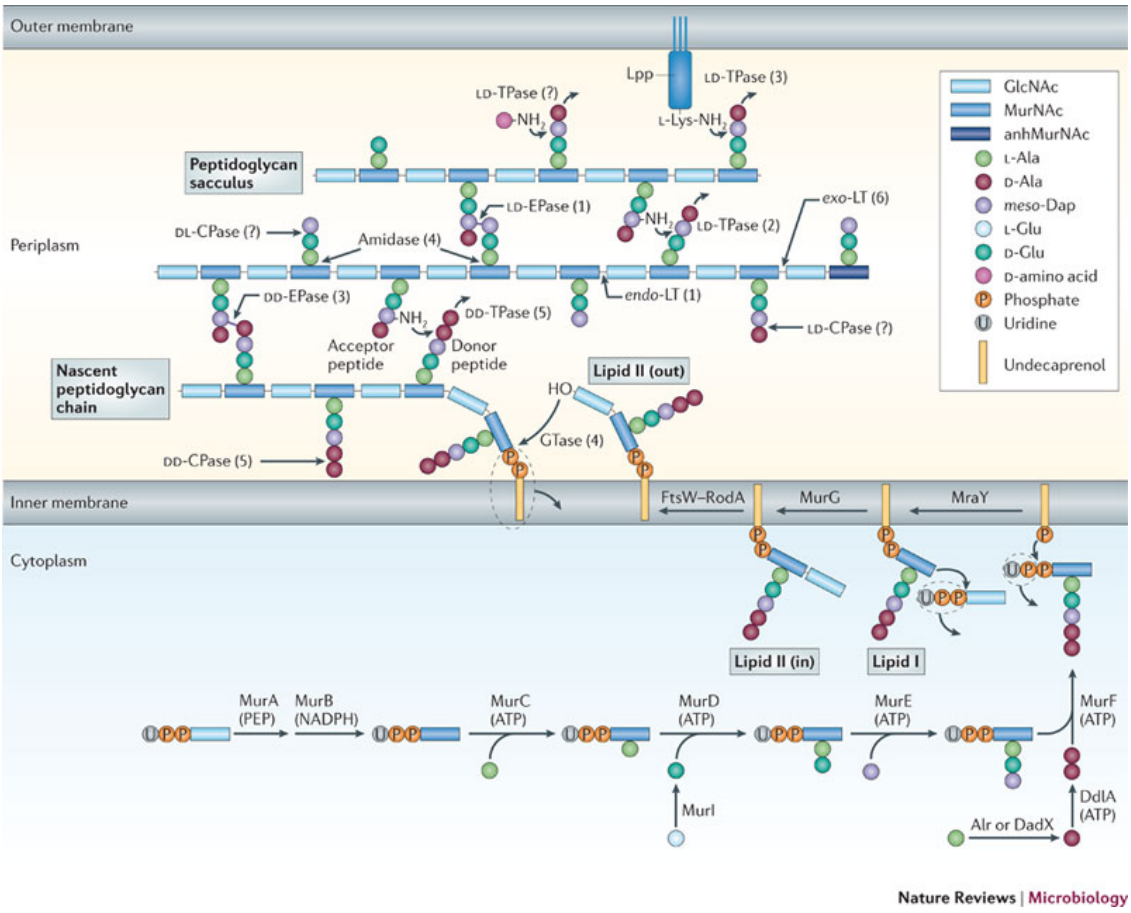

Figure 4.1: **Peptidoglycan synthesis.** This figure is reproduced from *Typas et. al.* [2]. The synthesis and attachment of a new peptidoglycan strand to the existing sacculus, with particular emphasis on the different synthetic and degrading enzymes. Precursors are synthesized in the cytoplasm, linked to the transport lipid (undecaprenyl phosphate) and flipped accross the inner membrane by FtsWRodA. A glycosyltransferase (GTase) catalyses polymerization of a nascent peptidoglycan chain from lipid II precursor at the inner membrane, followed by attachment of the new chain to the sacculus by a DD-transpeptidase (DD-TPase). Peptides are trimmed by DD-, LD- and DL-carboxypeptidases (CPases), and crosslinks are cleaved by the DD- and LD-endopeptidases (EPases). Amidases remove peptides from glycan chains, and exo- or endo-specific lytic transglycosylases (LTs) cleave in the glycan chain to form 1,6-anhydro-N-acetylmuramic acid (anhMurNAc) residues, which are the hallmark of glycan chain ends. LD-TPases are responsible for the formation of LD-crosslinks, the attachment of the major outer-membrane lipoprotein (Lpp), which is anchored in the outer membrane, and the binding of unusual D-amino acids. The number of known *Escherichia coli* enzymes for each group is shown in brackets, but this is probably an underestimate, as even in *E. coli* not all players are known and/or characterized.

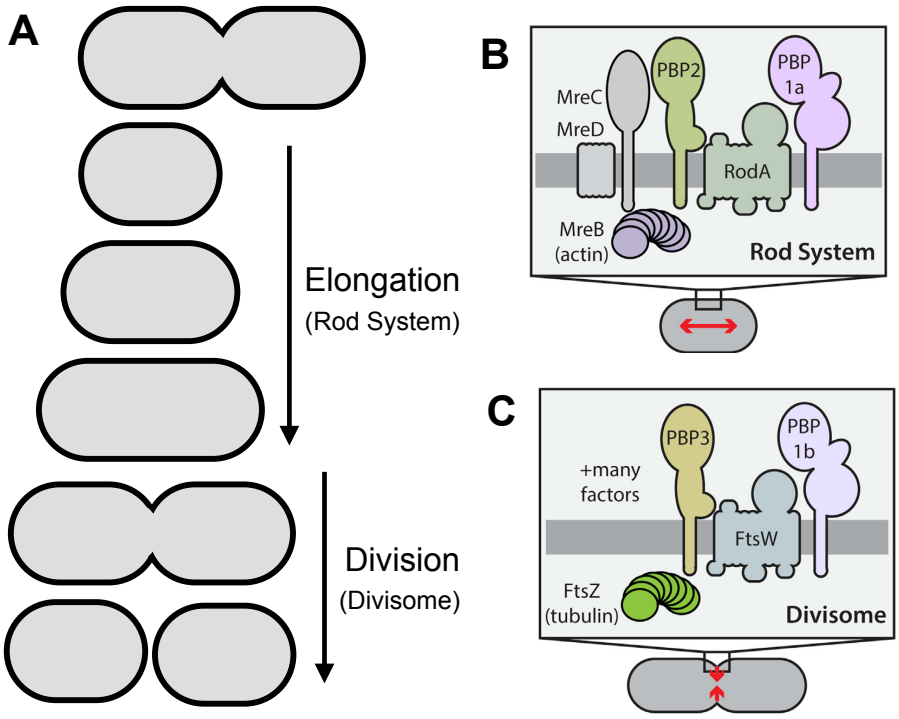

Figure 4.2: **Phases of cell wall synthesis.** (A) Soon after cell birth, the cell grow along its long axis. During this elongation process, PG is synthesized along the cell boundaries. The long cell then divides into two daughter cells; during division, PG is synthesized at the cell center to form a septum and split the mother cell into two daughter cells. (B-C). Schematics from *Cho et. al.* ([3]) highlighting the components of the two main PG synthetic systems in rod-shaped bacteria. Both systems require a dedicated bPBP (PBP2 or PBP3) and a SEDS (shape/elongation/division/sporulation) family protein (RodA or FtsW) for proper function.

in a *Bacillus subtilis* background, show that deletion of proteins in this system causes the production of small, round cells [111]. The proteins in the 'divisome' are responsible for synthesizing the septum which forms at the cell center to divide the cell into two daughter cells. The deletion of genes in this pathway in *B. subtilis* were shown to result in long, snake-like cells [111].

We are particularly interested in PG synthesis for several reasons. Firstly, this process occurs through a complex and essential pathway; this makes it an important cellular process to understand. In addition, many important antiobiotics target this pathway, giving it valuable medical implications. Finally, the development of a cell wall was an important evolutionary development of bacteria, making deletions in this pathway important for our understanding for the development of bacteria as we now known them.

4.1.2 PG synthesis studies: antibiotics

Studies of PG synthesis can have serious implications for the development of new antibiotics and therapies. Some common existing antibiotics, such as Ampicillin (and all penicillins), kill bacteria by inhibiting cell wall synthesis [112] . If we can identify new essential proteins in the PG synthesis pathway, we can potentially develop therapies or drugs that target those proteins to kill pathogenic cells, or manage them in a lab setting.

4.1.3 PG synthesis studies: evolution

The effects of PG synthesis on morphology may tell us an important story about the evolution of bacteria. Early proteo-bacteria possessed no cell wall, and their shapes were therefore less well-regulated. These amorphous cell wall deficient bacteria, called "L-forms," have extremely interesting cellular structure, and grow and divide very differently from modern bacteria [113, 114]. Rather than growing in a polar or uniform way, these cells form "blebs," or (seemingly) randomly placed protrusions growing from the cell body [115]. If the L-form is stable, eventually these blebs will close off from the mother cell, forming a daughter and allowing the population to proliferate.

The exact process by which bacterial cells evolved cell walls remains unknown; L-form bacteria could provide us with clues about general cellular organization of the primitive progenitors of bacteria and especially how they proliferated [113, 115]. We hope that studies of PG synthesis, and especially knockout morphology studies, can give us insight into the mechanisms bridging ancient proteobacteria with current cell-wall protected bacteria.

4.2 Acinetobacter baylyi

We performed these experiments in the bacterium *Acinetobacter baylyi*. There are three main reasons why this background is of particular advantage for these studies: (i) the medical significance of the *Acinetobacter* species, (ii) the experimental advantage of using *A. baylyi* for mutant studies due to its high rate of transformation competency, and (iii) the advantage of using *A. baylyi* for PG studies due to elongation pathway being non-essential.

4.2.1 Close relation to Acinetobacter baumannii

Acinetobacter baumannii bacteria are round, coccobacillus members of the *Acinetobacter* species. They are opportunistic pathogens, often infecting humans with compromised immune systems and causing high infection rates in healthcare settings [116]. These bacteria are difficult to contain, control, and eliminate due to their resistance to antibiotics [117], making them highly pertinent for genetic studies. Many consider *A. baylyi* to pose a high risk for causing an epidemic [118]; their infected hosts have a striking mortality rate of up to 75% [119]. A common way to kill bacteria is to target essential genes: those genes which are required for cell survival. By determining the function of essential genes in *A. baumannii*, researchers hope to find strategies to better combat *A. baumannii* infections.

A. baylyi are similarly round, coccobacillus members of the *Acinetobacter* species. They are closely related to *A. baumannii*, with 82% of *A. baumannii* genes having orthologs in *A. baylyi* [120]. However, *A. baylyi* are sensitive to antibiotics, making them easier to control and to study in the lab. It is hoped that, by studying these more easily manipulated species with many common genes, a therapy for *A. baumannii* infections can be developed. This

highly motivates the use of *A. baylyi* in any bacterial genetic study.

4.2.2 Efficient transformation: real-time visualization of morphological knock-out effects

Transformations in bacteria are typically highly inefficient (0.1%) [121]. As a result, rare transformed cells usually require a genetic selection on differential identifying medium which only allows transformed recipients to form colonies. A single colony can then be isolated, then grown in liquid before the cells can be imaged. Because of the length of this process to isolate the rare transformation recipients, only stable mutants can be successfully imaged.

For our studies here, *A. baylyi* offers a distinct advantage: this species has high natural competence [122]. On average, our transformations yield a notable 10% efficiency rate without any needed chemical, temperature or electrical shock for transformation success [122]. Because of this, we can immediate image cells which have been freshly transformed.

This is a remarkable advantage. In the case of an essential gene deletion, we can observed cell morphology as the cell transitions from healthy, to depleted for an essential function, to death; this process is outlined in detail in the Materials and Methods section of this chapter. The ability to watch this process in real time from transformation to cell death is a novel and remarkable technique, and it is impossible with bacteria that need to be streaked on plates for selection prior to imaging.

Our goal in this study is to knock out genes in the PG synthesis pathways, and watch the progression of cell morphology as cells lose proteins needed for cell wall synthesis. In this way, we hope to determine whether there exists a phenotypic coherence in morphology across PG synthesis pathways.

4.2.3 Elongation non-essentiality: simpler PG analysis

While the elongation pathway is essential in many rod-shaped bacteria (including *E. coli*), it is not essential for *A. baylyi*. This causes *A. baylyi* to be an attractive background for analyses further along in our study, in which we can characterize division mutants in an

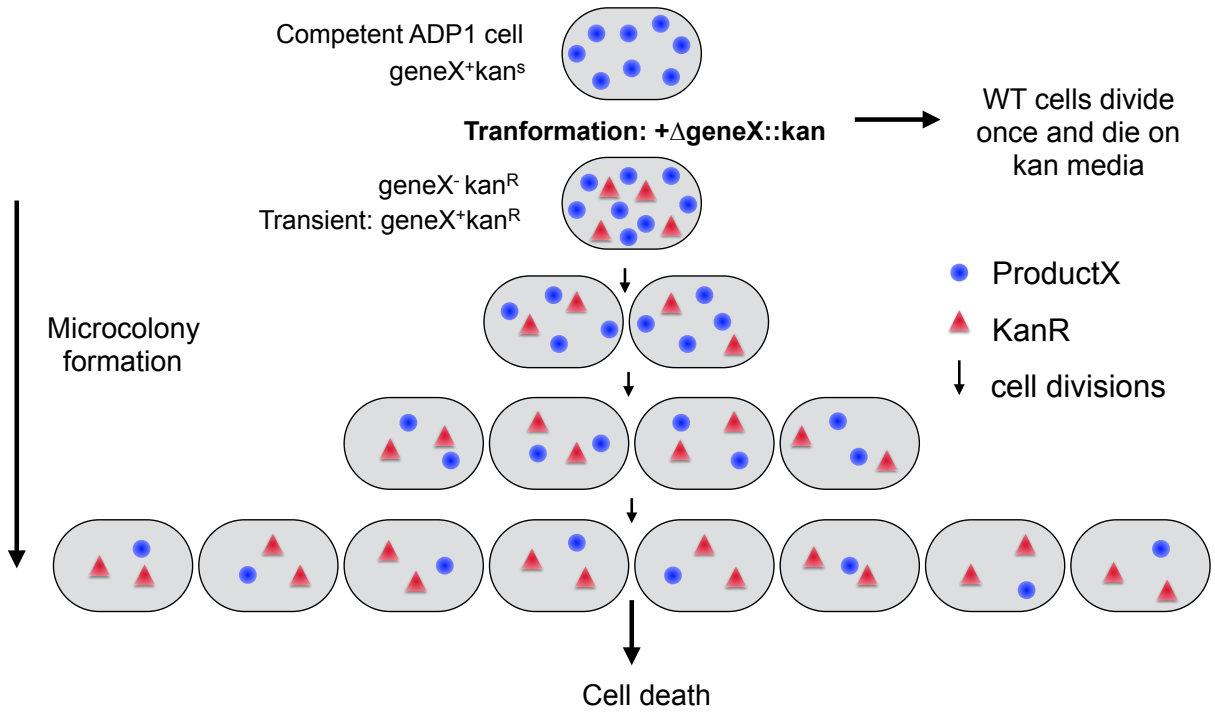

Figure 4.3: **Depletion of deleted genes for microcolony analysis.** We begin with a competent WT *A. baylyi* cell; it possesses geneX (geneX^{+}), and is sensitive to kanamycin (kan^{s}). A DNA fragment carrying a deletion of geneX (geneX^{-}) and addition of kanamycin resistance (kan^{R}) is inserted. The cells are then innoculated on a kan pad; any WT cells (cells which failed to receive and incorporate the insert) will divide once or twice, then die. Successful transformants will now be lacking geneX, but still carry any remaining copies of ProductX (blue circles) that were already produced in the cell. Each time this cell divides, its daughter cells will possess one half the remaining Product X. Since transformants are kan^{+} they each produce new product for kan resistance (red triangles). Eventually, the concentration of essential product ProductX will be too low for the cell to survive, and the daughter cells will die. This ends the proliferation of the mutant, but leaves a microcolony of mutant cells.

elongation-disabled background, so that all precursors are funneled solely into the division pathway.

4.3 Materials and Methods

The Manoil lab has developed a robust method for both determining the genes in bacterial genomes which are essential for cell life, and for creating large libraries bacterial mutants [123]. In addition, J. Bailey has conducted extensive studies to determine the optimum media for growth and proliferation of delicate PG mutants. In this project, we have paired their work with our *Supersegger* and `gateTool` software to achieve the imaging and characterization of myriad PG synthesis mutants [1, 16].

4.3.1 Strain construction

The Manoil lab created a library of inserts for the deletion of most *A. baylyi* genes during the development of their Tn-seq mutant library. The procedure for insert development employs a Tn-seq method based on the creation and amplification of single-strand loops carrying transposon junction sequences ("Tn-seq circle method") with an antibiotic resistance cassette. This technique can be used with virtually any transposon, making the method adaptable for all mutant strain construction [123].

4.3.2 Culture growth

Wild-type *A. baylyi* cells were grown in succinate media overnight, then inoculated with selected kna-resistant-marked inserts for 2 hours at 30 degrees prior to imaging. They were then diluted to OD 0.2.

All experiments in this study are conducted using succinate media. For initial tests, succinate media with no supplements except kanamycin for mutant selection was used. In later experiments, osmo-protective media was made by supplementing the succinate media with 0.5 M sucrose and 10 mM magnesium sulfate.

4.3.3 Time-lapse imaging and analysis

Cells were inoculated onto agarose pads containing succinate media and Km for the selective growth of transformed cells with the deletion insert. Cells were imaged using brightfield illumination every 2 min for 10 hours, often 10-20 *xy* positions. These images were then used to generate time-lapse videos of micro-colony development.

A visual outline of the process for depletion and micro-colony development is provided in Fig. 4.3. We begin with a competent WT *A. baylyi* cell; it possesses *geneX* (geneX$^+$), and is sensitive to kanamycin (kan^s). A DNA fragment carrying a deletion of *geneX* (geneX$^-$) and addition of kanamycin resistance (kan^R) is inserted. The cells are then innoculated on a kan pad; any WT cells (cells which failed to receive and incorporate the insert) will divide once or twice, then die. Successful transformants will now be lacking *geneX*, but still carry any remaining protein molecules of ProductX (blue circles) that were already expressed in the cell from the gene. Each time this cell divides, its daughter cells will possess one half the remaining Product X. Since transformants are kan$^+$ they each produce new product for kan resistance (red triangles). Eventually, the concentration of essential product ProductX will be too low for the cell to survive, and the daughter cells will die. This ends the proliferation of the mutant, but leaves a microcolony of mutant cells.

When possible, images were segmented using *SuperSegger* and analyzed using the `gateTool` image cytometry software.

4.4 Results

This project is still in its early stages of development, so I will only provide preliminary results here. I will introduce one mutant from each section of the PG synthesis pathways: (i) a precursor, (ii) an elongation mutant and (iii) a division mutant. For each pathway, I provide time-lapse images of the growth of a micro-colony, then describe the phenotypic patterns established so far through these videos and snapshot imaging.

4.4.1 PG synthesis precursor mutants

The early precursor MurA is a critical enzyme in the PG biosynthesis pathway, which can be inhibited by the phosphonic antibiotic fosfomycin [124]. Using the imaging procedure described in the last section, we followed the depletion of MurA in *A. baylyi* from early stages to cell death, observing the time-dependent changes in morphology of MurA-depleted cells. A selection of frames from a representative time-lapse video is provided in Fig. 4.4.

The changes documented in these time-lapse videos are fascinating; we begin with healthy, round cells with wildtype morphology. After a couple hours of very slow growth, the cells start to balloon. As they grow in all directions, they lose their rod-shape phentoype and become much more amorphous. The morphology and division dynamics of these giant cells is highly reminiscent of L-form cells.

At times, these giant cells attempt to divide. Unlike wildtype cells, they do not divide evenly through constriction at a well-defined cell center. Instead, the cells divide through blebbing; one region of the giant cell starts to pull away from the rest, forming a protrusion. Rarely, this bleb will successfully divide from the rest of the cell; the completion of division is often accompanied by a rapid translocation of the daughter cell away from the mother cell. Much more often, however, the mother (or daughter) cells lyse directly after the cell's attempt to divide.

Without the cell wall, the dividing cells lack the stability to maintain their structure. We believe that the presence of additional structural support of nearby cells or firm media acting as scaffolding may help protect the cells and promote successful division. While non-lysing daughter cell are rare, we find that they are most often the result of the division of a cell surrounded by other cells.

To-date, we have only produced timelapse videos of the *murA* mutant from the precursor pathway. However, we have completed snapshot imaging for several other mutants in the precursor pathway of PG synthesis. A compilation of representative snapshots from these mutants is provided in Figure 4.5. All mutants imaged from this pathway exhibit similar

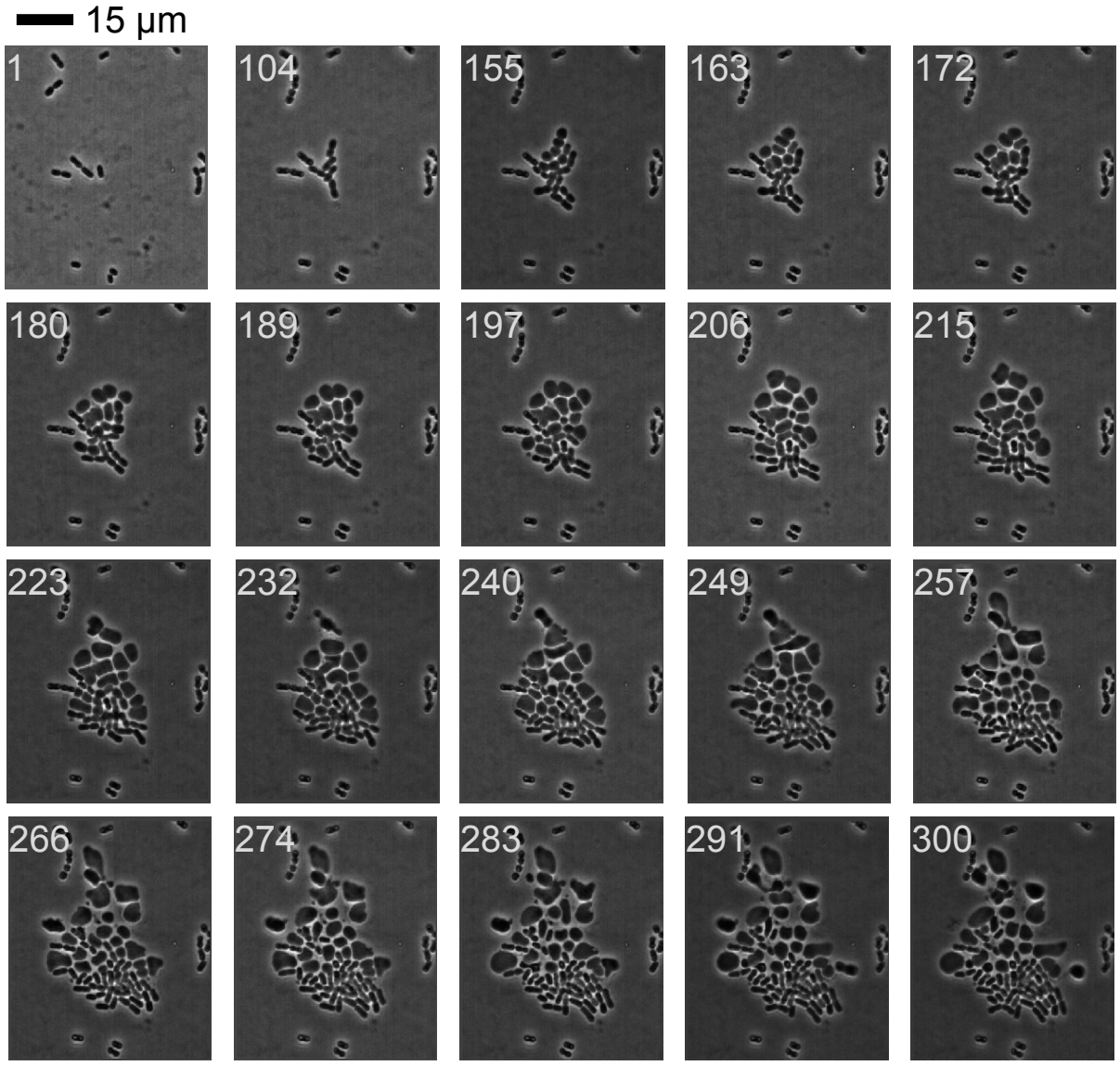

Figure 4.4: **Selected frames from ΔmurA timelapse imaging.**

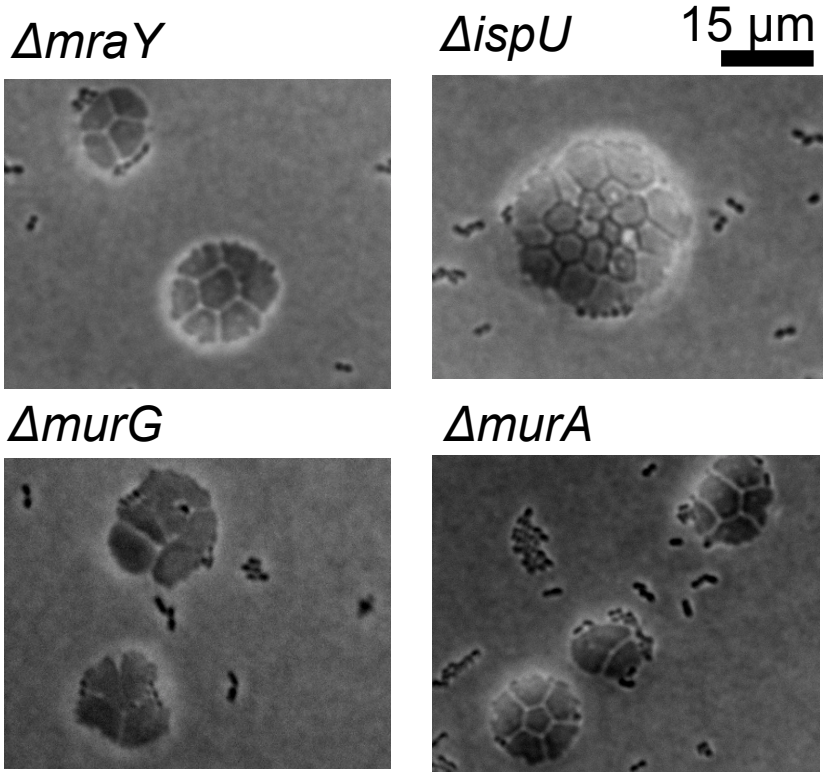

Figure 4.5: **Snapshot images of several mutants depleted in effectors from the precursor pathway for PG synthesis.**

morphologies; this suggests a coherent phenotype of giant, L-form-like cells for mutants in the precursor PG synthesis pathway.

4.4.2 Elongation mutants

Deletions of genes in the elongation pathway singly or in combinations are viable in *A. baylyi*. For this example, we will discuss a triple mutant with three of the key effectors in cell elongation PG synthesis deleted: Δ*pbpA* Δ*rodA* Δ*ponA* (for simplicity, we will refer to this strain at Δelong). The time-lapse imaging of these cells is therefore not a depletion study, but the imaging of a viable mutant strain. In Fig. 4.6, we provide a selection of frames from a representative time-lapse video of this elongation mutant.

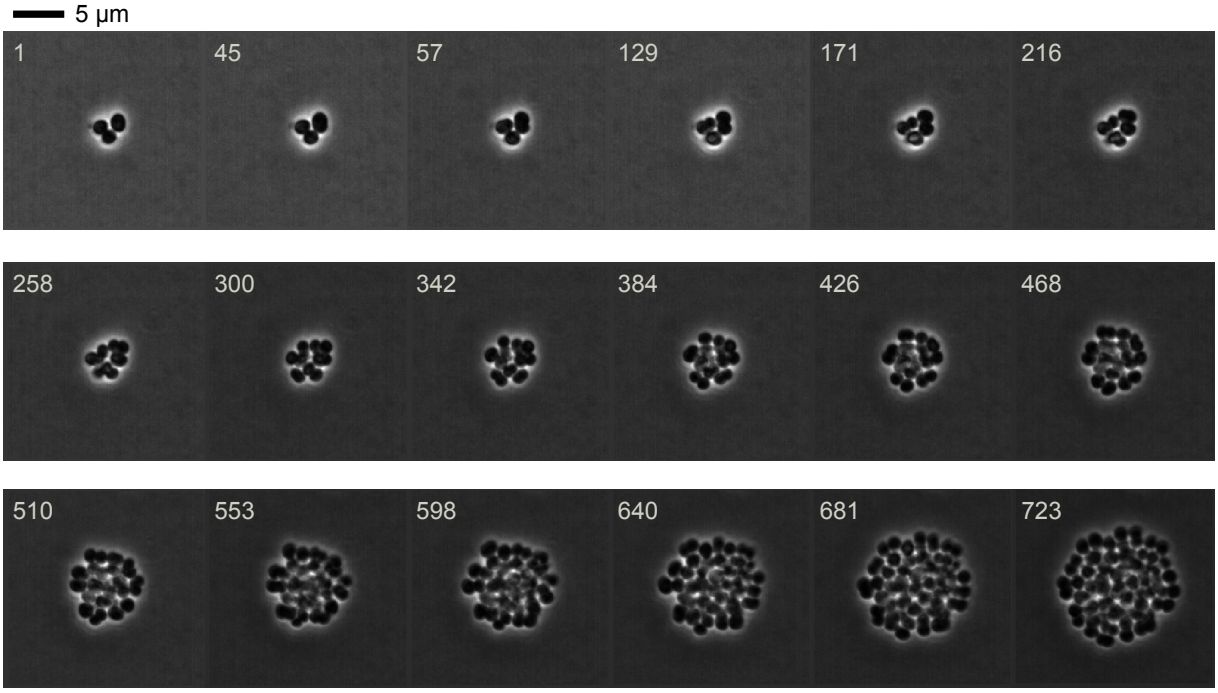

Figure 4.6: **Selected frames from Δelongation timelapse imaging.**

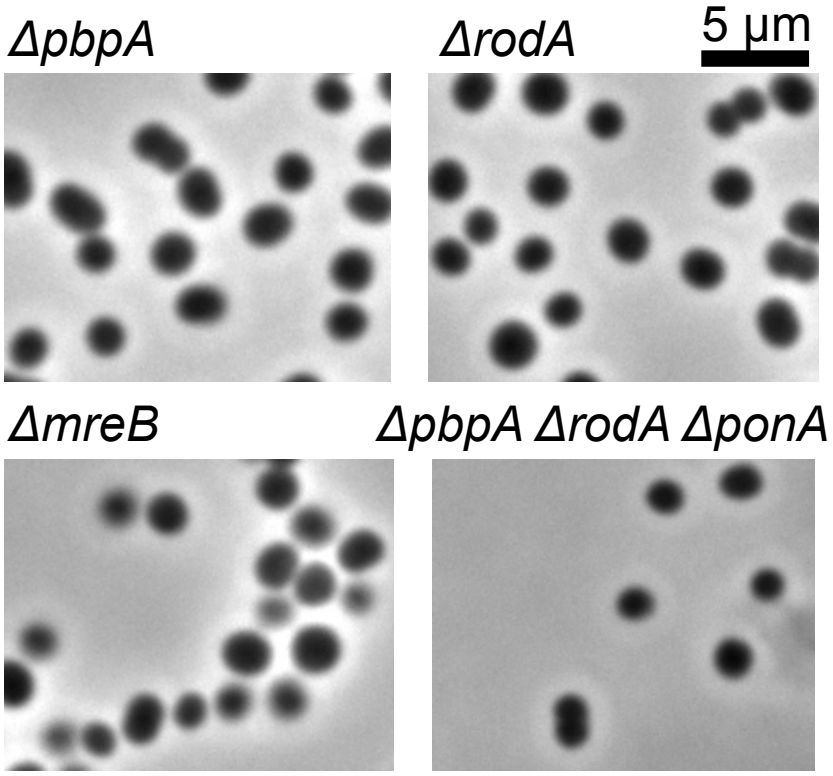

Figure 4.7: **Snapshot images of several mutants depleted in effectors from the elongation pathway for PG synthesis.**

Preliminary efforts by J. Bailey have verified that mutants from across this pathway produce cells with similar morphologies (Fig. 4.7); this suggests that there is a coherent phenotype of short, round cells for deletions across the elongation pathway.

From these frames, we can see that this viable, elongation-minus strain of *A. baylyi* has the expected short, round phenotype. We have only produced timelapse videos of the triple mutant from the elongation pathway, however we have conducted preliminary snapshot imaging several single mutants in the elongation pathway of PG synthesis. A compilation of representative snapshots from these mutants is provided in Figure 4.7. All mutants imaged from this pathway exhibit similar morphologies; this suggests a coherent phenotype of small, near-spherical cells for mutants in the elongation PG synthesis pathway.

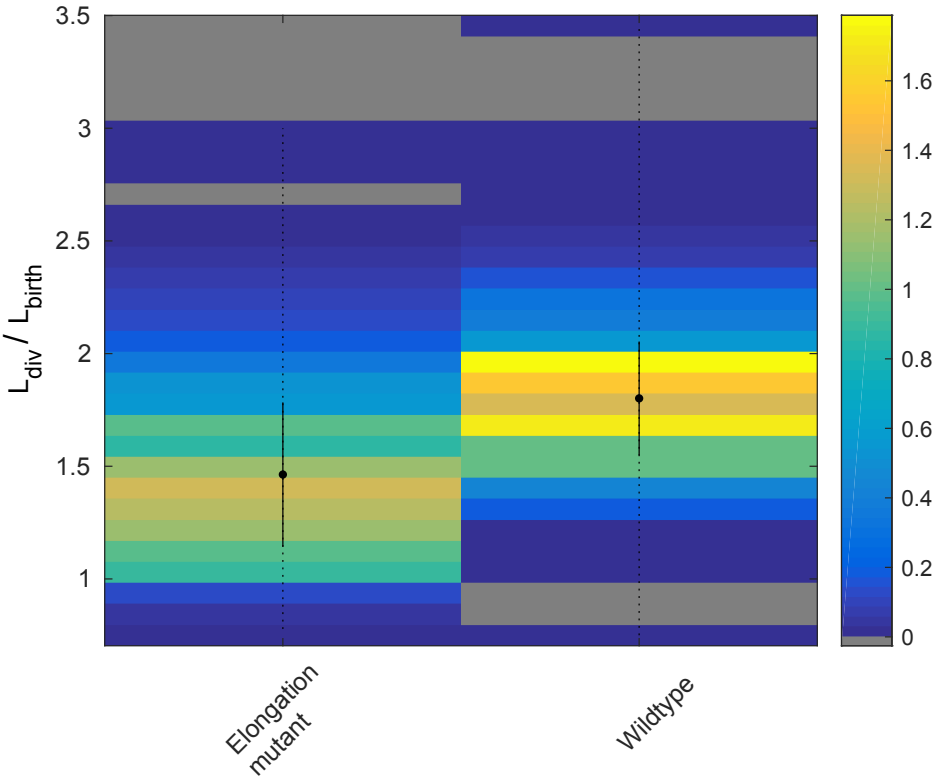

Figure 4.8: **Comparison of growth ratios for wildtype and Δelongation.**

Since the process for elongating cells has been disabled, a short cell morphology is fairly unsurprising. However, since the cells divide and continuously maintain a growing colony, we wondered whether the mere synthesis of the septum generated enough PG to maintain a generation-consistent cell length.

Because these cells are small and regular, they are easily segmented by *SuperSegger*. We therefore used `gateTool` to conduct a short analysis of dependence of birth length on birth time. Using `gateTool`, we isolated cells that were imaged for a complete cell cycle. We then compared the distribution of division length to birth length ratios for all Δelong and wildtype cells (Fig. 4.8). We expect that for a sustained population of cells, the length just before division should be twice twice the length at birth, or $L_{div}/L_{birth} \approx 2$. As seen in Fig. 4.8, the ratio is just shy of two ($\langle L_{div}/L_{birth} \rangle \approx 1.80$); for the Δelong strain, however, the ratio is significantly lower ($\langle L_{div}/L_{birth} \rangle \approx 1.46$).

Naively, this would suggest that each generation of cells is successively shorter (by 25%). However, we should note that even the wildtype cells do not quite double prior to complete cell division. We must keep in mind that additional PG is synthesized to complete the division of the cell between the final frame of mother cell life. For a long cell, the increase in length provided through the completions of septum synthesis is negligible ($< 10\% L_{birth}$); for a very short, Δ*elongation* cell, however, this additional PG would theoretically provide most of the cells growth. Based on this fact and Fig. 4.8, we can theorize 50% of the length increase occurs during septation prior to *SuperSegger* declaring the cell as completed divided. Followed that frame, the additional 50% length increase occurs through completion of septation. This would account for the complete doubling of the cell resulting from the completion of division.

4.4.3 Division mutants

Since the division pathway of PG synthesis is essential is *A. baylyi*, the time-lapse imaging of mutants lacking genes in this pathway are depletion studies. In this example, we consider FtsN: an essential cell division protein that activates septal PG synthesis and constriction

of the cell. Selected frames from a representative time-lapse video are providied in Fig. 4.9. As in the MurA depletion study, the cells begin with a wildtype phenotype. After a couple hours, however, they begin to grow rapidly along their long axes. The cells continue to grow until they are radically longer than their wildtype counterparts - they can grow up to 50x as long as WT cells! This phenotype is consistent with observations in other organisms.

While we have only produced timelapse videos of one division mutant (FtsN), we have conducted preliminary snapshot imaging several other mutants in the division pathway of PG synthesis. A compilation of representative snapshots from these mutants is provided in Figure 4.10. All mutants imaged from this pathway exhibit similar morphologies; this suggests a coherent phenotype of extremely long, snake-like cells for mutants in the division PG synthesis pathway.

4.5 Discussion and Future Directions

4.5.1 Phenotypic coherence across pathways

Our preliminary imaging has suggested coherence in large-scale features across precursor, elongation and division pathways of PG synthesis. For deletions in the elongation pathway, we observed short, round cell morphologies. Deletions in the division pathway consistently produce long, snake-like cells. The PG synthesis precursor mutants that we imaged also displayed a consistent large-scale morphological pattern across pathway deletions: all cells produced giant cells reminiscent of L-forms, which divide through blebbing, often resulting in cell lysis. Based on these results, we are now determining the future directions for this project- in the next section I will describe these plans.

4.5.2 Further studies

There is much work left to be done in this study, and this exciting project could be built-upon in several different directions, depending on upcoming results. Before we can fully develop our conclusions from this study, we will need to image more diverse mutants from each of

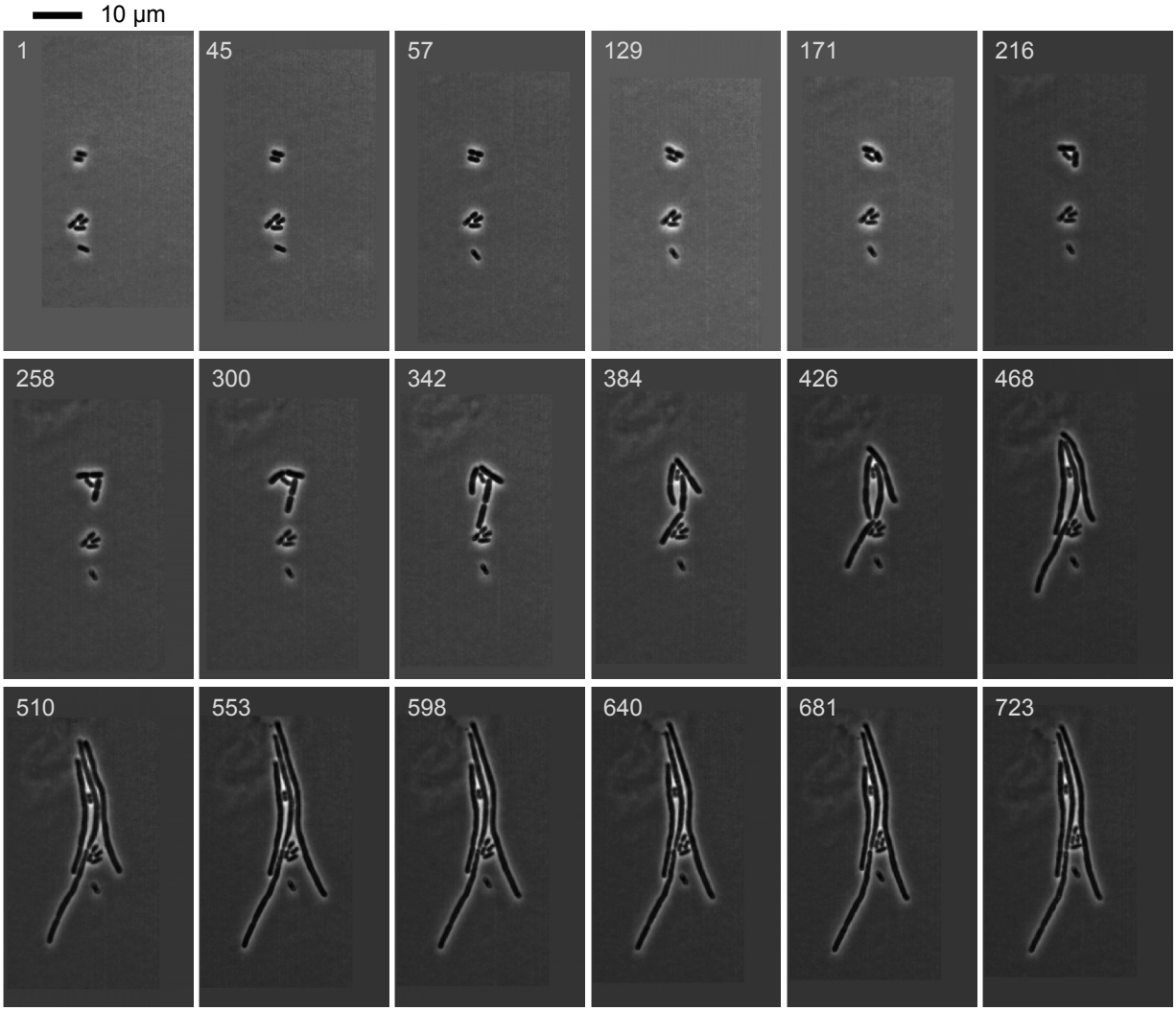

Figure 4.9: **Selected frames from ΔftsN timelapse imaging.**

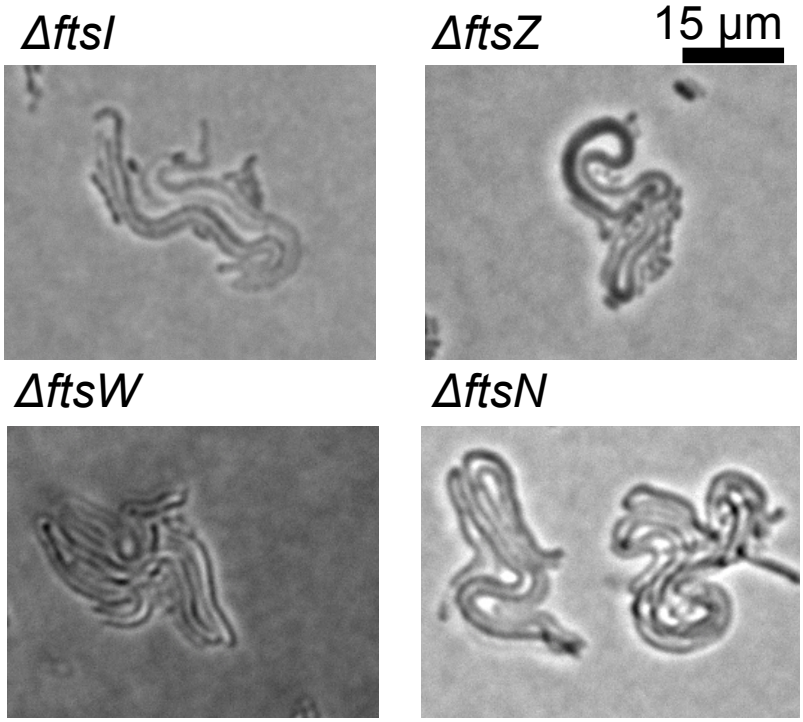

Figure 4.10: **Snapshot images of several mutants depleted in components from the division pathway for PG synthesis.**

the pathways.

Additionally, we hope to increase daughter cell stability, to better observe the long-term proliferation of these micro-colonies. This may be achieved through adjustments to the media used; in particular, current control experiments are testing media optimization through variations in the concentrations of Mg, sucrose and kanamycin.

Our hope in this project is to apply the analysis tools outlined in the second chapter of this thesis to quantitate mutant morphology and dynamics, and mathematically analyze the coherence mutant cell morphologies. Currently, the *SuperSegger* segmentation algorithms very accurately identify smaller, round and rod-shaped bacteria (i.e. those with morphologies close to wild-type). However, our efforts to segment giant cells in a colony have been unsuccessful; we are able to identify all regions that are completely isolated (such as the outer edges of a micro-colony), but segmentation of the colony into individual giant cells has been a major challenge. We are hopeful that in our future work, either adjustments to the segmentation code, or methods for analyzing micro-colony shapes will allow for a more quantitative analysis of mutant morphologies in the PG synthesis pathways.

Chapter 5
CONCLUSIONS AND FUTURE WORK

While major advances in the last century have provided novel insight into the structure of bacterial cells, many questions about the mechanisms driving structure remain. As a computational biophysics lab, we develop high-throughput techniques for probing that structure by either working around, or taking advantage of, cell-to-cell variations. In this chapter, I will summarize the results of this thesis, and describe how the experiments on chromosomal structure and morphological structure could be expanded on.

5.1 Chromosomal structure and dynamics in E. coli

In this project, we studied the chromosomal structure of *E. coli* in two major parts: (i) we developed a cell-cycle dependent model for chromosome structure in MG1655, and (ii) we used this mapping to probe the as-yet uncharacterized mechanism for chromosome segregation. I will now summarize each of these experiments, and how they could be further developed in the future.

5.1.1 Cell-cycle dependent chromosome dynamics

We developed time-lapsed mapping of sub-cellular localization for seven chromosomal loci. This was used to create a full-chromosome mapping of cell-cycle dependent dynamics. This mapping revealed a folded-over *ori-ter* structure for the MG1655 chromosome, conflicting with *left-right* structure observed in AB1157. Thus, models for cell-cycle dependent chromosome dynamics remain incoherent across growth conditions and strains. An important driving force in future work studying chromosomal structure will be the resolution of these inconsistencies, or an understanding of what causes them. An important next step in ad-

vancing this study will be to develop a two-colored strain of MG1655 to allow us to look at both chromosomal arms simultaneously. Directly observing the colocalization of loci on either arm is an important step in bolstering the *ori-ter* model for chromosome organization.

Earlier in this thesis I discussed how studies of chromosome structure dependence on growth stage could help link these disparate models. We observed *oriC* localization in stationary cells both imaged on PBS and media pads. This analysis reveals that 6% of stationary cells on PBS pads possessed sister chromosomes which segregated apart from one another. When we compared the mean-squared-displacements for segregating chromosomes of stationary cells imaged on PBS and media pads, we had the surprising result that the cells with no nutrients (PBS pads) exhibited a higher MSD scaling coefficient than the cells imaged with nutrients, suggesting a more directed motion. These experiments are still in their initial stages, and there is much work left to be done to continue that line of study. In the future, our lab will compare the MSD and diffusion coefficient results obtained for the chromosomal loci in these stationary cells to that of large RNA-protein complexes. This will allow us to determine whether the cytoplasm exhibits a crystalline state or a fluid one. This will help us to isolate and characterize the growth stage-dependent structure of the chromosome. Additionally, we will repeat this study in all seven of the GFP-labeled locus strains, to reconstruct a mapping of chromosome structure in stationary cells for comparison to the growing cell chromosome map.

Our model revealed that all seven loci moved towards the cell center and split into two resolvable sister loci near midcell. This helped to shed light on another debated aspect of chromosomal dynamics, as this result is consistent with the factory model of replication. We are in the process of further addressing the replication factory model by considering a two-colored strain with both *oriC* and dnaN labeled. By repeating our experiments in this strain, we hope to determine whether the stationary cells that exhibited segregation are replicating. If they are, studying the relative localization of the forks in stationary cells would provide exciting new insight into both bacterial bet-hedging and the replication factory model.

5.1.2 Chromosome segregation mechanism

Using the data for cell-cycle dependent modeling of chromosomal dynamics, we developed a quantitative study of chromosome segregation dynamics. This study suggested that all seven studied chromosomal loci segregate with a universal initial drift velocity. This suggested that, whatever the unknown mechanism for segregation in *E. coli*, that mechanism is effecting the initial segregation of all loci the same way. We also observed that the deletion of *seqA*, a sequestering protein and candidate for segregation involvement, decreased late-segregation displacement of loci but had no effect on the initial segregation velocities.

The observance of universal segregation dynamics is both exciting and frustrating. On one hand, it is intriguing and unexpected that *oriC* would exhibit the same segregation dynamics as later-segregated loci. On the other hand, it is easier to develop hypotheses about a unique behavior than a universal one. One possibility for future studies, though, is to repeat the knockout experiment conducted in a *seqA* background with other knockout strains. By screening for a major impact on initial segregation velocities, we could search a protein heavily involved in the initiating the segregation process. This is, however, is painfully reminiscent of searching for a needle in a haystack.

5.2 Morphological structure of A. baylyi

In the last chapter, I outlined an ongoing study probing the coherence of morphological features associated with PG synthesis mutants. As described previously, this experiment is still in the data collection phase, so there is a great deal of work left to be done. Here I will recapitulate the key results and future plans for this sprawling and exciting project.

5.2.1 Time-lapse imaging of A. baylyi PG mutants

We hope to repeat time-lapse imaging of microcolony phenotypes during protein depletion for a larger variety of mutants from both the elongation and the division pathways. The more morphological characterizations of mutants we can generate, the better we can establish the

coherence or sub-classes of cellular shapes in each pathway.

We also hope to further this project by providing greater structural support unstable L-form-like giant cells. As described in the previous chapter, most successful giant cell divisions were observed in cells that were in contact with neighboring cells. In addition, imaging on pads with a higher agarose concentration also produce a higher number of successful giant cell divisions. Together, these observations suggest to us that the presence of surrounding structures to act as a scaffolding to support divisions that don't result in immediate cell lysis. Preliminary tests have also suggested that some media can be more protective for giant cells, decreasing the osmotic forces that drive giant cell lysis. As a result, we have been experimenting with media supplements, to bolster micro-colony development.

5.2.2 Quantitative analysis of PG mutant cells and colonies

We hope to support our hypothesis of phenotypic coherence across PG synthesis pathways with quantitative analysis of mutant cell morphologies and dynamics. In Chapter Two, I described in detail the high-throughput image cytometry tool developed by our lab; the *Clist* and `gateTool` provide support for time-dependent analysis of such cell features as long and short axis lengths, area, growth rate, and eccentricity. We hope to apply our `gateTool` analysis techniques to PG synthesis mutants. This could lead to exciting characterization of major time-dependent transitions in cell morphology, and allow for direct quantitative comparison of cellular features in search of morphological coherence.

As discussed in the previous chapter, our ability to progress with quantitative `gateTool` analysis is limited by *SuperSegger*'s ability to segment the cells. Currently, *SuperSegger* can segment smaller, rod-shaped or round cells with extreme accuracy; the giant cells produced through deletions in the division pathway, however, are currently very difficult for *SuperSegger* to identify. While *Supersegger* accurately locates the boundaries of isolated objects, such as the outer edges of a micro-colony or the edges of a cell with no neighbors. It is our hope to either determine a method for using this ability to complete our quantitative analysis, or to improve the *SuperSegger* segmentation algorithms to be robust enough to segment

giant cells. The former is more feasible, but requires us to sacrifice the single-cell tracking of depletion ; the latter is a major challenge but would enable the extremely exciting analysis of single cell time-dependent morphology changes in depletion experiments.

5.2.3 Genes of unknown function

Our analysis of the PG synthesis pathways is one branch of a larger effort. The Manoil and Wiggins labs are collaborating to gain insight into how the depletion imaging of *A. baylyi* can provide insight into bacterial structure and functions. One of the major goals of this project is create a library of mutant morphologies for a large set of *A. baylyi* deletions. First, we hope to create a library of morphologies for genes of known function, in hopes of establishing coherent patterns across pathways. The PG synthesis pathway is one example of this process. Next, we hope to develop a library of cell morphologies for genes of *unknown* function. It is our hope that through high-throughput and sensitive image cytometry analysis, we can map the morphology of any particular mutant of unknown gene deletion to a class of morphologies for genes whose function or pathway is known. In this way, we could develop a predictive model for determining unknown gene function based on cell morphology. This model would be powerful in developing our knowledge of cellular structure, and could provide insight into the function of homologous unknown genes in the pathenogenic relative *A. baumannii*. Understanding the function of unknown genes could therefore be used to help develop antibiotics that could successfully target and manage this highly resilient bacterium (e.x. [111]).

5.3 Concluding notes on cell-to-cell variations

In this thesis, I introduced our high-throughput image cytometry tool developed for quantitative studies of bacteria. While image analysis is possible without automated software, cell-to-cell variation causes a large amount of noise in biological data, necessitating large datasets to decrease error. Our image cytometry software, `gateTool` was designed to streamline the analysis of large data sets to enable the quantitation of many cells to harness this cell-to-cell

variation.

By enabling the fast and efficient analysis of large datasets, `gateTool` facilitates high-throughput image cytometry and bolsters the researchers ability to minimize the effects of cell-to-cell noise in their results. By this perspective, `gateTool` treats cell-to-cell variations as a 'bug' in our experiments, and allows us to get around this challenge.

Conversely, `gateTool` treats cell-to-cell variation as a feature, or beneficial aspect, of our experiments. The diversity in values for the same cellular variable (i.e. length, mean fluorescence, etc) across a population allow us to use a single population as a grounds for testing hypotheses. For example, if we are interested in testing whether a covarying relationship exists between length and fluorescence, could compare the lengths of cells from individually engineered strains with varying fluorescence expression levels. With `gateTool`, we can approach this questions more efficiently, by analyzing the relationship between length and fluorescence within a single population. In this way, the variations within a single strain can offer valuable insight into exciting biological phenomena.

Following the introduction of the `gateTool` imagery software, I provided examples of how cellular structure can be probed. In Chapter Three, I demonstrated how we harnessed variations within a population to study DNA structure and dynamics on the sub-cellular level, and investigated the segregation mechanisms controlling DNA dynamics following replication. In Chapter Four, I described our plans and preliminary data for analyzing morphological traits of various PG synthesis mutants, and studying how their shapes are controlled by genes in the cell wall synthesis pathways.

In conclusion, my graduate work has shed light on two major mechanisms by which the cell maintains control over its structure, treating cell-to-cell variations as both a feature and a bug in high-throughput quantitative biophysics.

BIBLIOGRAPHY

[1] Stella Stylianidou, Connor Brennan, Silas B Nissen, Nathan J Kuwada, and Paul A Wiggins. Supersegger: robust image segmentation, analysis and lineage tracking of bacterial cells. *Mol Microbiol*, 102, Aug 2016.

[2] Banzhaf M. Gross C.A. Vollmer W. Typas, A. From the regulation of pg synthesis to bacterial growth and morphology. *Nat. Rev. Microbiology*, 10:123–136, 2011.

[3] Uehara T. Cho, H. and T.G. Bernhardt. Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery. *Cell*, 159:1300–1311, 2014.

[4] Stella Stylianidou, Nathan J Kuwada, and Paul A Wiggins. Cytoplasmic dynamics reveals two modes of nucleoid-dependent mobility. *Biophys J*, 107(11):2684–92, Dec 2014.

[5] J. Seger and H.J. Brockmann. What is bet-hedging? *Oxford Surveys in Evolutionary Biology*, pages 182–211, 1987.

[6] C. Robinow and E. Kellenberger. The bacterial nucleoid revisited. *Microbiol. Rev.*, 58:211–232, 1994.

[7] Belmont A.S. Robinett C.C. Straight, A.F. and A.W. Murray. Gfp tagging of budding yeast chromosomes reveals that proteinprotein interactions can mediate sister chromatid cohesion. *Curr. Biol.*, 6:15991608, 1996.

[8] Ciosk R. Michaelis, C. and K. Nasmyth. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, 91:35–45, 1997.

[9] C.C. Robinett. In vivo localization of dna sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.*, 135:1685–1700, 1996.

[10] Sergueev K. Li, Y. and S. Austin. The segregation of the *Escherichia coli* origin and terminus of replication. *Mol. Microbiol.*, 46:985–996, 2002.

[11] A A Teleman, P L Graumann, D C Lin, A D Grossman, and R Losick. Chromosome arrangement within a bacterium. *Curr Biol*, 8(20):1102–9, Oct 1998.

[12] V. F. Holmes and N. R Cozzarelli. Closing the ring: links between smc proteins and chromosome partitioning, condensation, and supercoiling. *Proc. Natl Acad. Sci. USA*, 97:1322–1324, 2000.

[13] N.J. Trun and J.F. Marko. Architecture of a bacterial chromosome. *Am. Soc. Microbiol. News*, 64:276–283, 1998.

[14] Frits Zernike. Phase contrast, a new method for the microscopic observation of transparent objects. *Physica*, 9.

[15] S. Bradbury and P. Evennett. Fluorescence microscopy, contrast techniques in light microscopy. *BIOS Scientific Publishers, Ltd. (Oxford, U.K.*, 1996.

[16] Stylianidou S. Kuwada N.J. Traxler B. Cass, J. and P.A. Wiggins. Probing bacterial cell biology using image cytometry. *Mol. Micro.*, 103:818–828, 2017.

[17] Di Liu Yi Xiao, Christopher H Bowen and Fuzhong Zhang. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nature Chem. Bio.*, 12:339–344, 2016.

[18] Sven van Teeffelen, Joshua W Shaevitz, and Zemer Gitai. Image analysis in fluorescence microscopy: bacterial dynamics as a case study. *Bioessays*, 34(5):427–36, May 2012.

[19] Nathan J Kuwada, Keith C Cheveralls, Beth Traxler, and Paul A Wiggins. Mapping the driving forces of chromosome structure and segregation in *Escherichia coli*. *Nucleic Acids Res*, 41(15):7370–7, Aug 2013.

[20] Nathan J Kuwada, Beth Traxler, and Paul A Wiggins. Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. *Mol Microbiol*, 95(1):64–79, Jan 2015.

[21] Julie A. Cass, Nathan J. Kuwada, Beth Traxler, and Paul A. Wiggins. *Escherichia coli* chromosomal loci segregate from midcell with universal dynamics. *Biophysical Journal*, 110(12):2597–2609, 2016/06/21 2016.

[22] Adrien Ducret, Ellen M. Quardokus, and Yves V. Brun. Microbej, a tool for high throughput bacterial cell detection and quantitative analysis. *Nature Microbiology*, 1:16077 EP –, 06 2016.

[23] Ahmad Paintdakhi, Bradley Parry, Manuel Campos, Irnov Irnov, Johan Elf, Ivan Surovtsev, and Christine Jacobs-Wagner. Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol*, 99(4):767–77, Feb 2016.

[24] Paul A Wiggins, Keith C Cheveralls, Joshua S Martin, Robert Lintner, and Jané Kondev. Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc Natl Acad Sci U S A*, 107(11):4991–5, Mar 2010.

[25] Michele LeRoux, Justin A De Leon, Nathan J Kuwada, Alistair B Russell, Delia Pinto-Santini, Rachel D Hood, Danielle M Agnello, Stephen M Robertson, Paul A Wiggins, and Joseph D Mougous. Quantitative single-cell characterization of bacterial interactions reveals type VI secretion is a double-edged sword. *Proc Natl Acad Sci U S A*, 109(48):19804–9, Nov 2012.

[26] Michele LeRoux, Robin L Kirkpatrick, Elena I Montauti, Bao Q Tran, S Brook Peterson, Brittany N Harding, John C Whitney, Alistair B Russell, Beth Traxler, Young Ah Goo, David R Goodlett, Paul A Wiggins, and Joseph D Mougous. Kin cell lysis is a danger signal that activates antibacterial pathways of *Pseudomonas aeruginosa*. *Elife*, 4, 2015.

[27] Nathan J Kuwada, Beth Traxler, and Paul A Wiggins. High-throughput cell-cycle imaging opens new doors for discovery. *Curr Genet*, 61(4):513–6, Nov 2015.

[28] Michael R Lamprecht, David M Sabatini, and Anne E Carpenter. Cellprofiler: free, versatile software for automated biological image analysis. *Biotechniques*, 42(1):71–5, Jan 2007.

[29] E Bi and J Lutkenhaus. Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *J Bacteriol*, 175(4):1118–25, Feb 1993.

[30] Stuart A. West, Stephen P. Diggle, Angus Buckling, Andy Gardner, and Ashleigh S. Griffin. The social lives of microbes. *Annu. Rev. Ecol. Evol. Syst. Annual Review of Ecology, Evolution, and Systematics*, 38(1):53–77, 2007.

[31] Brenda Youngren, Henrik Jörk Nielsen, Suckjoon Jun, and Stuart Austin. The multifork *Escherichia coli* chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes Dev*, 28(1):71–84, Jan 2014.

[32] Emily Helgesen, Solveig Fossum-Raunehaug, Frank Sætre, Kay Oliver Schink, and Kirsten Skarstad. Dynamic *Escherichia coli* SeqA complexes organize the newly replicated DNA at a considerable distance from the replisome. *Nucleic Acids Res*, 43(5):2730–43, Mar 2015.

[33] Masanari Kitagawa, Takeshi Ara, Mohammad Arifuzzaman, Tomoko Ioka-Nakamichi, Eiji Inamoto, Hiromi Toyonaga, and Hirotada Mori. Complete set of orf clones of *Escherichia coli* ASKA library (a complete set of *E. coli* k-12 orf archive): unique resources for biological research. *DNA Res*, 12(5):291–9, 2005.

[34] M B Elowitz and S Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–8, Jan 2000.

[35] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–6, Aug 2002.

[36] Jeffery Errington, Richard A Daniel, and Dirk-Jan Scheffers. Cytokinesis in bacteria. *Microbiol Mol Biol Rev*, 67(1):52–65, table of contents, Mar 2003.

[37] Joe Lutkenhaus. Assembly dynamics of the bacterial MinCDE system and spatial regulation of the z ring. *Annu Rev Biochem*, 76:539–62, 2007.

[38] Eric J Stewart, Richard Madden, Gregory Paul, and François Taddei. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol*, 3(2):e45, Feb 2005.

[39] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of *Escherichia coli*. *Curr Biol*, 20(12):1099–103, Jun 2010.

[40] David T Kysela, Pamela J B Brown, Kerwyn Casey Huang, and Yves V Brun. Biological consequences and advantages of asymmetric bacterial growth. *Annu Rev Microbiol*, 67:417–35, 2013.

[41] Houra Merrikh, Yan Zhang, Alan D Grossman, and Jue D Wang. Replication-transcription conflicts in bacteria. *Nat Rev Microbiol*, 10(7):449–58, Jul 2012.

[42] Jing Ge, David K Wood, David M Weingeist, Somsak Prasongtanakij, Panida Navasumrit, Mathuros Ruchirawat, and Bevin P Engelward. Standard fluorescent imaging of live cells is highly genotoxic. *Cytometry A*, 83(6):552–60, Jun 2013.

[43] J Cairns. The bacterial chromosome and its manner of replication as seen by autoradiography. *J Mol Biol*, 6:208–13, Mar 1963.

[44] Shun Adachi, Masamichi Kohiyama, Toshinari Onogi, and Sota Hiraga. Localization of replication forks in wild-type and mukb mutant cells of escherichia coli. *Mol Genet Genomics*, 274(3):264–71, Oct 2005.

[45] Tanneke den Blaauwen, Mirjam E G Aarsman, Linda J Wheeler, and Nanne Nanninga. Pre-replication assembly of e. coli replisome components. *Mol Microbiol*, 62(3):695–708, Nov 2006.

[46] Xindan Wang, Rodrigo Reyes-Lamothe, and David J Sherratt. Modulation of *Escherichia coli* sister chromosome cohesion by topoisomerase iv. *Genes Dev*, 22(17):2426–33, Sep 2008.

[47] Christian Lesterlin, Emmanuelle Gigant, Frédéric Boccard, and Olivier Espéli. Sister chromatid interactions in bacteria revealed by a site-specific recombination assay. *EMBO J*, 31(16):3468–79, Aug 2012.

[48] Mohan C Joshi, David Magnan, Timothy P Montminy, Mark Lies, Nicholas Stepankiw, and David Bates. Regulation of sister chromosome cohesion by the replication fork tracking protein seqa. *PLoS Genet*, 9(8):e1003673, 2013.

[49] Patrick H Viollier, Martin Thanbichler, Patrick T McGrath, Lisandra West, Maliwan Meewan, Harley H McAdams, and Lucy Shapiro. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial dna replication. *Proc Natl Acad Sci U S A*, 101(25):9257–62, Jun 2004.

[50] Henrik J Nielsen, Yongfang Li, Brenda Youngren, Flemming G Hansen, and Stuart Austin. Progressive segregation of the *Escherichia coli* chromosome. *Mol Microbiol*, 61(2):383–93, Jul 2006.

[51] Rodrigo Reyes-Lamothe, Xindan Wang, and David Sherratt. *Escherichia coli* and its chromosome. *Trends Microbiol*, 16(5):238–45, May 2008.

[52] Henrik J Nielsen, Jesper R Ottesen, Brenda Youngren, Stuart J Austin, and Flemming G Hansen. The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol Microbiol*, 62(2):331–8, Oct 2006.

[53] Xindan Wang, Xun Liu, Christophe Possoz, and David J Sherratt. The two *Escherichia coli* chromosome arms locate to separate cell halves. *Genes Dev*, 20(13):1727–31, Jul 2006.

[54] Xun Liu, Xindan Wang, Rodrigo Reyes-Lamothe, and David Sherratt. Replication-directed sister chromosome alignment in escherichia coli. *Mol Microbiol*, 75(5):1090–7, Mar 2010.

[55] Michèle Valens, Stéphanie Penaud, Michèle Rossignol, François Cornet, and Frédéric Boccard. Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J*, 23(21):4330–41, Oct 2004.

[56] H Niki, Y Yamaichi, and S Hiraga. Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev*, 14(2):212–23, Jan 2000.

[57] Jean-Christophe Meile, Romain Mercier, Mathieu Stouf, Carine Pages, Jean-Yves Bouet, and François Cornet. The terminal region of the e. coli chromosome localises at the periphery of the nucleoid. *BMC Microbiol*, 11(1):28, Feb 2011.

[58] Olivier Espéli, Romain Borne, Pauline Dupaigne, Axel Thiel, Emmanuelle Gigant, Romain Mercier, and Frédéric Boccard. A matp-divisome interaction coordinates chromosome segregation with cell division in e. coli. *EMBO J*, 31(14):3198–211, May 2012.

[59] David Bates and Nancy Kleckner. Chromosome and replisome dynamics in *E. coli*: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, 121(6):899–911, Jun 2005.

[60] D C Lin, P A Levin, and A D Grossman. Bipolar localization of a chromosome partition protein in bacillus subtilis. *Proc Natl Acad Sci U S A*, 94(9):4721–6, Apr 1997.

[61] C D Webb, A Teleman, S Gordon, A Straight, A Belmont, D C Lin, A D Grossman, A Wright, and R Losick. Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of b. subtilis. *Cell*, 88(5):667–74, Mar 1997.

[62] C D Webb, P L Graumann, J A Kahana, A A Teleman, P A Silver, and R Losick. Use of time-lapse microscopy to visualize rapid movement of the replication origin region of the chromosome during the cell cycle in bacillus subtilis. *Mol Microbiol*, 28(5):883–92, Jun 1998.

[63] P Glaser, M E Sharpe, B Raether, M Perego, K Ohlsen, and J Errington. Dynamic, mitotic-like behavior of a bacterial protein required for accurate chromosome partitioning. *Genes Dev*, 11(9):1160–8, May 1997.

[64] M E Sharpe and J Errington. A fixed distance for separation of newly replicated copies of oric in bacillus subtilis: implications for co-ordination of chromosome segregation and cell division. *Mol Microbiol*, 28(5):981–90, Jun 1998.

[65] K P Lemon and A D Grossman. Localization of bacterial dna polymerase: evidence for a factory model of replication. *Science*, 282(5393):1516–9, Nov 1998.

[66] Melanie B Berkmen and Alan D Grossman. Spatial and temporal organization of the bacillus subtilis replication cycle. *Mol Microbiol*, 62(1):57–71, Oct 2006.

[67] Xindan Wang, Paula Montero Llopis, and David Z Rudner. Bacillus subtilis chromosome organization oscillates between two distinct patterns. *Proc Natl Acad Sci U S A*, 111(35):12877–82, Sep 2014.

[68] Michael A Fogel and Matthew K Waldor. Distinct segregation dynamics of the two vibrio cholerae chromosomes. *Mol Microbiol*, 55(1):125–36, Jan 2005.

[69] Michael A Fogel and Matthew K Waldor. A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev*, 20(23):3269–82, Dec 2006.

[70] D A Mohl and J W Gober. Cell cycle-dependent polar localization of chromosome partitioning proteins in caulobacter crescentus. *Cell*, 88(5):675–84, Mar 1997.

[71] Rainer M Figge, Jesse Easter, and James W Gober. Productive interaction between the chromosome partitioning proteins, para and parb, is required for the progression of the cell cycle in caulobacter crescentus. *Mol Microbiol*, 47(5):1225–37, Mar 2003.

[72] Yoshiharu Yamaichi, Michael A Fogel, Sarah M McLeod, Monica P Hui, and Matthew K Waldor. Distinct centromere-like pars sites on the two chromosomes of vibrio spp. *J Bacteriol*, 189(14):5314–24, Jul 2007.

[73] Yoshiharu Yamaichi, Michael A Fogel, and Matthew K Waldor. par genes and the pathology of chromosome loss in vibrio cholerae. *Proc Natl Acad Sci U S A*, 104(2):630–5, Jan 2007.

[74] Conrad W Shebelut, Jonathan M Guberman, Sven van Teeffelen, Anastasiya A Yakhnina, and Zemer Gitai. Caulobacter chromosome segregation is an ordered multistep process. *Proc Natl Acad Sci U S A*, 107(32):14194–8, Aug 2010.

[75] K Ireton, N W Gunther, 4th, and A D Grossman. spo0j is required for normal chromosome segregation as well as the initiation of sporulation in bacillus subtilis. *J Bacteriol*, 176(17):5320–9, Sep 1994.

[76] D C Lin and A D Grossman. Identification and characterization of a bacterial chromosome partitioning site. *Cell*, 92(5):675–85, Mar 1998.

[77] Olivier Espeli, Romain Mercier, and Frédéric Boccard. DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol Microbiol*, 68(6):1418–27, Jun 2008.

[78] Stephanie C Weber, Andrew J Spakowitz, and Julie A Theriot. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys Rev Lett*, 104(23):238102, Jun 2010.

[79] Stephanie C Weber, Julie A Theriot, and Andrew J Spakowitz. Subdiffusive motion of a polymer composed of subdiffusive monomers. *Phys Rev E Stat Nonlin Soft Matter Phys*, 82(1 Pt 1):011913, Jul 2010.

[80] Avelino Javer, Nathan J Kuwada, Zhicheng Long, Vincenzo G Benza, Kevin D Dorfman, Paul A Wiggins, Pietro Cicuta, and Marco Cosentino Lagomarsino. Persistent super-diffusive motion of escherichia coli chromosomal loci. *Nat Commun*, 5:3854, 2014.

[81] Steven Elmore, Michiel Müller, Norbert Vischer, Theo Odijk, and Conrad L Woldringh. Single-particle tracking of oric-gfp fluorescent spots during chromosome segregation in *Escherichia coli*. *J Struct Biol*, 151(3):275–87, Sep 2005.

[82] Jay K Fisher, Aude Bourniquel, Guillaume Witz, Beth Weiner, Mara Prentiss, and Nancy Kleckner. Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells. *Cell*, 153(4):882–95, May 2013.

[83] Nastaran Hadizadeh Yazdi, Calin C Guet, Reid C Johnson, and John F Marko. Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions. *Mol Microbiol*, 86(6):1318–33, Dec 2012.

[84] Mathieu Stouf, Jean-Christophe Meile, and François Cornet. Ftsk actively segregates sister chromosomes in escherichia coli. *Proc Natl Acad Sci U S A*, 110(27):11157–62, Jul 2013.

[85] Ivy F Lau, Sergio R Filipe, Britta Søballe, Ole-Andreas Økstad, Francois-Xavier Barre, and David J Sherratt. Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol Microbiol*, 49(3):731–43, Aug 2003.

[86] Rodrigo Reyes-Lamothe, Christophe Possoz, Olessia Danilova, and David J Sherratt. Independent positioning and action of *Escherichia coli* replisomes in live cells. *Cell*, 133(1):90–102, Apr 2008.

[87] Jean-Philippe Bouchaud and Antoine Georges. Anomalous diffusion in disordered media: Statistical mechanisms, models, and physical applications. *Physics Reports*, 195:127–293, 1990.

[88] Mohan C Joshi, Aude Bourniquel, Jay Fisher, Brian T Ho, David Magnan, Nancy Kleckner, and David Bates. *Escherichia coli* sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc Natl Acad Sci U S A*, 108(7):2765–70, Feb 2011.

[89] Aretha Fiebig, Kinneret Keren, and Julie A Theriot. Fine-scale time-lapse analysis of the biphasic, dynamic behaviour of the two *Vibrio cholerae* chromosomes. *Mol Microbiol*, 60(5):1164–78, Jun 2006.

[90] Romain Mercier, Marie-Agnès Petit, Sophie Schbath, Stéphane Robin, Meriem El Karoui, Frédéric Boccard, and Olivier Espéli. The matp/mats site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell*, 135(3):475–85, Oct 2008.

[91] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[92] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference.* Springer-Verlag New York, Inc., 2nd. edition, 1998.

[93] Bo Wang, James Kuo, Sung Chul Bae, and Steve Granick. When brownian diffusion is not gaussian. *Nat Mater*, 11(6):481–5, Jun 2012.

[94] Bradley R Parry, Ivan V Surovtsev, Matthew T Cabeen, Corey S O'Hern, Eric R Dufresne, and Christine Jacobs-Wagner. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell*, 156(1-2):183–94, Jan 2014.

[95] Thomas J Lampo, Nathan J Kuwada, Paul A Wiggins, and Andrew J Spakowitz. Physical modeling of chromosome segregation in *Escherichia coli* reveals impact of force and DNA relaxation. *Biophys J*, 108(1):146–53, Jan 2015.

[96] Stephanie C Weber, Michael A Thompson, W E Moerner, Andrew J Spakowitz, and Julie A Theriot. Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. *Biophys J*, 102(11):2443–50, Jun 2012.

[97] Stephanie C Weber, Andrew J Spakowitz, and Julie A Theriot. Nonthermal atp-dependent fluctuations contribute to the in vivo motion of chromosomal loci. *Proc Natl Acad Sci U S A*, 109(19):7338–43, May 2012.

[98] Xindan Wang, Christophe Possoz, and David J Sherratt. Dancing around the divisome: asymmetric chromosome segregation in *Escherichia coli*. *Genes Dev*, 19(19):2367–77, Oct 2005.

[99] Elena Minina and Axel Arnold. Induction of entropic segregation: the first step is the hardest. *Soft Matter*, 10(31):5836–41, Aug 2014.

[100] M Lu, J L Campbell, E Boye, and N Kleckner. Seqa: a negative modulator of replication initiation in *E. coli*. *Cell*, 77(3):413–26, May 1994.

[101] K Ohsumi, M Yamazoe, and S Hiraga. Different localization of seqa-bound nascent DNA clusters and mukf-muke-mukb complex in *Escherichia coli* cells. *Mol Microbiol*, 40(4):835–45, May 2001.

[102] Cedric Cagliero, Ralph S Grand, M Beatrix Jones, Ding J Jin, and Justin M O'Sullivan. Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription. *Nucleic Acids Res*, 41(12):6058–71, Jul 2013.

[103] Jean-Yves Bouet, Mathieu Stouf, Elise Lebailly, and François Cornet. Mechanisms for chromosome segregation. *Curr Opin Microbiol*, 22:60–5, Dec 2014.

[104] Cass J. A. Merrikh H. Mangiameli, S. M. and P.A. Wiggins. Bacterial replisomes show factory-like behavior. *Current Genetics Review In preparation.*

[105] Esteban Toro, Sun-Hae Hong, Harley H McAdams, and Lucy Shapiro. Caulobacter requires a dedicated mechanism to initiate chromosome segregation. *Proc Natl Acad Sci U S A*, 105(40):15435–40, Oct 2008.

[106] Suckjoon Jun and Bela Mulder. Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. *Proc Natl Acad Sci U S A*, 103(33):12388–93, Aug 2006.

[107] Nancy Kleckner, Denise Zickler, Gareth H Jones, Job Dekker, Ruth Padmore, Jim Henle, and John Hutchinson. A mechanical basis for chromosome function. *Proc Natl Acad Sci U S A*, 101(34):12592–7, Aug 2004.

[108] J A Sawitzke and S Austin. Suppression of chromosome segregation defects of *Escherichia coli* muk mutants by mutations in topoisomerase i. *Proc Natl Acad Sci U S A*, 97(4):1671–6, Feb 2000.

[109] Olessia Danilova, Rodrigo Reyes-Lamothe, Marina Pinskaya, David Sherratt, and Christophe Possoz. Mukb colocalizes with the oric region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves. *Mol Microbiol*, 65(6):1485–92, Sep 2007.

[110] Elnaz Alipour and John F Marko. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res*, 40(22):11202–12, Dec 2012.

[111] Handuo Shi Tomasz L. Czarny Matthew H. Larson Spencer Wong John S. Hawkins Candy H.S. Lu Byoung-Mo Koo Elizabeth Marta Anthony L. Shiver Evan H. Whitehead Jonathan S. Weissman Eric D. Brown Lei S. Qi Kerwyn Casey Huang Jason M. Peters, Alexandre Colavin and Carole A. Gross. A comprehensive, crispr-based functional analysis of essential genes in bacteria. *Cell*, 165:14931506, 2016.

[112] Onusko EM Holten KB. Appropriate prescribing of oral beta-lactam antibiotics. *American Family Physician.*, 62:611–620, 2000.

[113] Hoischen C. Allan, E.J. and Gumpert J. Bacterial l-forms. *Advances in App. Micro.*, 68:1–39.

[114] Y. Errington J. Mercier R., Kawai. Wall proficient e. coli capable of sustained growth in the absence of the z-ring division machine. *eLife*, 3, 2014.

[115] Errington J. L-form bacteria, cell walls and the origins of life. *Open Biol.*, 3, 2013.

[116] Juni E. Genetics and physiology of acinetobacter. *Annu Rev Microbiol.*, 32:349–371, 1978.

[117] Nemec A. Dijkshoorn, L. and H. Seifert. An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*, journal = Nat. Rev. Microbiology, volume = 5, pages = 939-951, year = 2007.

[118] Gianoulis T.A. Pukatzki S. Mekalanos J.J. Ornston L.N. Gerstein M. Smith, M.G. and M. Snyder. New insights into acinetobacter baumannii pathogenesis revealed by high-density pyrosequencing and transposon mutagenesis. *Genes Dev.*, 21:604–614, 2007.

[119] J. Chastre and J.L. Trouillet. Problem pathenogens. *Semin Respir Infect.*, 15:287–298, 2000.

[120] Goglin K. Molyneaux N. Hujer K.M. Lavender H. Jamison J.J. MacDonald I.J. Martin K.M. Russo T. Campagnari A.A. Hujer A.M.-Bonomo R.A. Gill S.R. Adams, M.D. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*, journal = J. Bacteriol., volume = 190, issue = 24, pages = 8053-8064, year = 2008.

[121] John M. Martinko Kelly S. Bender Daniel H. Buckley Madigan, Michael T. and David Allan Stahl. *Brock Biology of Microorganisms.* Boston: Pearson, 14 edition.

[122] Kathryn T. Elliott and Ellen L. Neidle. Acinetobacter baylyi adp1: Transforming the choice of model organism. *IUBMB Life*, 63:1075–1080, 2011.

[123] Jay Shendure Larry A. Gallagher and Colin Manoil. Genome-scale identification of resistance functions in pseudomonas aeruginosa using tn-seq. *MBio.*, 2, 2011.

[124] Christopher T. Walsh Eric D. Brown, 1 Eugenio I. Vivas and Roberto Kolter. Mura (murz), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in escherichia coli. *J. Bacteriol.*, 177:41944197.

[125] Y. Errington J. Mercier R., Kawai. General principles for the formation and proliferation of a wall-free (l-form) state in bacteria. *Nature Microbiology Letters*, 3.

[126] Christopher Fuller Juliet Girard Joanna Lipinski-Kruszka Dan Lu Thomas Noriega Geoffrey Rollins Russell Spitzer Michael Todhunter Emma McCullagh, Justin Farlow and Hana El-Samad. Not all quiet on the noise front. *Nature Chem. Bio.*, 5:700–704, 2009.

[127] Connor Brennan and Paul A. Wiggins. 2016. An informative geometric measure of fluorescence localization for cell biology. *Submitted to Biophys. J.*, 2016.

[128] Beat Christen, Michael J Fero, Nathan J Hillson, Grant Bowman, Sun-Hae Hong, Lucy Shapiro, and Harley H McAdams. High-throughput identification of protein localization dependency networks. *Proc Natl Acad Sci U S A*, 107(10):4681–6, Mar 2010.

[129] Cecilia Garmendia-Torres, Alexander Skupin, Sean A Michael, Pekka Ruusuvuori, Nathan J Kuwada, Didier Falconnet, Gregory A Cary, Carl Hansen, Paul A Wiggins, and Aimée M Dudley. Unidirectional p-body transport during the yeast cell cycle. *PLoS One*, 9(6):e99428, 2014.

[130] Orit Gefen, Ofer Fridman, Irine Ronin, and Nathalie Q Balaban. Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity. *Proc Natl Acad Sci U S A*, 111(1):556–61, Jan 2014.

[131] Zemer Gitai. The new bacterial cell biology: moving parts and subcellular architecture. *Cell*, 120(5):577–86, Mar 2005.

[132] Ido Golding and Edward C Cox. RNA dynamics in live Escherichia coli cells. *Proc Natl Acad Sci U S A*, 101(31):11310–5, Aug 2004.

[133] Jonathan M Guberman, Allison Fay, Jonathan Dworkin, Ned S Wingreen, and Zemer Gitai. Psicic: noise and asymmetry in bacterial division revealed by computational image analysis at sub-pixel resolution. *PLoS Comput Biol*, 4(11):e1000233, Nov 2008.

[134] M. Guizar-Sicairos, G.-S. Manuel, S.T. Thurman, and J.R. Fienup. Efficient subpixel image registration algorithms. *Opt. Lett.*, 33:156, 2008.

[135] Robert Ietswaart, Florian Szardenings, Kenn Gerdes, and Martin Howard. Competing para structures space bacterial plasmids equally over the nucleoid. *PLoS Comput Biol*, 10(12):e1004009, Dec 2014.

[136] Shinya Inoué and Kenneth R. Spring. *Video Microscopy. The Fundamentals.* Plenum Press: New York and London, 1997.

[137] Nathan J Kuwada, Martin J Zuckermann, Elizabeth H C Bromley, Richard B Sessions, Paul M G Curmi, Nancy R Forde, Derek N Woolfson, and Heiner Linke. Tuning the performance of an artificial protein motor. *Phys Rev E Stat Nonlin Soft Matter Phys*, 84(3 Pt 1):031922, Sep 2011.

[138] Michele LeRoux, S Brook Peterson, and Joseph D Mougous. Bacterial danger sensing. *J Mol Biol*, 427(23):3744–53, Nov 2015.

[139] James C W Locke and Michael B Elowitz. Using movies to analyse gene circuit dynamics in single cells. *Nat Rev Microbiol*, 7(5):383–92, May 2009.

[140] Daniel Madar, Erez Dekel, Anat Bren, Anat Zimmer, Ziv Porat, and Uri Alon. Promoter activity dynamics in the lag phase of escherichia coli. *BMC Syst Biol*, 7:136, 2013.

[141] Michael T. Madigan, John M. Martinko, David Stahl, and David P. Clark. *Brock Biology of Microorganisms (13th Edition).* Benjamin Cummings, 13 edition, December 2010.

[142] Sarah Mangiameli, Chris Merrikh, Paul A. Wiggins, and Houra Merrikh. In vivo visualization of replication-transcription conflicts reveals pervasive replisome instability. *Revised for Molecular Cell.*, 0, 2016.

[143] Sarah Mangiameli, Chris Merrikh, Houra Merrikh, and Paul A. Wiggins. Tracking replisome movement throughout the cell cycle. *In preparation for PNAS.*, 0, 2016.

[144] F. Meyer and M. Fernand. Topographic distance and watershed lines. *Signal Processing*, 38:113–125, 1994.

[145] Jennifer R O'Connor, Nathan J Kuwada, Varisa Huangyutitham, Paul A Wiggins, and Caroline S Harwood. Surface sensing and lateral subcellular localization of wspa, the receptor in a chemosensory-like system leading to c-di-gmp production. *Mol Microbiol*, 86(3):720–9, Nov 2012.

[146] Matthew D Rolfe, Christopher J Rice, Sacha Lucchini, Carmen Pin, Arthur Thompson, Andrew D S Cameron, Mark Alston, Michael F Stringer, Roy P Betts, József Baranyi, Michael W Peck, and Jay C D Hinton. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol*, 194(3):686–701, Feb 2012.

[147] Alistair B Russell, Rachel D Hood, Nhat Khai Bui, Michele LeRoux, Waldemar Vollmer, and Joseph D Mougous. Type vi secretion delivers bacteriolytic effectors to target cells. *Nature*, 475(7356):343–7, Jul 2011.

[148] Alistair B Russell, Michele LeRoux, Krisztina Hathazi, Danielle M Agnello, Takahiko Ishikawa, Paul A Wiggins, Sun Nyunt Wai, and Joseph D Mougous. Diverse type vi secretion phospholipases are functionally plastic antibacterial effectors. *Nature*, 496(7446):508–12, Apr 2013.

[149] Jon Russell. (personal communication), 2016.

[150] Sven van Teeffelen, Joshua W Shaevitz, and Zemer Gitai. Image analysis in fluorescence microscopy: bacterial dynamics as a case study. *Bioessays*, 34(5):427–36, May 2012.

[151] Jan-Willem Veening, Wiep Klaas Smits, and Oscar P Kuipers. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol*, 62:193–210, 2008.

[152] Quanli Wang, Jarad Niemi, Chee-Meng Tan, Lingchong You, and Mike West. Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry A*, 77(1):101–10, Jan 2010.

[153] Jonathan W Young, James C W Locke, Alphan Altinok, Nitzan Rosenfeld, Tigran Bacarian, Peter S Swain, Eric Mjolsness, and Michael B Elowitz. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc*, 7(1):80–8, Jan 2012.

[154] Zhenmao Wan, Joseph Varshavsky, Sushma Teegala, Jamille McLawrence, and Noel L Goddard. Measuring the rate of conjugal plasmid transfer in a bacterial population using quantitative PCR. *Biophys J*, 101(1):237–44, Jul 2011.

[155] Elena Cabezón, Jorge Ripoll-Rozada, Alejandro Peña, Fernando de la Cruz, and Ignacio Arechaga. Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev*, 39(1):81–95, Jan 2015.

[156] Emily Helgesen, Solveig Fossum-Raunehaug, and Kirsten Skarstad. Lack of the h-ns protein results in extended and aberrantly positioned dna during chromosome replication and segregation in escherichia coli. *J Bacteriol*, 198(8):1305–16, Apr 2016.

[157] H M Davey and D B Kell. Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiol Rev*, 60(4):641–96, Dec 1996.

[158] C L Woldringh, P G Huls, and N O Vischer. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/α as determined by image cytometry. *J Bacteriol*, 175(10):3174–81, May 1993.

[159] K D Wittrup. Protein engineering by cell-surface display. *Curr Opin Biotechnol*, 12(4):395–9, Aug 2001.

[160] Anne-Lyse Ducrest, Mario Amacker, Joachim Lingner, and Markus Nabholz. Detection of promoter activity by flow cytometric analysis of GFP reporter expression. *Nucleic Acids Res*, 30(14):e65, Jul 2002.

[161] Tobias Sawetzki, Charles D Eggleton, Sanjay A Desai, and David W M Marr. Viscoelasticity as a biomarker for high-throughput flow cytometry. *Biophys J*, 105(10):2281–8, Nov 2013.

[162] W A Bonner, H R Hulett, R G Sweet, and L A Herzenberg. Fluorescence activated cell sorting. *Rev Sci Instrum*, 43(3):404–9, Mar 1972.

[163] Cassandra Stowe, Arnold Pizzey, Tammy Kalber, Adam Badar, Mark Lythgoe, and Martin Pule. Flow-based single cell deposition for high-throughput screening of protein libraries. *PLoS One*, 10(11):e0140730, 2015.

[164] Claire Simonnet and Alex Groisman. High-throughput and high-resolution flow cytometry in molded microfluidic devices. *Anal Chem*, 78(16):5653–63, Aug 2006.

[165] Myron R Melamed, T Lindmo, and Mortimer L Mendelsohn. *Flow cytometry and sorting*. Wiley-Liss, New York, 2nd ed edition, 1990.

[166] Howard M Shapiro. *Practical flow cytometry*. Wiley-Liss, New York, 4th ed edition, 2003.

[167] Qin Huang, Chenggong Yu, Xiaoqi Zhang, and Raj K Goyal. Comparison of dna histograms by standard flow cytometry and image cytometry on sections in barrett's adenocarcinoma. *BMC Clin Pathol*, 8:5, 2008.

[168] Yuanyuan Han and Yu-Hwa Lo. Imaging cells in flow cytometer using spatial-temporal transformation. *Sci Rep*, 5:13267, 2015.

[169] Yu-ichi Ozaki, Shinsuke Uda, Takeshi H Saito, Jaehoon Chung, Hiroyuki Kubota, and Shinya Kuroda. A quantitative image cytometry technique for time series or population analyses of signaling networks. *PLoS One*, 5(4):e9955, 2010.

[170] F F Mandy, M Bergeron, and T Minkus. Principles of flow cytometry. *Transfus Sci*, 16(4):303–14, Dec 1995.

[171] J V Watson. The early fluidic and optical physics of cytometry. *Cytometry*, 38(1):2–14; discussion 1, Feb 1999.

[172] D J Recktenwald. Introduction to flow cytometry: principles, fluorochromes, instrument set-up, calibration. *J Hematother*, 2(3):387–94, 1993.

[173] M Brown and C Wittwer. Flow cytometry: principles and clinical applications in hematology. *Clin Chem*, 46(8 Pt 2):1221–9, Aug 2000.

[174] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nat Methods*, 9(7):671–5, Jul 2012.

[175] Oleksii Sliusarenko, Jennifer Heinritz, Thierry Emonet, and Christine Jacobs-Wagner. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Mol Microbiol*, 80(3):612–27, May 2011.

[176] Anjana Badrinarayanan, Christian Lesterlin, Rodrigo Reyes-Lamothe, and David Sherratt. The *Escherichia coli* smc complex, mukbef, shapes nucleoid organization independently of DNA replication. *J Bacteriol*, 194(17):4669–76, Sep 2012.

[177] Anjana Badrinarayanan, Rodrigo Reyes-Lamothe, Stephan Uphoff, Mark C Leake, and David J Sherratt. In vivo architecture and action of bacterial structural maintenance of chromosome proteins. *Science*, 338(6106):528–31, Oct 2012.

[178] Edward J Banigan, Michael A Gelbart, Zemer Gitai, Ned S Wingreen, and Andrea J Liu. Filament depolymerization can explain chromosome pulling during bacterial mitosis. *PLoS Comput Biol*, 7(9):e1002145, Sep 2011.

[179] Esteban Toro and Lucy Shapiro. Bacterial chromosome organization and segregation. *Cold Spring Harb Perspect Biol*, 2(2):a000349, Feb 2010.

[180] Sun-Hae Hong and Harley H McAdams. Compaction and transport properties of newly replicated *Caulobacter crescentus* DNA. *Mol Microbiol*, 82(6):1349–58, Dec 2011.

[181] R B Jensen, S C Wang, and L Shapiro. A moving DNA replication factory in *Caulobacter crescentus*. *EMBO J*, 20(17):4952–63, Sep 2001.

[182] David Bates, Jessica Epstein, Erik Boye, Karen Fahrner, Howard Berg, and Nancy Kleckner. The *Escherichia coli* baby cell column: a novel cell synchronization method provides new insight into the bacterial cell cycle. *Mol Microbiol*, 57(2):380–91, Jul 2005.

[183] Rodrigo Reyes-Lamothe, David J Sherratt, and Mark C Leake. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science*, 328(5977):498–501, Apr 2010.

[184] Ana Babic, Ariel B Lindner, Marin Vulic, Eric J Stewart, and Miroslav Radman. Direct visualization of horizontal gene transfer. *Science*, 319(5869):1533–6, Mar 2008.

[185] Mark-Anthony Bray, Martha S Vokes, and Anne E Carpenter. Using cellprofiler for automatic identification and measurement of biological objects in images. *Curr Protoc Mol Biol*, 109:14.17.1–13, 2015.

Appendix A

APPENDIX

A.1 Adaption of KL Divergence for comparing empirical distributions

Although the KL Divergence is well defined on function p and q as:

$$K(p||q) \equiv \int \mathrm{d}x \; p(x) \log \tfrac{p(x)}{q(x)}, \tag{A.1}$$

a problem arrises with empirical distributions generated from a finite number of observations. For convenience, we binned the data using a histogram and then used the discrete representation of the KL Divergence

$$K(p||q) \equiv \sum_i p_i \log \tfrac{p_i}{q_i}. \tag{A.2}$$

Finally to avoid problems in poorly sampled regions of the histogram, we summed only over bins i in which both distributions had at least one observation. To compute the error in the KL Divergence, we then bootstrapped from the empirical distribution.

Locus	Position (bp)
oriC	3,909,402
R1	4,413,507
R2	515,143
R3	1,081,356
ter1	1,806,680
L2	2,616,013
L1	3,739,211

Table A.1: **Genomic positions** of all loci (*parS* insertion sites) reported in base pairs in the strain MG1655.

A.2 "Genomic" and "Pole" Orientation of trajectories

Many previous studies of chromosome structure were conducted through snapshot experiments, where a heterogenous population of cells is imaged for a single time point. As a result, cells were oriented along their long axis using the positions of the fluorescent foci. This a priori orientation method, which we refer to as "genomic orientation", entails splitting the long axis of each cell in two halves: if the locus is sequentially located on the left arm of the chromosome, the cell was oriented with that locus in the left half of the cell, and vice versa.

Our complete-cell-cycle imaging approach offers a new method for cell orientation: tracking cells through division into daughter cells allows us to identify which cell pole is newly formed through the division event (new pole) and which pole was pre-existing (old pole). As a result, we were able to orient our cells by using pole orientation, such that the new pole of the cell is always on the bottom of the kymograph. Although any choice in orientation may introduce an unavoidable bias, this technique does not rely on an anticipated chromosome structure or dynamics, and thus we feel is the most unambiguous method for combining dynamic position information of from multiple cells.

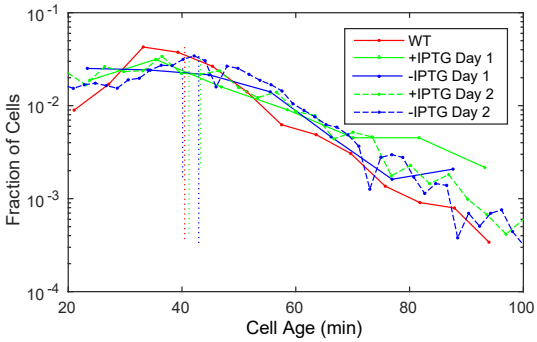

Figure A.1: **Histogram of cell cycle duration for wild-type MG1655 cells (WT), and for MG1655 cells with ParB-*parS* labeling system, both with and without induction (-IPTG, +IPTG).** Dotted lines show population mean. Long-lived cohesion between sisters would delay the cell division increasing the duration of the cell cycle. The variation observed in the population mean is consistent with day-to-day variation. The observation of the shorter tail (cell age > 70) of the wild-type cells suggests that there may be a small subset of the population of the ParB-*parS* cells in which the cell cycle is delayed, but the vast majority of cells are unaffected.

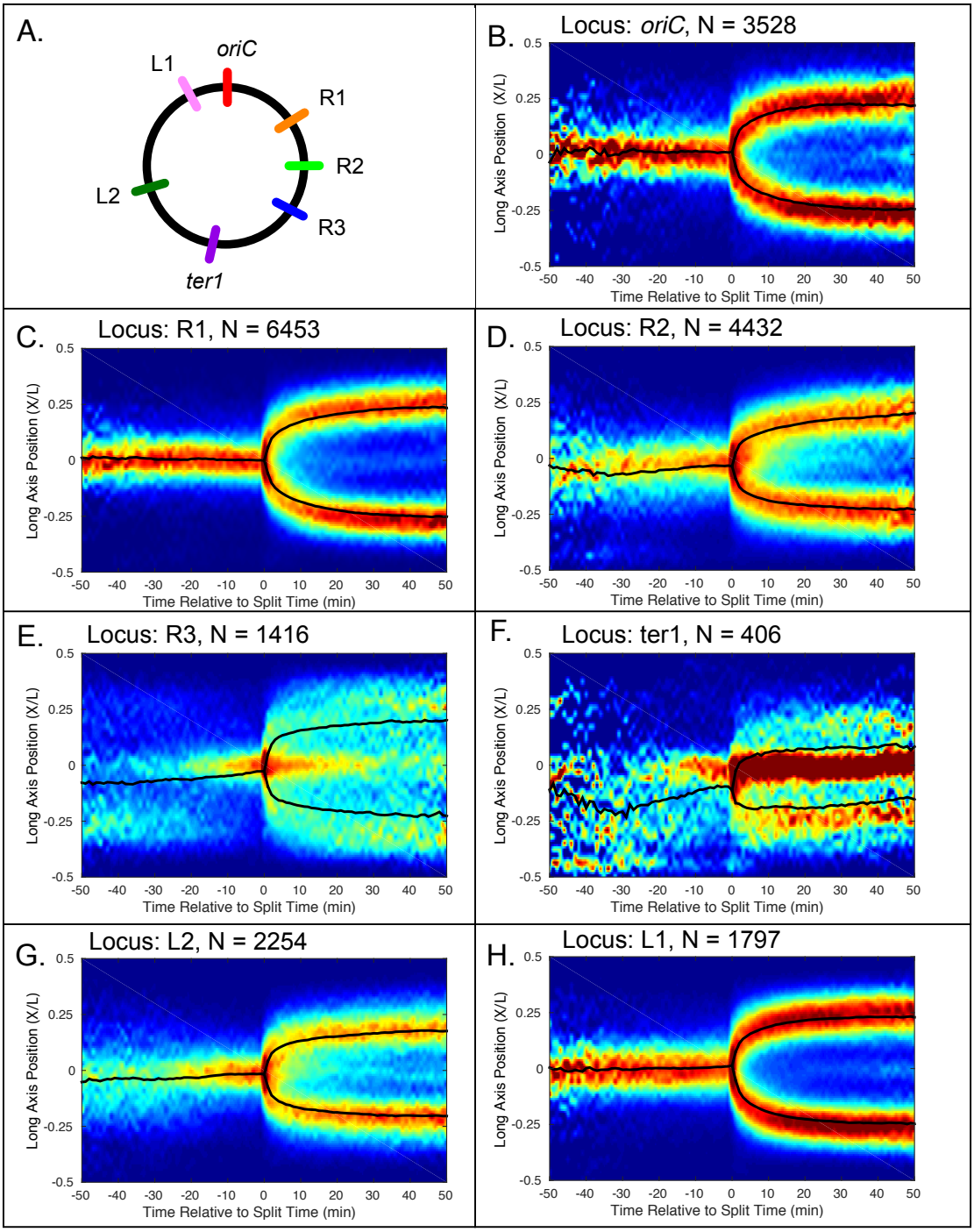

Figure A.2: **Fluorescence histograms for seven loci.** (A) Genetic map of parS insertion sites or loci. (B-H) Histograms of all overlaid synchronized, normalized, oriented trajectories throughout the cell cycle for all chromosomal loci. Tracks are oriented with $L > 0$ indicating the old pole and $L < 0$ indicating the new pole side of the cell.

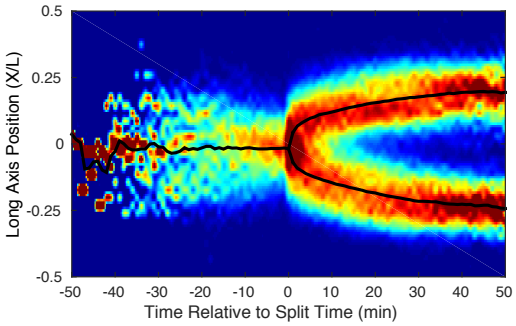

Figure A.3: **Histogram of *lac* trajectories in background strain AB1157.** Trajectories in this histogram have been synchronized, normalized and oriented. Despite being genomicly located in a similar position to "R3" used in this study, the lacI locus in AB1157 possess striking symmetry, consistent with the difference in chromosomal orientation in MG1655 and AB1157.

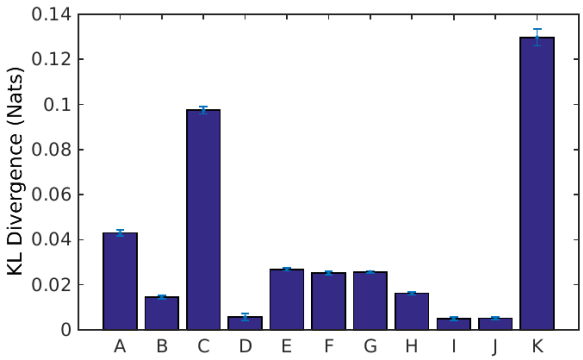

Figure A.4: **KL Divergence values for step-size distributions $p(x)$ and $q(x)$.** The KL Divergence $K(p||q)$ is a qualitative measure of the distance between a true distribution p and a model q. Qualitatively, the number of observations required to make the difference between the distribution significant is $N \approx K^{-1}$. A list of the step-size distributions $p(x)$ and $q(x)$ are given in Table 2. The error bars in the divergences are estimated by bootstrapping from the empirical distribution.

Plot Bar	True distribution $p(x)$	Model distribution $q(x)$	Value
A	*oriC* pre-split	*oriC* post-split	0.0430 ± 0.0010
B	*oriC* pre-split	*oriC* post-split M.S.	0.0146 ± 0.0007
C	*oriC* post-split, forward	*oriC* post-split, reverse	0.0974 ± 0.0036
D	*oriC* post-split, forward M.S.	*oriC* post-split, reverse M.S.	0.0057 ± 0.0017
E	*oriC* pre-split	Gaussian	0.0296 ± 0.0018
F	*oriC* post-split M.S.	Gaussian	0.0253 ± 0.0010
G	*oriC* post-split	Gaussian	0.0257 ± 0.0016
H	*oriC* pre-split	L2 pre-split	0.0162 ± 0.0018
I	*oriC* post-split M.S.	L2 post-split M.S.	0.0050 ± 0.0010
J	*oriC* post-split	L2 post-split	0.0051 ± 0.0016
K	MS2-GFP-mRNA	Gaussian	0.1297 ± 0.0010

Table A.2: **Distributions $p(x)$ and $q(x)$** used to calculate KL divergences in Figure 7. "M.S." indicates distributions which have been mean-subtracted. "Gaussian" indicates a Gaussian fit for the associated $p(x)$.

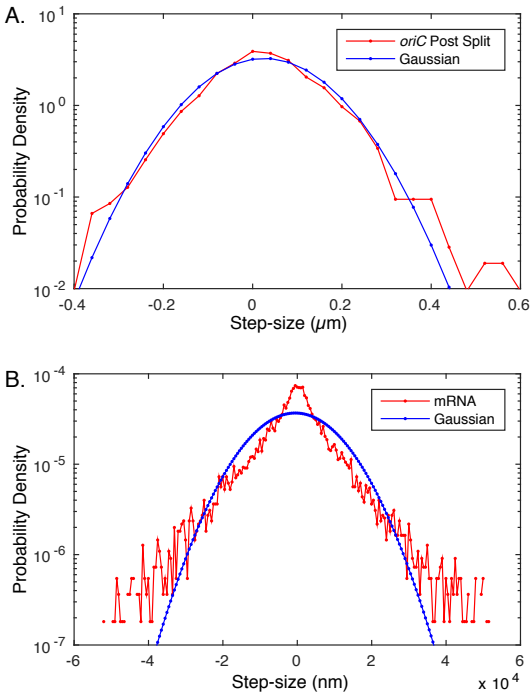

Figure A.5: **Non-Gaussian step-size distributions compared for chromosomal loci and large complexes.** The observed locus step-size distributions all have *Fat tails*: The probability of large magnitude steps is greater than predicted by a Gaussian step-size distribution with the observed mean and variance (of the true distribution). As can be seen in Panel A, the mismatch between the true step-size distribution (red) and the gaussian fit (blue) is quite small compared to the mismatch observed for MS2-GFP-mRNA complexes (Panel B). For MS2-GFP-mRNA complexes, there is a dramatic mismatch between the observed step-size distribution and the Gaussian model [4]. The size of this mismatch is measured by the KL Divergence and shown in Fig. SI A.4.

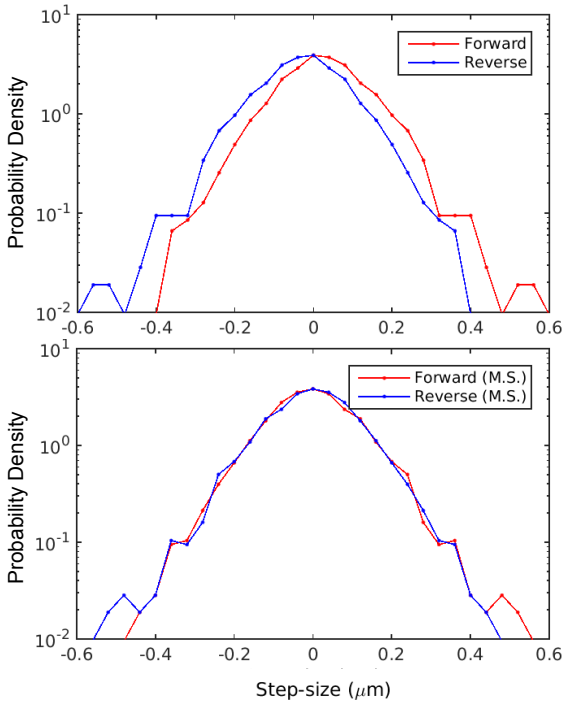

Figure A.6: **Segregation results from a drift velocity.** Consider the distributions of forward, $p(x)$, and reverse, $p(-x)$, biased motion during the first ten minutes of segregation, shown in Panel A. The displacement in the distributions can be understood as the change in the physical character of the motion that results in segregation. To test whether the dominant contribution is consistent with a shift of the distribution in the direction of the bias (a drift velocity), we compute the distribution of mean subtracted steps $p'(\delta x)$ and compare this to the reverse bias distribution $p'(-\delta x)$, as shown in Panel B. The similarity between the these distributions is consistent with the drift velocity representing the dominant contribution to segregation. The difference between the distributions is quantitated using the KL Divergence in Fig. SI A.4.

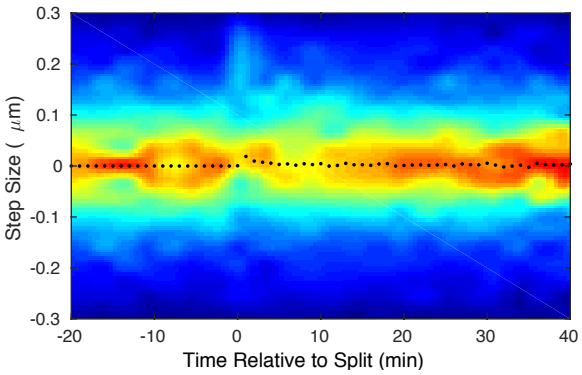

Figure A.7: **Time dependent step-size distribution for *ter1*.** As with the *oriC* time-dependent step-size distribution provided in Figure 6B, the *ter1* step-size distribution is well approximated by a Gaussian distribution with a time dependent bias.

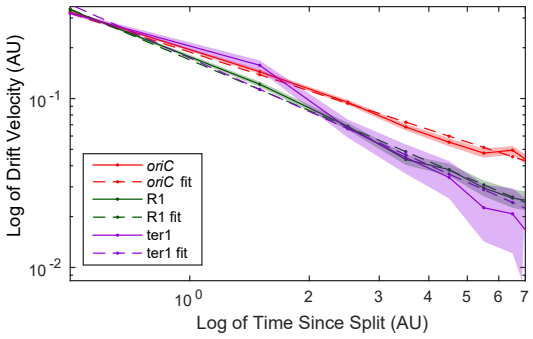

Figure A.8: **Powerlaw fitting of locus drift velocities.** When drift velocities for loci are plotted in log-log scale, we apply powerlaw fitting, Eqn. A.6. For *oriC*, R1 and ter1 the resulting resulting fit slopes are $\alpha = 0.48, 0.0,$ and -0.2 respectively.

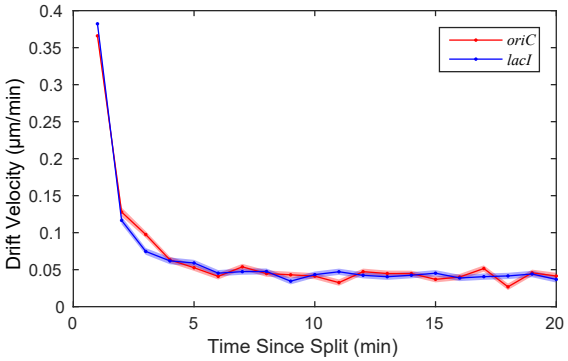

Figure A.9: **Drift velocities for *oriC* and *lac* in model strain AB1157.** The similarities in the profiles of these two locus drift velocities suggests that while structure may vary between strains, universal segregation dynamics are also observed in AB1157 under slow growth conditions.

A.3 Mean velocity as an estimator of instantaneous velocity

In the experiments described in the text, we measure the displacement between foci x at fixed times intervals $\Delta t = \tau = 1$ min where t is the time since the split. We report the velocity as

$$v_*(t + \tfrac{1}{2}\tau) \equiv \tfrac{1}{\tau}\left[x(t+\tau) - x(t)\right], \tag{A.3}$$

where we denote the velocity with a star to denote its definition in terms of a finite difference. This is equivalent to measuring the time-averaged velocity over the time interval $[t, t+\tau]$:

$$v_*(t) \equiv \int_{t-\frac{1}{2}\tau}^{t+\frac{1}{2}\tau} \mathrm{d}t'\ v(t'). \tag{A.4}$$

We can interpret the time-averaged velocity v_* as an estimator of the instantaneous velocity v. The relative error of this estimator is defined:

$$\varepsilon(t) \equiv \frac{v_*(t) - v(t)}{v(t)}. \tag{A.5}$$

To estimate the relative error, we use a parametric model for the drift velocity that we have previously proposed [95]:

$$\overline{v}(t) = v_{\mathrm{app}}\ t^{\alpha/2-1}, \tag{A.6}$$

which will facilitate an explicit estimate of the relative error of the velocity estimator. An integration gives the relative error:

$$\varepsilon(t) = \frac{2t}{\alpha\tau}\left[\left(1+\frac{\tau}{2t}\right)^{\frac{\alpha}{2}} - \left(1-\frac{\tau}{2t}\right)^{\frac{\alpha}{2}}\right], \tag{A.7}$$

which is plotted for $\alpha = 0.6$ in Figure A.10. As expected, the error is small except for the first step $t \approx \tau$.

Additional uncertainty is added by the unknown timing of the split. We assign the split time to the last integer frame in which only a single locus is resolved, but the true splitting time may be any time between this frame and the next frame when two foci are resolved. An analogous computation facilitates the estimate of this error.

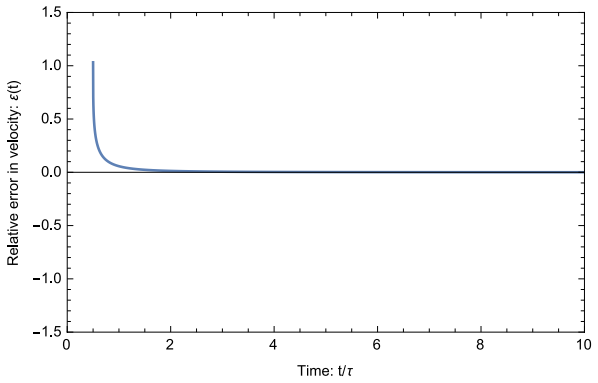

Figure A.10: **Model for the relative error of the mean velocity as an estimator for instantaneous velocity.**

A.4 Software Availability and Documentation

The website for the SuperSegger software [1], which contains the analysis tool `gateTool`, can be found at `http://mtshasta.phys.washington.edu/website/SuperSegger.php`. The software can be downloaded at the GitHub repository `https://github.com/wiggins-lab/SuperSegger/`. The GitHub wiki (found by following the 'Wiki' tab of the GitHub page) contains a detailed outline of how to use the SuperSegger software; this includes instructions for downloading, and details for

both GUI and command-line use of the segmentation (`SuperSegger`) and `gateTool` analysis tool. Tutorials and sample datasets for these processes are also available at `http://mtshasta.phys.washington.edu/website/tutorials.php`.

The "View gateTool tutorial" link brings you to a scroll-through html version of the tutorial. The version begins with an electronic Table of Contents, so that you can jump directly to a section of interest. The "Download Tutorial" link allows you to download the tutorial script and all associated files, so that you may run through the tutorial yourself. This interactive method allows you to select your own gates by clicking to draw gates on 1D and 2D plots, just as you would when using the `gateTool` in your analysis.

Feature	Description
SuperSegger	Automated MATLAB-based trainable cell segmentation, fluorescence quantification and analysis suite, well-suited for high-throughput time lapse fluorescence microscopy of *in vivo* bacterial cells
`gateTool`	MATLAB-based image cytometry tool, designed to be part of the same complete package as *SuperSegger*, but able to be used independently; intakes *Clists* for plotting, gate-setting through interactive plots, and other analysis
Clist	Matrix summary of single-valued cell descriptors for all cells
3D Clist	3D matrix summary of cell descriptors for all cells, throughout the cell-cycle
Single Cell Towers	Time-lapse images of a single cell, stacked in chronological order, with consistent pole orientation
Consensus Images	Mean cell-cycle dependent fluorescence localization from sets of single cell towers, spatially scaled to the same cell dimensions

Table A.3: **Summary of *SuperSegger* and its analysis features.**

A.5 Clist structure

All *Clist* information is collected in a single structure-array; it possesses structure-array fields for holding single-valued and time-dependent *Clist* data, cell descriptor definitions, and gates. These fields are defined as follows:

- `data`: stores single-valued measurements for all cells; each row represents an individual cell tracked through the time-course and the columns represent a subset of the > 70 cell descriptors.

- `def`: defines the `data` column number where each cell descriptors is stored.

- `data3D`: stores the time-dependent *3D Clist* data; it contains measurements for a subset of cell descriptors at each frame. The matrix contains Not-a-Number (NaNs) in frames where the cell does not exist.

- `def3D`: defines the `data3D` column number where each time-dependent cell descriptor is stored.

- `gate`: holds information about which cell descriptors have been gated (stored as field `x`) and what the gate values are (stores as field `ind`).

A.6 Clist definition fields

As shown in Section II, the `def` and `def3D` fields hold definitions of the cell descriptors stored in `data` and `data3D`, respectively. (For all fluorescent focus cell descriptors, FocusX(Y) refers to focus number Y found in channel X, and FluorZ refers to fluorescence channel Z.) Below are the `def` field definitions for the *Clist*:

1. Cell ID
2. Region num birth
3. Region num death
4. Cell birth time
5. Cell death time
6. Cell age
7. Old pole age
8. Error frame
9. stat0
10. Long axis (L) birth
11. Long axis (L) death
12. Short axis birth
13. Short axis death
14. Area birth
15. Area death
16. Region score birth
17. Region score death
18. X position birth
19. Y position birth
20. Fluor1 sum
21. Fluor1 mean
22. Fluor2 sum
23. Fluor2 mean
24. Num of neighbors
25. Region gray value
26. Focus1(1) long axis birth
27. Focus1(1) short axis birth
28. Focus1(1) score birth
29. Focus1(1) intensity birth
30. Focus1(2) long axis birth
31. Focus1(2) short axis birth
32. Focus1(2) score birth
33. Focus1(2) intensity birth
34. Focus1(2) long axis birth
35. Focus1(3) short axis birth
36. Focus1(3) score birth

37. Focus1(3) intensity birth
38. Focus1(4) long axis birth
39. Focus1(4) short axis birth
40. Focus1(4) score birth
41. Focus1(4) intensity birth
42. Focus1(5) long axis birth
43. Focus1(5) short axis birth
44. Focus1(5) score birth
45. Focus1(5) intensity birth
46. Focus1(1) long axis pole align
47. Focus1(1) long axis norm pole align
48. Focus1(1) long axis normalized
49. Focus1(2) long axis normalized
50. Focus1(3) long axis normalized
51. Focus1(4) long axis normalized
52. Focus1(5) long axis normalized
53. Focus1(1) short axis normalized
54. Focus1(2) short axis normalized
55. Focus1(3) short axis normalized
56. Focus1(4) short axis normalized
57. Focus1(5) short axis normalized
58. Focus1(1) gaussian fit width
59. Focus1(2) gaussian fit width
60. Focus1(3) gaussian fit width
61. Mother ID
62. Daughter1 ID
63. Daughter2 ID
64. dL max
65. dL min
66. L death / L birth
67. Fluor1 sum death
68. Fluor1 mean death
69. Fluor2 sum death
70. Fluor2 mean death
71. Focus1(1) long axis death
72. Focus1(1) short axis death
73. Focus1(1) score death
74. Focus1(1) intensity death

75. Focus1(2) long axis death
76. Focus1(2) short axis death
77. Focus1(2) score death
78. Focus1(2) intensity death
79. Focus1(2) long axis death
80. Focus1(3) short axis death
81. Focus1(3) score death
82. Focus1(3) intensity death
83. Focus1(4) long axis death
84. Focus1(4) short axis death
85. Focus1(4) score death
86. Focus1(4) intensity death
87. Focus1(5) long axis death
88. Focus1(5) short axis death
89. Focus1(5) score death
90. Focus1(5) intensity death
91. Focus1(1) gaussian fit width death
92. Focus1(2) gaussian fit width death
93. Focus1(3) gaussian fit width death
94. Long axis/Short axis birth
95. Long axis/Short axis death
96. Neck width
97. Maximum width
98. Cell dist to edge
99. Growth Rate

Below are the `def3D` field definitions for the *3D Clist*:

1. `Cell ID`
2. `Long axis (L)`
3. `Short axis`
4. `Area`
5. `Fluor1 sum`
6. `Fluor1 mean`
7. `Fluor2 sum`
8. `Fluor2 mean`
9. `Focus1(1) long axis`
10. `Focus1(1) short axis`
11. `Focus1(1) score`
12. `Focus1(1) intensity`
13. `Focus1(2) long axis`
14. `Focus1(2) short axis`
15. `Focus1(2) score`
16. `Focus1(2) intensity`
17. `Focus1(1) long axis pole align`
18. `Focus1(1) long axis norm pole align`

The cell descriptors included in both the *Clist* and *3D Clist* are easily customizable. Users may do this by editing the *SuperSegger* script `trackOptiClist.m`, or simply by using the `add` and `add3d` commands of the `gateTool` (See Table A.4).

A.7 `gateTool` *Implementation*

The `gateTool` function allows the user to manage all *Clist* modification and visualization by passing commands through a single `gateTool` function. Table A.4 gives a description of all `gateTool` commands for modifying the *Clist*; this includes options such as merging *Clists*, loading data from the *Clist*, and setting gates. Table A.5 gives a description of all `gateTool` commands for viewing the *Clist* data; this includes options such as producing histograms, dot plots, KDE plots, and setting axis scaling. Finally, Table A.6 gives a description of all additional `gateTool` commands for such options as saving the *Clist* and displaying the *Clist* channel definitions. To use these commands, the user simply passes in the *Clist* (or a set of *Clists*), along with the command and any accompanying arguments (i.e. `[output] = gateTool(`*`Clist`*`, 'command', argument);`

). Several commands can (and some commands *must*) be used together in a single call to `gateTool`. A tutorial for using the `gateTool` can be found at `http://mtshasta.phys.washington.edu/website/gatetoolTutorial/sample.html`.

If you prefer to use a GUI's, you can open an easy-to-use `gateTool` GUI using the script `gateToolGUI`. An image of this GUI is shown in Figure A.11.

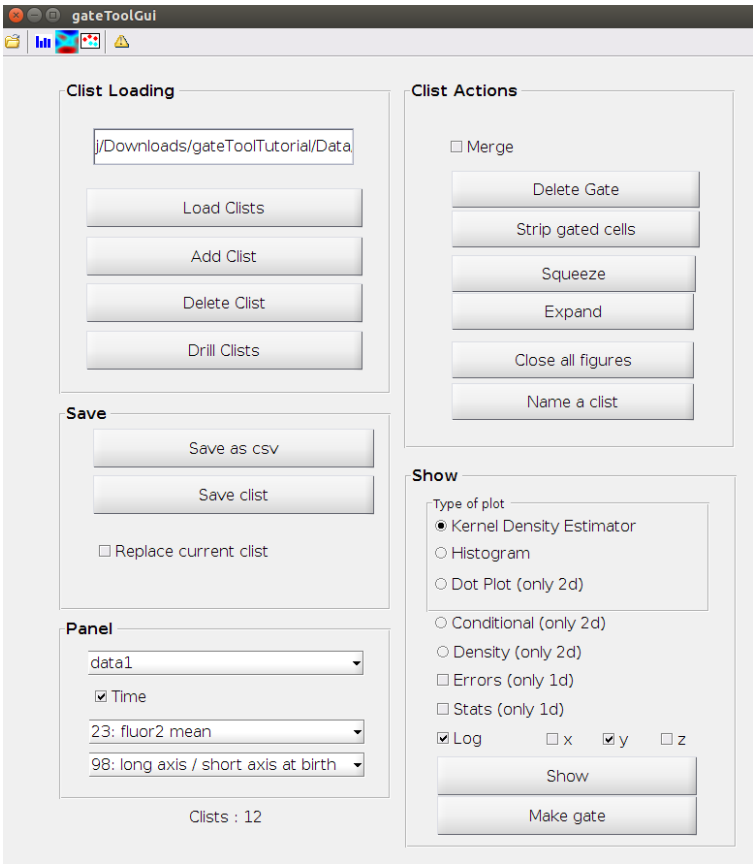

Figure A.11: **Image of the `gateTool` GUI.**

Command	Arg(s)	Description
'merge'		Merge and output all input *Clists* into a single *Clist*
'strip'		Remove all un-gated cells from the *Clist* data
'squeeze'		Treat all *Clist* entries identically
'expand'		Treat all *Clist* entries as different conditions
'name'	name	Provide a label for a *Clist* (appears in plot legends)
'color'	cc	Provide a color for a *Clist* (used for plotting *Clist* data)
'add'	data, name	Add a field (name) to the *Clist* data, using input values (data)
'add3d'	data, name	Add a field (name) to the 3D *Clist* data, using input values (data)
'get'	ind	Output data from data column ind
'make'	ind	Gate all input *Clists* on index (indices) in ind. (ind is a vector with either 1 or 2 elements.)
'getgate'	ind	Outputs the indexed gate

Table A.4: ***Clist* modification commands.**

Command	Arg(s)	Description
'show'	ind	Make a figure for viewing the *Clist*. ind is either an index or pair of indices specifying what is to be visualized. If no ind is passed, all gates are displayed on a single *Clist* input
'time'	ind	Make a temporal plot of single cell dynamics for one index
'hist'	ind	Make a 1 or 2D histogram (depending on dimension of ind)
'dot'	ind	Make a 2D dot plot (dim. of ind must be 2)
'kde'	ind	Make a 1 or 2D KDE (depending on dimension of ind)
'bin'	bin	Set binning for hist, kde plots. bin: (1D) scalar is number of bins, vector is set of bin centers, (2D) vector is set of bin numbers for each dim, cell array is two vectors of bin centers
'den'		Normalize hist,kde plots to show probability density
'cond'		Normalize hist, kde plots to show conditional probability density
'rk'	rk	Set radius of the Gaussian kernel for kde plots
'rm'	rm	Set radius of the point mask (size of plotted point) for kde plots
'mult'	mult	Set resolution for kde plots (i.e. set number of pixels in image)
'err'		Show error in 1D hist, kde plots
'inv'		Invert 2D hist, kde plots for printing
'stat'		Show statistics for a 1D show command
'log'	axes	Set of axes to set with log scales. axes = [1,2,3] will set x, y and color axes to log scale
'no clear'		Do not run clear figure (clf) before drawing
'new fig'		Draws new figures for each input *Clist*

Table A.5: ***Clist* visualization commands.** Note: these commands do not modify the *Clist*. All commands in this table must be used in conjunction with the 'show' command. All commands specifying aspects of 'kde' and 'hist' plots must be used in conjunction with the 'kde' and 'hist' commands.

Command	Argument(s)	Description
`'def'`		Show all *Clist* channel definitions at the command line
`'def3D'`		Show all *3D Clist* channel definitions at the command line
`'xls'`	`filename`	Export an excel document with *Clist* data (Excel must be installed)
`'csv'`	`filename`	Export a csv document with *Clist* data
`'save'`	`filename`	Save *Clist* to a .mat file
`'units'`	`units`	Set a multiplier for the data to set desired units
`'drill'`		Use recursive loading through a directory tree to any level

Table A.6: **Additional commands.**

A.8 Strains and data collection

The analysis for this work was applied to data taken from a variety of new and previous studies. A list of all strains used is provided in Table A.7. Data for the "Length as a proxy for cell age" and "Overexpression phenotypes" experiments were initially collected for previous studies (respectively, [21] and [20, 27]). For all experiments, slides were prepared using pads of 0.2% LM- agarose (Cat. no. 16520-020). Cells are spotted onto the pads, covered with a coverslip and sealed using VaLP (1:1:1 vaseline, lanolin and paraffin). Imaging was completed using a Nikon Ti-E inverted wide-field fluorescence microscope with a large format sCMOS camera (Andor NEO) and controlled by NIS-Elements. Samples were kept at 30 degrees C throughout the imaging process using an environmental chamber. Image processing and analysis was completed using custom MATLAB (Natick, MA) software

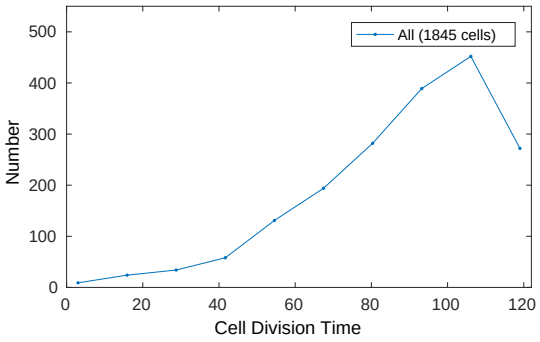

Figure A.12: **Histogram of division times for cells in "Length as a proxy for cell age" experiment.**

Bacterium	Strain	Vector	Experiment
A. baylyi	ADP1		Morphology gating
E. coli	AB1157	pZA12-GFP	Morphology gating
E. coli	MG1655	pALA2705	Length as age proxy
E. coli	K-12	ASKA	Overexpression phenotpyes
E. coli	MG1655		Cell-cycle duration

Table A.7: **Strains.**

A.9 Kernel Density Estimates (KDEs)

Kernel density estimates provide a binning-independent, smoothed representation, or estimate, of the probability density of data. The histogram of values is normalized to provide the probability density, and the resulting values are convolved with a Gaussian kernel, smoothing the resulting values to prevent artifacts from binning. Two example 2D KDE plots are provided in Fig. A.13.

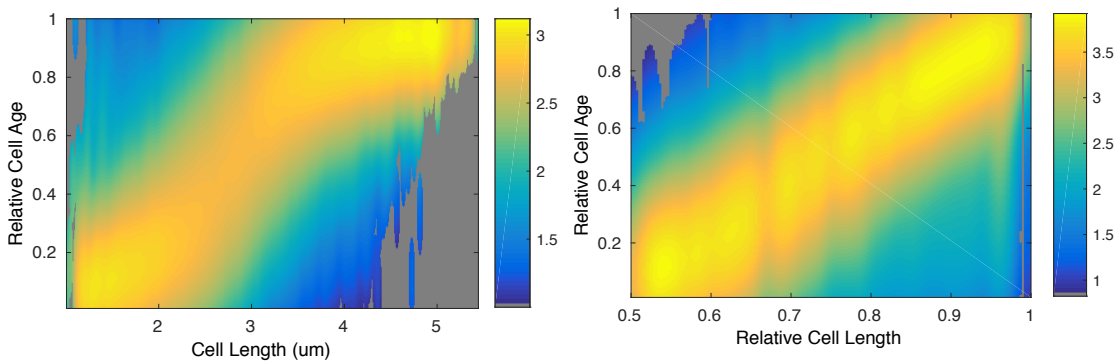

Figure A.13: **Conditional Probabilities for Relative Variables.** For the "Length as a proxy for cell age" experiment, we provide the conditional probabilities for cell length vs relative cell age (left) and for relative cell length vs relative cell age (right). Here "relative" means relative to the end-of-cell-cycle value (i.e. length and age at the time of cell division.

A.10 Computation of p-values for the determinants of growth rate

To compute the p-values for the different determinants of growth rate, we used the non-parameter Kolmogorov–Smirnov Test to compare the distributions and test whether they were generated from the same PDF. We used the MATLAB implementation `kstest2`. Due to the small number of determinants tested, we did not apply a Bonferroni correction to correct for multiplicity.

A.11 Estimation of mutual information

The mutual information is defined:

$$M(X,Y) \equiv \mathbb{E}_{\substack{X,Y \\ p(\cdot,\cdot)}} \log \frac{p(X,Y)}{p(X)\ p(Y)}, \tag{A.8}$$

where p are the pdf and joint pdf distributions for random variables X and Y. To estimate the probability distributions, we define the empirical probability distribution using a Kernel Density Estimate:

$$\hat{p}(x|x^N;\sigma) \equiv N^{-1} \sum_{i=1}^{N} K(x|x_i,\sigma), \tag{A.9}$$

where K is a Gaussian kernel centered on training observation x_i and σ is the kernel width. The leave-one-out-cross-validated mutual information estimate is:

$$\hat{M}(x^N, y^N; \sigma) \equiv N^{-1} \sum_{i=1}^{N} \log \frac{\hat{p}(x_i, y_i | x^{\neq i}, y^{\neq i})}{\hat{p}(x_i | x^{\neq i})\ \hat{p}(y_i | y^{\neq i})}, \tag{A.10}$$

where $x^{\neq i}$ and $y^{\neq i}$ refer to the data set without the ith observation. Finally we maximize M with respect to the kernel density widths with respect to X and Y.

Descriptor	P-value
Phase Brightness	0.901
Dist. to Colony Edge	0.857
Number of Old Poles	0.012
Cell Length at Birth	0.026

Table A.8: **P-values for growth rate distributions gated by cell descriptor** P-values are calculated for the growth distributions in Fig. 4E. For each cell descriptor, the growth rate distributions for the highest (+) and lowest (-) subpopulations are compared.