IMAGE SEGMENTATION AND MODELLING OF HOST-PATHOGEN DYNAMICS OF *SALMONELLA*

by

Taylor Dunn

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2016

Table of Contents

List of Tables

List of Figures

Abstract

The pathogenic bacterium *Salmonella enterica* has recently been shown to hyper-replicate (HR) within the cytoplasm of epithelial cells. To better understand the mechanisms of HR, time-lapse fluorescence microscopy images, taken by our experimental collaborators, of *Salmonella* infecting HeLa cells were segmented. We have developed a robust approach to identifying the presence of the HR phenotype cell-by-cell and, across three independent experiments, found that HR fraction is independent of bacterial load.

Related to *Salmonella* invasion dynamics, but separate from the hyper-replication phenomenon, a mean-field model of infection was developed. The model involves a set of coupled differential equations governed by rates of interactions between a population of hosts and pathogens. Through fitting the model to data from previously published experiments, we have demonstrated that the *Salmonella*-HeLa system may be simply describable with time-independent rates, regardless of experimental conditions.

List of Abbreviations and Symbols Used

GMM Gaussian mixture model

HPI hours post-infection

HR hyper-replication

LAP linear assignment problem

MOI multiplicity of infection

PRR pattern-recognition receptor

SCV *Salmonella*-containing vacuole

Sif *Salmonella*-induced filament

SPI *Salmonella* pathogenicity island

T3SS type III secretion system

Acknowledgements

First and foremost, I would like to thank my supervisor Andrew Rutenberg for the opportunity of working on this project, and for all the opportunities that came with it. Thank you for your support, guidance and analogies that made sense most of the time. Thank you especially for your patience and your (sometimes brutal) honesty which made me a better physicist. I would also like to thank my supervisory committee, Laurent Kreplak and Manfred Jericho, for their helpful feedback and discussion.

Much of this work was made possible by Jennifer Fredlund and the lab of Jost Enninga at l'Institut Pasteur. I owe Jenn many thanks, not only for all of her laboratory work, but for the numerous Skype chats and emails of discussion on the latest results. Thank you to Jost, Virginie, Camille and Laura for their generous hospitality during my visit to Pasteur, and for allowing me to share my work.

Finally, thank you to my family and friends. Thank you to my parents for their love and unwavering support, and for their many attempts to understand what exactly my research is about. To my girlfriend Kristen, thank you for motivating and inspiring me daily. Your love and encouragement made these past two years possible. Thank you especially for reminding me of what's really important and to take a break once in a while.

Chapter 1

Introduction

1.1 Host-Pathogen Interactions

Host-pathogen interactions encompass a variety of phenomena that occur at the molecular level, the single cell level, the single organism level and the population level. From the Greek *pathos*, meaning "suffering", pathogens are agents of disease and include species of bacteria, viruses and fungi. While many of these microorganisms (microbes) are harmless, or even beneficial to their host, infectious disease is a leading cause of death among humans worldwide [1].

First contact between a host and pathogen involves adherence to an epithelial surface, followed by colonization or penetration to replicate within the tissue [2]. In vertebrates, this process is often cut short due to host innate immunity, which involves identification via germline-encoded pattern-recognition receptors (PRRs) and recruitment of different immune cells, such as macrophages, dendritic cells, and neutrophils [3]. When this fails to eliminate an invader, the adaptive immune system acts as the second line of defense. The adaptive immune system is a more complex host defense that is initiated by the presence of foreign proteins called antigens (such as toxins secreted by bacteria). These antigens bind to receptors on T and B lymphocytes which respond to the specific pathogen and also produce memory cells to improve efficiency against the pathogen in the future [4, 5].

Highly virulent pathogens adopt complex strategies to subvert cellular defense mechanisms and proliferate in their host [6]. Antimicrobial research has made great strides in the last half of the 20th century to provide a variety of treatment options for

most infectious diseases [7]. In recent years, however, the tendency for pathogens to adapt to antimicrobial drugs (called antimicrobial resistance) has threatened to render commonly used treatments ineffective [8, 9, 10]. Understandably, this phenomenon has raised the alarm on a global scale, demanding cooperation from the public, health professionals and policy makers to stem the flow of drug-resistance through more careful handling and distribution of antimicrobials [11]. Tackling the problem from the research and development side will require further insights into host-pathogen interactions to establish novel approaches to treatment.

If not eliminated, pathogens can replicate in a variety of compartments [12]. Extracellular pathogens may occupy interstitial space, blood or lymph, or simply remain on the epithelial surface [2]. Foreign invaders in this area are highly susceptible to phagocytosis and thus must evolve the means to avoid engulfment. One example of an extracellular bacterium is *Vibrio cholerae* which colonizes the small intestine and secretes toxins that lead to the disease cholera [13]. Conversely, intracellular pathogens invade and replicate inside of host cells, and can be further subdivided into two major types based on their location: vacuolar and cytoplasmic [14]. The majority of intracellular pathogens enter via endocytosis and replicate inside a vacuole composed of host-derived membranes [15, 16]. To avoid intracellular autophagy (or more specifically xenophagy), pathogen-derived proteins modify the internalization vacuole to prevent or delay fusion with lysosomes (organelles that contain enzymes for degradation) [17]. Alternatively, viruses and certain bacteria bypass this defense mechanism by entering the host cell cytosol directly or by quickly escaping the vacuole (within 30 minutes of internalization) [18]. When it comes to cytosolic bacteria, much less is known compared to their vacuolar counterparts. It is thought that the cytosol is a nutrient-rich environment and protected from certain aspects of host immune killing, such as recognition by circulating antibodies. If this were the whole story, then it might be expected that more pathogens exploit this niche. However, a comparatively

small number of bacteria, including *Listeria monocytogenes* and *Shigella flexneri*, have adapted to thrive in the cytosol [18].

1.2 *Salmonella enterica*

Salmonella enterica is Gram-negative bacterium that infects both animals and humans. There are six subspecies of *S. enterica*, which are further divided into over 2000 serovars classified by their surface antigens [19]. Typically ingested orally via contaminated food or water, the majority of the serovars cause Salmonellosis (acute enterocolitis), while the serovar Typhi causes typhoid fever in humans [20, 21]. These two represent the most common illnesses associated with *S. enterica*, causing an estimated 155,000 [22] and 217,000 [23] deaths per year, respectively. *Salmonella enterica* serovar Typhimurium is among the most common in causing Salmonellosis, and will be the focus of this thesis and the remainder of this section.

The stages of invasion into epithelial cells taken by *Salmonella* are illustrated in Fig 1.1A. To regulate process, *S.* Typhimurium uses two type III secretion systems (T3SS), encoded within *Salmonella* pathogenicity islands SPI-1 and SPI-2, that inject proteins called effectors into the host cell cytosol [24, 25]. These effectors alter host physiology for enhanced bacterial entry and survival. Endocytosis is mainly attributed to first secretion system T3SS-1 and its effectors, which trigger cytoskeletal rearrangments resulting in bacterial internalization and the formation of prominent membrane ruffles [26]. Interestingly, these ruffles have been shown to act as topological obstacles that enhance local stopping, binding and docking [27]. *Salmonella* invasion into epithelial cells is therefore a cooperative effort where the first bacterium to invade facilitates the process for others swimming nearby.

Upon entry within a modified phagosome called the *Salmonella*-containing vacuole (SCV), effectors primarily translocated by the T3SS-2 are required for the maturation and biogenesis of the SCV [28, 29]. One of these effectors is SifA, which induces the formation of Sifs (*Salmonella*-induced filaments) and has been shown to be necessary for vacuolar integrity [30]. When the gene expressing SifA is knocked out, the SCV cannot be maintained and mutant bacteria are released into the epithelial cell cytosol where they grow at a faster rate [31]. This was intriguing discovery because *Salmonella* had been traditionally categorized as a solely vacuolar pathogen, yet replication seemed to be enhanced when cytosolic. Later studies confirmed that a subpopulation of internalized bacteria lyse their nascent ($\leq$ 90 minutes) SCV to escape into epithelial cell cytosol [32, 33, 34]. Knodler et al. [35] coined the term hyper-replication (HR) to describe cytosolic *Salmonella* which replicated to large numbers (initially defined as >50 bacteria/cell) at late times, an effect they observed in approximately 10% of infected epithelial cells (the HR fraction). Subsequent studies have found HR fractions in the range 5–15% [36, 37]. Single-cell analyses by Malik-Kale et al. [36] revealed that net intracellular growth is a product of both vacuolar and cytosolic replication, meaning that *Salmonella* can occupy both replicative niches in a host simultaneously.

There are still many open questions on the topic of *Salmonella* pathogenesis, especially the recently discovered hyper-replicative phenotype. When first exposed to the cytosol, *Salmonella* acquires various autophagy proteins which function to deliver invaders to the lysosome for degradation [38, 33]. Autophagy certainly restricts cytosolic growth [32], but clearly fails in some cases to prevent hyper-replication. In fact, some proteins indicative of autophagy (LC3 and p62) still associate with hyper-replicative *Salmonella* and, surprisingly, may even enhance growth [39] Though the exact mechanisms are a mystery, *Salmonella* appears to greatly alter host cell function through its T3SS and exploit the autophagic machinery for its own benefit.

To gain insight into these molecular-level details, we seek to answer a simple cellular-level question: does a greater bacterial load (more internalized bacteria) result in a greater chance of hyper-replication? As we have discussed, *Salmonella* enter the host in a cooperative manner through ruffle recruitment. Despite this, it has been shown that each *Salmonella* is contained in its own SCV, which divides along with the bacterium [40]. Based upon this finding, we propose a simple model of independent escape, shown in Fig 1.1B, where each bacterium has some chance q of vacuolar escape, autophagy evasion, and replication in the cytosol. The probability of an infected cell harbouring at least one hyper-replicating bacterium is then

$$P(\text{HR}) = 1 - P(\text{no HR}) = 1 - (1-q)^N \tag{1.1}$$

where N is the initial number of SCV-bound bacteria internalized in the host, or the bacterial load.

For *in vitro* invasion assays, the natural way to explore bacterial load is to vary the ratio of inoculating bacteria to host cells (the multiplicity of infection or MOI). Compounded by cooperative attachment to ruffles, the number of invading bacteria per cell quickly rises with MOI, greatly exceeding numbers predicted by independent invasion [27]. We therefore expect a wide range of MOI to result in an even wider range of bacterial loads. Exploring MOI effects on the HR phenotype, and thereby testing the independent escape hypothesis, will be the focal point of Chapter 2.

1.3 Cellular Decision Making

Regardless of genetic or environmental properties, living cells tend to differentiate into specialized subtypes, altering their cellular fate to serve functionally important roles [41]. This so-called cellular decision making is being observed in an increasing variety of cell types, such as the bacterium *Bacillus subtilis* which stochastically differentiates

into resilient, non-growing spores in the face of limited nutrients [42]. In cases like these, differentiation is a preventative measure for survival. Another example is *Escherichia coli*, which appears to hedge its bet against potential antibiotics by having a small subpopulation enter a persister state [43]. Adopting this niche avoids complete eradication of an invasive population, allowing further proliferation within the host [44, 45]. This strategy for survival may be the reason for *Salmonella* Typhimurium's recently discovered bimodal lifestyle.

Related to bacterial decision making is quorum sensing, a system of cell-to-cell signaling that is density-dependent. Bacteria utilizing this system secrete signaling molecules called autoinducers that are detected and bound to receptors on other bacteria [46]. The marine bacterium *Vibrio fischeri* was the first studied case of quorum sensing [47]. As *V. fischeri* divide and their population density (and therefore the density of secreted autoinducers) reaches a threshold, bioluminescence within the organism is activated. Since this discovery, quorum sensing has been shown to regulate many behaviours in a wide variety of bacteria, including virulence genes in *Escherichia coli* and *Staphylococcus aureus* [48]. By communicating via quorum sensing, populations of bacteria exhibit a sort of multicellularity, or a population-level focused behaviour, to alter their fate through density-dependent decision making. For instance, quorum sensing may be used to prevent bacterial populations from growing to levels that are unsustainable in their environment. *Salmonella* Typhimurium has been shown to require the *luxS* gene (responsible for autoinducer production) for up-regulation of certain virulence factors related to the T3SS-1 and required for bacterial internalization [49]. Any disagreement with the independent escape hypothesis of HR expression requires some form of non-random bacterial decision making, and would raise the provocative possibility that *Salmonella* are coordinating expression of the HR phenotype through intracellular communication.

1.4 Quantitative Modelling

Heterogeneity within bacterial populations, resulting from stochastic differentiation and influenced by quorum sensing, lends itself well to quantitative modelling. As Mogilner et al. [50] explain, cell biology encompasses behaviour at many length and time scales, and can become exceedingly complex when considering the breadth of microscopic detail available. The first decision when attempting to model a system is therefore how detailed a model to employ. Regardless of the level of coarse-graining, however, the most important feature of a quantitative model is that it can capture the behaviour of interest, because this indicates that we may have identified all of the relevant players and interactions. The insight gained through exploratory quantitative modelling can improve qualitative understanding of the system, as well as highlight the relative importance of the different aspects of the system.

One of the first success stories in modelling biological systems was a mathematical model of budding yeast cell cycle [51]. The authors distilled data from many independent studies into a consensus model of the cell cycle control mechanisms in budding yeast. They then converted these mechanisms into a set of differential and algebraic equations that determine the time evolution of the system. Out of 131 experimental yeast mutant strains compared, 120 could be explained by the model - certainly a successful implementation and suggests that their model captures essential phenomena and gives good estimates of hard-to-measure reaction rates.

Our model of *Salmonella* invasion into HeLa cells, detailed in Chapter 3, is similar in that it involves a system of ordinary differential equations to model various interactions. The time evolution of the system depends on constant interaction rates, as well as the changing states of host and pathogen populations. Sets of experimental data from independent sources in the literature were fit to parameterize the model and to

give estimate rates of the host-pathogen interactions. From a survey of the literature, there does not appear to be an approach to host-pathogen modelling like this. Other studies have attempted to model protein interactions [52], populations of host organisms [53], comparative genomics [54], and stochastic growth of bacterial populations [55]. At the single-cell and cellular population level, infection dynamics have mostly been modelled very broadly. An early examples comes from Anderson and May [56], who constructed a simple predator-prey model to describe how microbes regulate their survival in hosts. To study the importance of space for infection dynamics, Aguiar et al. [57] employed a spatial lattice model of a general host-pathogen population involving initial infection, reproduction, host death and transmission to neighbours. At the extremely-detailed end of the spectrum, whole-cell modelling attempts to incorporate every molecule and known gene of an organism [58], such as the computational model of the lifecycle of *Mycoplasma genitalium* [59]. Our mean field model is detailed in that it is specific to interactions between *Salmonella* and HeLa cells, but is broad in that it does not attempt to model subcellular processes. This level of detail allows it to be easily extendable to other systems, as it involves many of the interactions and stages of infection shared by other host-pathogen systems.

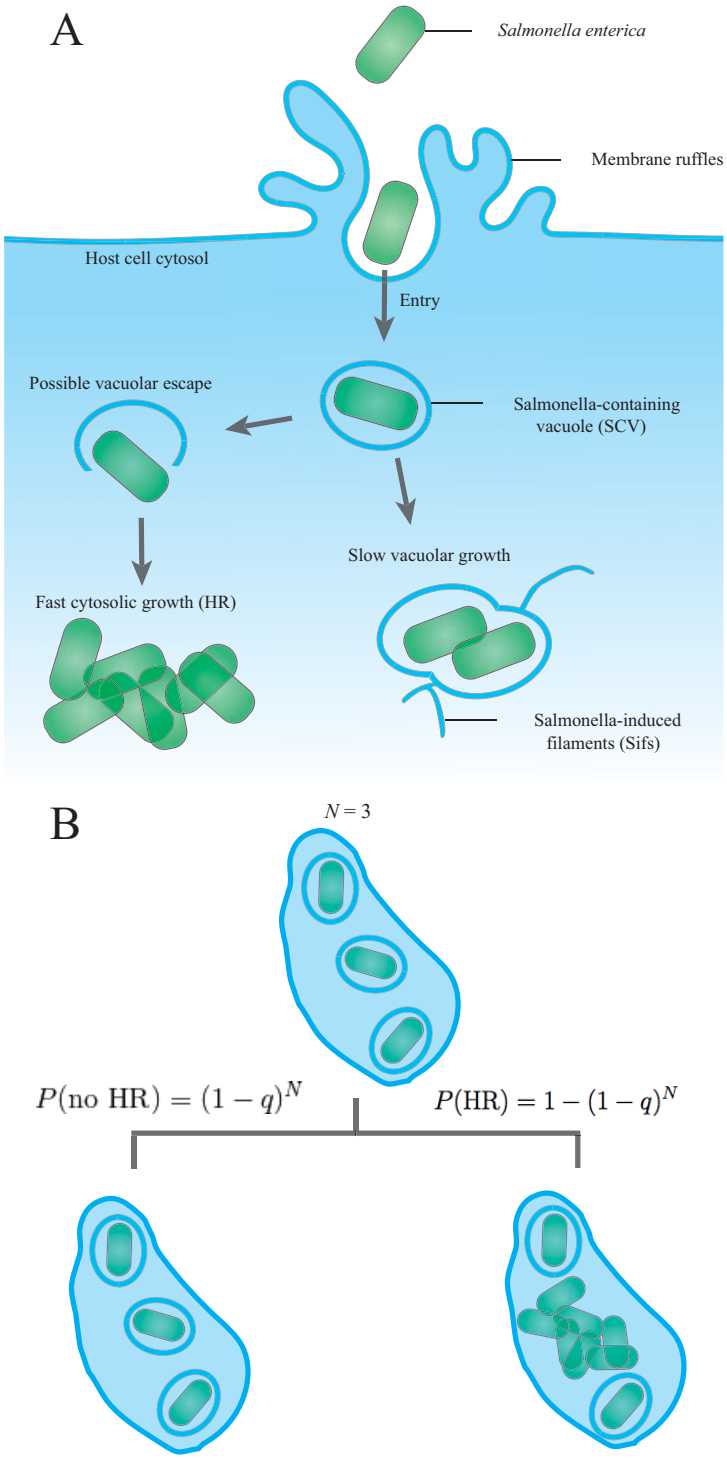

Figure 1.1: (A) Illustration of the stages of invasion by *Salmonella* into epithelial cells, and the branching paths of vacuolar and cytosolic replication. (B) Illustration of the independent escape hypothesis for a bacterial load of $N = 3$. In this context, is the probability of any one bacterium hyper-replicating is q and the probability of an infected cell harbouring HR bacteria is $P(\text{HR})$.

Chapter 2

Image Segmentation and Hyper-Replication

To study the hyper-replication of *Salmonella* Typhimurium, we have worked in collaboration with Dr. Jennifer Fredlund, a former postdoc of the lab of Jost Enninga at the Institut Pasteur. This chapter will first go over the details of Jennifer's experimental methods and our image analysis pipeline. We then introduce an automatic approach of determining HR growth in infected cells and present the estimated HR fraction for a wide range of MOI.

2.1 Timelapse Fluorescence Microscopy

HeLa epithelial cells were seeded at a variety of confluencies in 24-well culture plates one day prior to infection. Wildtype *Salmonella* Typhimurium expressing the DsRed marker gene were then added and allowed to infect for 25 minutes before being washed out. Timelapse images were acquired on an Opera (Perkin Elmer) high throughput confocal microscope at objectives of 10x and 40x, and times ranging from 1.5 to 5.5 hours post-infection (HPI). All images were taken by Jennifer Fredlund and provided to us as TIF files of a single plane with two fluorescent channels: one for DsRed *Salmonella* and the other for Draq5 nuclear staining.

To demonstrate reproducibility, three independent invasion experiments were performed in this way but with slightly different experimental procedures, as detailed in Table 1. For the first experiment, near 100% confluence was achieved and a 10x scope was used to take images of invasion at 25, 50 and 100 MOI. This resulted in a high sample size but difficulty in segmentation and tracking. A lower confluency and

higher objective lens were used in subsequent experiments to rule out segmentation errors biasing the analysis. In addition, these experiments exployed a wider range of MOI (from 25 to 800) to thoroughly explore the effects of bacterial load. To visualize the difference in confluency and scope objective between the three experiments, Fig 2.1 shows example raw images of the nuclear and bacterial channels.

	Experiment 1	Experiment 2	Experiment 3
Confluency	$2 \cdot 10^5$	$2 \cdot 10^4$	$1.5 \cdot 10^4$
Scope objective	10x	40x	40x
MOI	25, 50, 100	25, 50, 100, 200, 400, 800	25, 50, 100, 200, 400, 800
Wells per MOI	3	2-4	2-4
Fields per well	9	17	17
Cells per well	3960 ± 183	682 ± 10	459 ± 8
Time points (HPI)	1.5, 2, 2.5, 3, 3.5, 4.5, 5.5	1.5, 3, 5, 5.5	1.5, 3, 4, 5, 5.5

Table 2.1: Differences between the independent invasion experiments for this study. Confluency is the number of HeLa cells seeded per well. Cells per well is the mean $\pm$ standard error of the numbers of cells segmented and tracked throughout the timelapse.

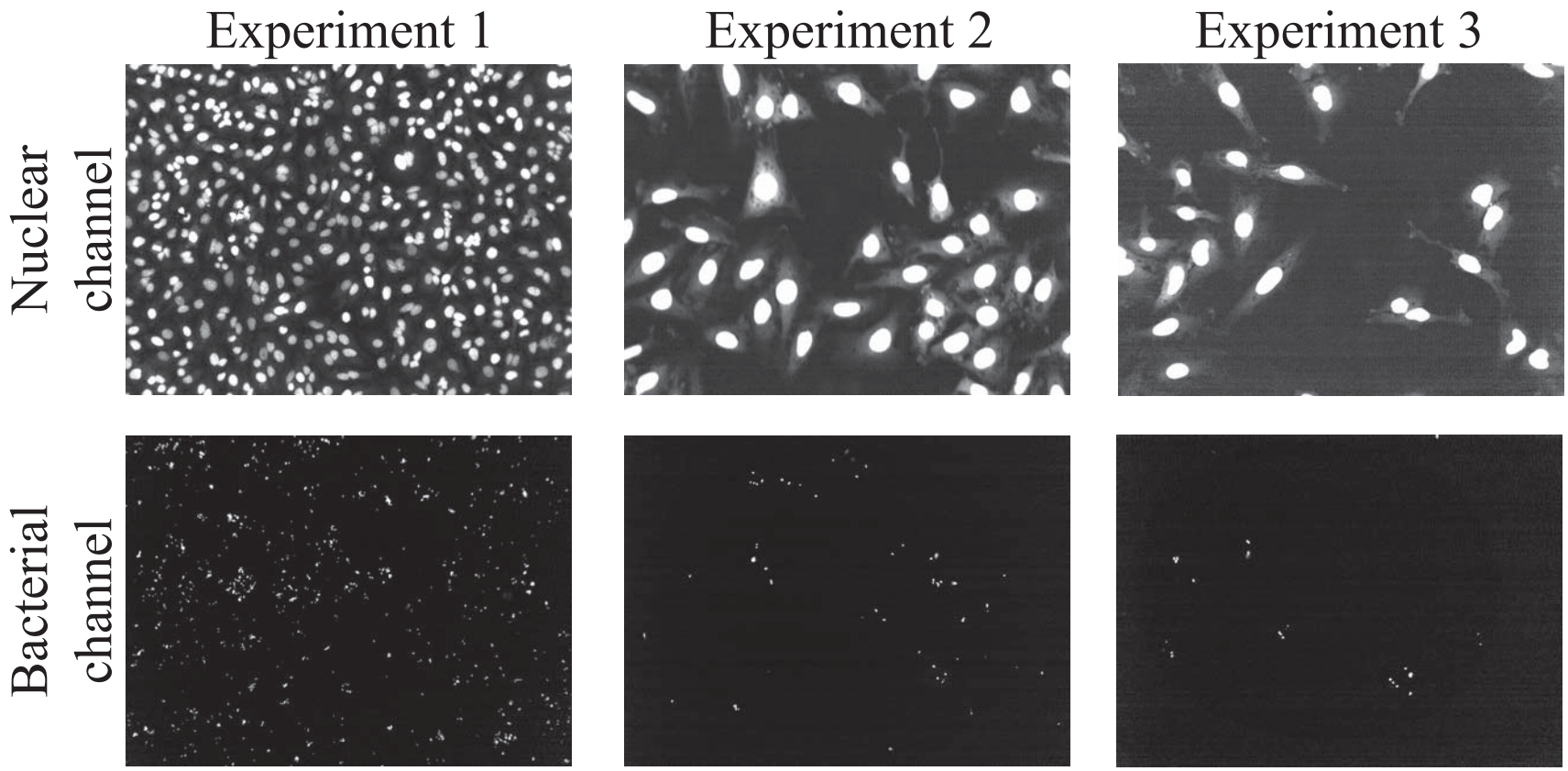

Figure 2.1: Example images of the nuclear and bacterial channels from the *Salmonella*-HeLa timelapse fluoresence microscopy experiments. Each was taken at 1.5 hours post-infection, and is of a single field from a well infected at 25 MOI. Images courtesy of Jennifer Fredlund.

2.2 Image Segmentation and Growth Analysis

The timelapse images taken by Jennifer were analyzed by us using the open source CellProfiler segmentation software [60]. Example bacterial, nuclear and cellular outlines resulting from the segmentation can be seen in Fig 2.2. A significant amount of time was spent learning to use CellProfiler and tuning the modules to produce high quality segmentation. We will give a brief explanation of the segmentation pipeline (set of sequential instructions) here but for a more detailed overview, refer to Appendix A.1.

The first step of our CellProfiler pipeline was to clean up the images and facilitate downstream segmentation. The Smooth module was used to apply a Gaussian filter, which essentially blurs the image slightly to obscure noise and image artifacts smaller than some diameter set by the user. For the bacterial channel, this choice of diameter was chosen carefully (3 pixels) as to not mistakenly conceal bacteria (as small as 5 pixels).

HeLa cell nuclei were detected by their Draq5 staining with the IdentifyPrimaryObjects module employing the three-class Otsu thresholding method [61]. A necessary setting in this module was to enable declumping of close or touching nuclei, so that they are identified separately. Distinguishing by shape proved to be the most effective method due to the characteristic rounded nuclear shape. For experiments 2 and 3, the IdentifySecondaryObjects module (Propagation algorithm) was used to establish cell edges belonging to each nuclei. For experiment 1, on the other hand, the higher confluency, lower resolution images and weaker staining of the cytoplasm caused difficulties in identifying edges. Instead, cells were approximated as Voronoi regions, where all nuclei were expanded until touching one another.

Of fundamental importance when measuring growth at the single-cell level is accurate tracking of host cells throughout the timelapse. An error in tracking could easily

result in an unusually large change in perceived bacterial load, causing misidentification of HR. Cell movement, close packing of monolayers, and high bacterial loads overruning infected cells posed difficulties when linking unique tracks together between frames. To solve these issues, the LAP (linear assignment problem) two-pass method [62] of the TrackObjects module in CellProfiler was used. There are a number of options available for the LAP method, and all must be tailored to the images, but the most important options are the costs of splitting and merging tracks, as well as the the search radius (the maximum distance you expect an object to move between frames). Some of the options are unfortunately unintuitive, so a large amount of trial and error was necessary. For each experiment, tracking parameter knobs were tuned meticulously to find high quality (by eye) tracking of host cell nuclei.

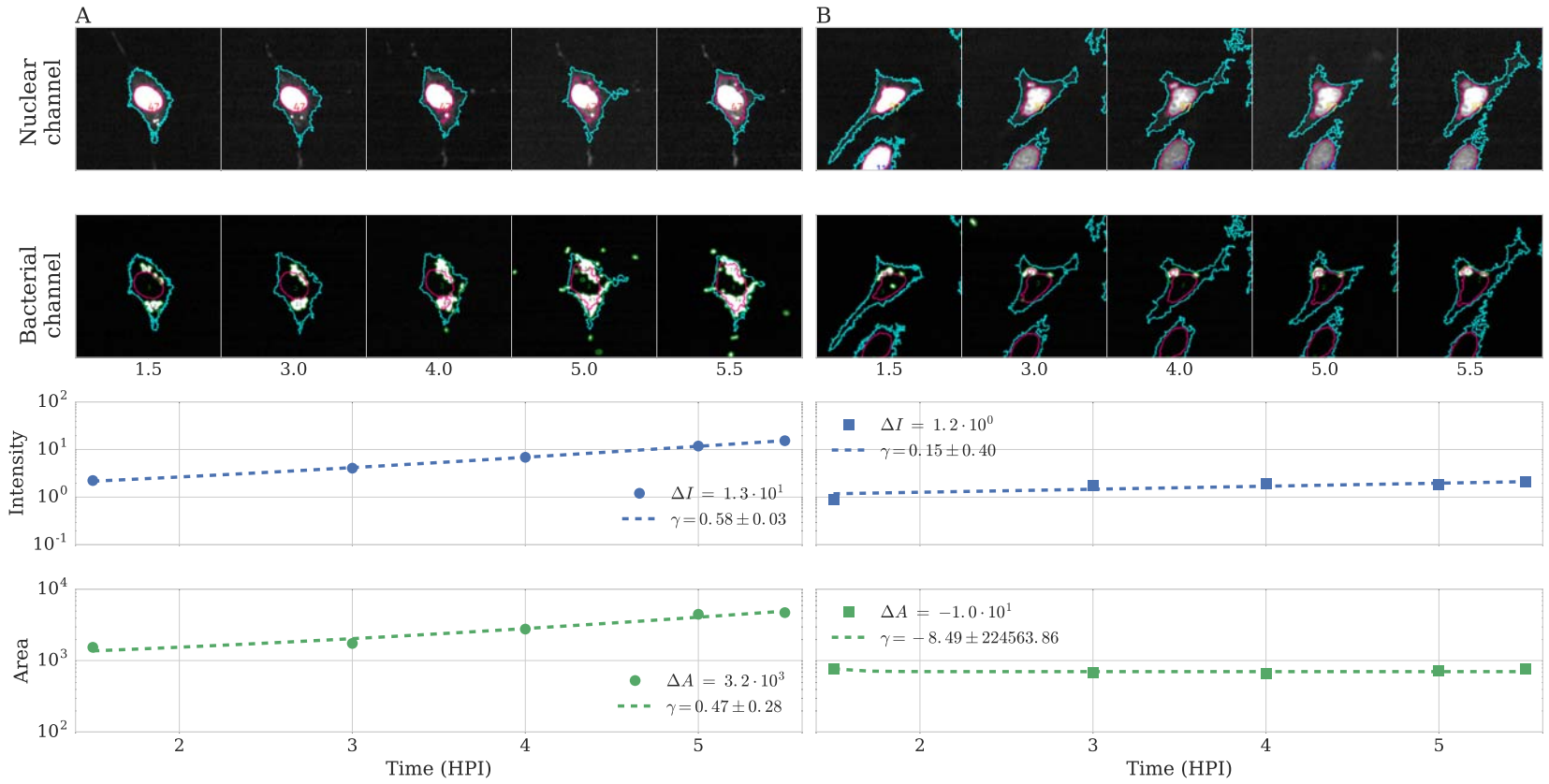

Figure 2.2: **Segmentation examples of slow and fast growth.** HeLa cells infected with *Salmonella* were segmented and tracked over various time points post infection. Growth was measured in terms of bacterial intensity and area, and fit to exponential growth plus a background. Object outline images (top two rows) were output from CellProfiler. Nuclei are outlined in red, cell edges in blue, and bacteria clumps in green. Time is in terms of hours post-infection (HPI). These cells, belonging to experiment 3, were chosen in particular to juxtapose hyper-replication and slow vacuolar growth.

With the levels of resolution, single focal plane, and size of bacterial loads (especially in the case of hyper-replication) used for this study, individual bacteria could not be resolved or counted. DsRed-expressing *Salmonella* were instead identified as clumps of brightness (IdentifyPrimaryObjects module) and assigned to an overlapping cell identified upstream. As a proxy for number of bacteria, both integrated intensity (sum of pixel brightness) and area (total number of pixels) of bacterial objects were measured in infected cells. In many cases, the difference between slow and fast modes of growth is so significant that they can easily be distinguished by eye (compare infected cells in Fig 2.2A and Fig 2.2B for example). However, especially at large initial bacterial load where net intracellular replication may be combination of both vacuolar and cytosolic growth [36], classification of growth can be ambiguous. To avoid the subjectivity and tedium of hand-selecting HR cells, we instead use measurements of net intensity growth ΔI and net area growth ΔA from 1.5 to 5.5 hours post-infection. Additionally, intensity and area growth were fit to an exponential plus non-growing background:

$$I(t) = I_c e^{\gamma} + I_v \tag{2.1}$$

$$A(t) = A_c e^{\gamma} + A_v. \tag{2.2}$$

This is meant to model a starting subpopulation of cytosolic *Salmonella* I_c, A_c growing at the rate γ, along with a non-growing (or relatively slow) vacuolar component I_v, A_v. A least-squares fit was used on all cells which harboured bacteria throughout the entire timelapse. This extra step in analysis served two purposes: the fit captures the entire course of infection as opposed to net growth, and also captures potential bimodality of growth. Any fits which returned greater error than estimated exponent ($\Delta\gamma > \gamma$) were excluded from HR consideration (but still counted in the infected population), as they were very unlikely to contain a high-growth subpopulation I_c,

A_c. Note that this alone isn't sufficient to solely classify HR and non-HR cells, since even slow growth may fit well to exponential growth. The filtering of cells does, however, greatly reduce the number of infected cells to consider, making the features of HR more prominent and easily extracted from a population.

2.3 Identifying Hyper-Replication

If we are to investigate the relationship between bacterial load and HR fraction, then it is important to be clear on how HR is defined and determined. When first discovered, hyper-replicative host cells were categorized as those with >50 bacteria per cell [35]. At 10 hours post-infection (HPI), and at ~50 MOI, 11 ± 4% of infected HeLa cells fell under this definition. The same authors later changed this definition to ≥100 bacteria per cell at 8 HPI, and found 9.2 ± 3.2% of infected cells (at an MOI between 50 and 100) were hyper-replicative [37]. Although the purpose of these studies was not necessarily to determine a concrete HR fraction, but rather demonstrate the existence of a fast-growing subpopulation, these categorizations do not seem appropriate for all situations. For instance, a cell with a large starting population of slow-growing vacuolar bacteria could potentially reach 50 or 100 bacteria at late times and be misidentified as hyper-replicative. A more robust definition would involve net growth of bacteria in infected cells. Malik-Kale et al. [36], for instance, calculated net fold change in area of bacterial load from 2 to 8 HPI, and binned the data into three groups: low (<2 fold increase), moderate (2–10) and high (>10) growth. Taking the high growth subpopulation to be HR, the authors found an HR fraction of 11 ± 3% at 50 MOI.

To date, these definitions of hyper-replication (50 bacteria per cell, 100 bacteria per cell, 10 fold increase in area) have been rather arbitrary. The dividing lines between modes of growth should not be hand-selected, but rather extracted from the distribution of growth. With this in mind, we developed a robust and automatic

approach to classifying distinct modes of bacterial growth within infected cells. Fig 2.3 shows net bacterial intensity and area growth of infected cells at different MOI (from experiment 1), and a clear bimodal separation can be seen in the distributions of ΔI and ΔA. To quantify the separation, the Python package Scikit-learn [63] was used to fit the data to a two-component, two-dimensional Gaussian mixture model (GMM). A GMM is a probability distribution that is a linear combination of two or more differently-shaped and positioned Gaussian distributions. Mathematically, this can be represented as

$$p(x) = \sum_{k=1}^{K} w_k \mathcal{N}(x|\mu_k, \Sigma_k) \tag{2.3}$$

where K is the number of components (Gaussians), w_k are the weights of the components ($0 \leq w_k \leq 1$), μ_k are the means and Σ_k are the covariance matrices. In our case, we expect a Gaussian at small values of ΔA and ΔI (the slow-growing infected cells), and one at large values (the hyper-replicative cells). As previously mentioned, infected cells were excluded from HR determination based on their single-cell growth fit ($\Delta\gamma > \gamma$). This resulted in a reduced weight w_k of the slow-growing class and made it much easier for the fitting routine to separate the data. After this filtration step, the GMM takes as input all of the $x_n = (\Delta A_n, \Delta I_n)$ data points, with only the knowledge that they should be separated into $K = 2$ classes, and computes the maximum likelihood estimates of the model parameters. The resulting probability distribution overlays the data in Fig 2.3.

The two classes are well fit, and show little overlap. Aligning with our expectations, one class (green) shows both larger intensity and area change and we designate it "high growth." The other class (red) shows lower change and we designate it "slow growth."

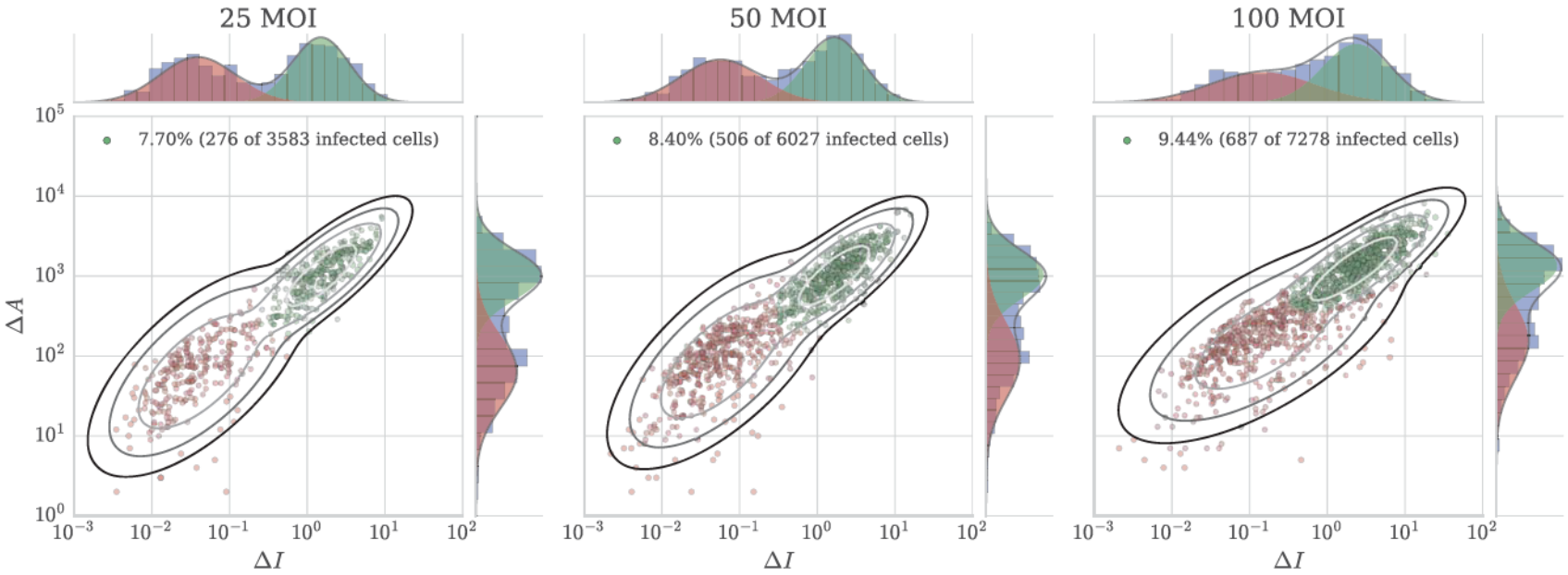

Figure 2.3: **Bimodal growth can be modelled to quantify HR fraction.** Distributions of net intensity and area growth of infected HeLa cells at three different MOI from experiment 1. Overlaying the data is the fit to a Gaussian mixture model. The data is coloured based on its predicted classification: high growth (green) or slow growth (red). Infected cells plotted here were screened based on their least-squares fit to an exponential growth plus background (Eq 2.1). The number of cells which passed the screening ($\gamma > \Delta\gamma$) are 538, 981 and 1248 for 25, 50 and 100 MOI, respectively.

From the model fit, each cell is predicted to belong to the slow or fast-growing class. The percentage of infected cells categorized as fast-growing is our estimate of the HR fraction, and from this preliminary experiment of just three MOI, appears to be increasing slightly or possibly constant in the range 7-10%, consistent with previously reported values [35, 36, 37]. It is worth noting that the separation between the two modes of growth becomes less pronounced with increasing MOI, resulting in greater overlap between the bimodal distributions. This effect likely arises due to variations in population growth rates as MOI and bacterial load increase. Since infected cells can harbour both vacuolar and cytosolic bacteria [36], net intracellular replication may be a combination of both slow and fast growth. If an infected cell has a disproportionately large population of vacuolar bacteria compared to the hyper-replicative population, then the HR growth may be masked when looking at net replication. This is another area where the filtering of infected cells by fitting to exponential growth (fast-growing cytosolic bacteria) plus background (relatively slow-growing vacuolar bacteria) helped to identify the HR phenotype. Regardless, the two-component GMM seems to capture the data even at high MOI. With confidence in our method, the process was repeated for experiments 2 and 3 and the resulting growth distributions and fits can be seen in Fig 2.4.

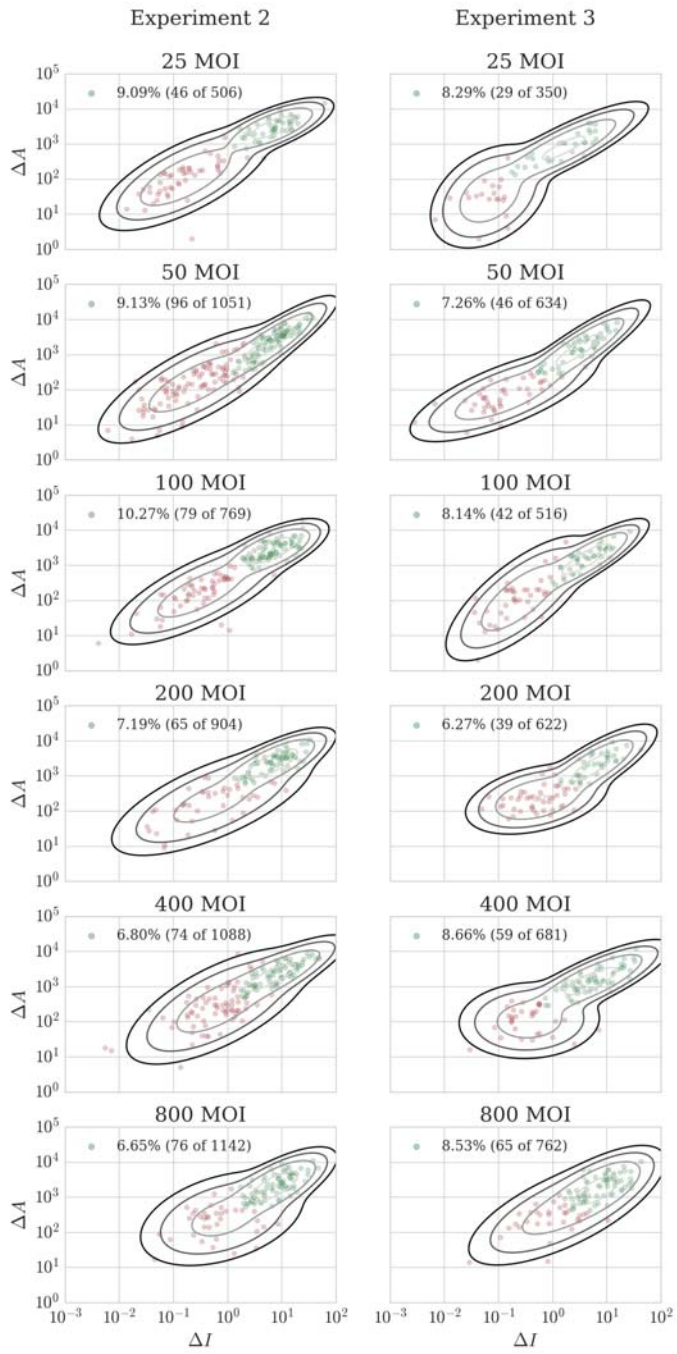

Figure 2.4: **Bimodal growth fitting for experiments 2 and 3.** Scatter plots of net intensity and area growth of infected HeLa cells for a wide range of MOI. Overlaying the data is the fit to our two-class Gaussian mixture model. The data is coloured based on its predicted classification: high growth (green) or slow growth (red).

2.4 HR Fraction and Bacterial Load

Fig 2.5 shows the HR fraction estimation of all three experiments versus infectivity (which correlates with MOI and initial bacterial load). The striking feature of this plot is plain to see: across MOI and between independent experiments, HR fraction appears to be constant around 8%. Consequently, it would seem hyper-replication is independent of bacterial load despite *Salmonella* invading in a cooperative manner that benefits greater numbers. To confirm this, we investigated the change in bacterial load with MOI. In Fig 2.6, we have compared initial bacterial load $I(1.5)$ and $A(1.5)$ and found it to approximately double with a doubling in MOI. Ideally, our first time point would be at an earlier time than 1.5 hours post-infection, but there is evidence to suggest that replication of internalized *Salmonella* does not begin until shortly after this time [37]. We can therefore conclude that, across the range of MOI tested here, the average initial bacterial load doubled multiple times. If it were the case the bacteria escaped and hyper-replicated independently, we would have expected the HR fraction to quickly rise from the baseline bacterial load $N = 1$ to $N = 2$ and beyond. This rules out the simple independent escape hypothesis, which is plotted in Fig 2.5 assuming an escape probability per bacterium of $q = 8\%$.

Early after escape and exposure to the cytosol, *Salmonella* colocalizes with autophagy proteins, such as LC3, which target the bacterium for lysosomal pathogen elimination [38]. This autophagic response is not completely understood currently, but it is believed that dysfunctional autophagy is to be blamed for cytosolic proliferation [64]. This may be a result of the *Salmonella* Typhimurium T3SS disrupting host cell pathways, as it has been shown to do during SCV maturation [28]. Our results, however, seem to indicate that more internalized bacteria does not lead to a greater chance of cytosolic survival and hyper-replication. Thus, a simple explanation for a constant HR fraction is that it is a host cell-dependent phenomena, wherein a

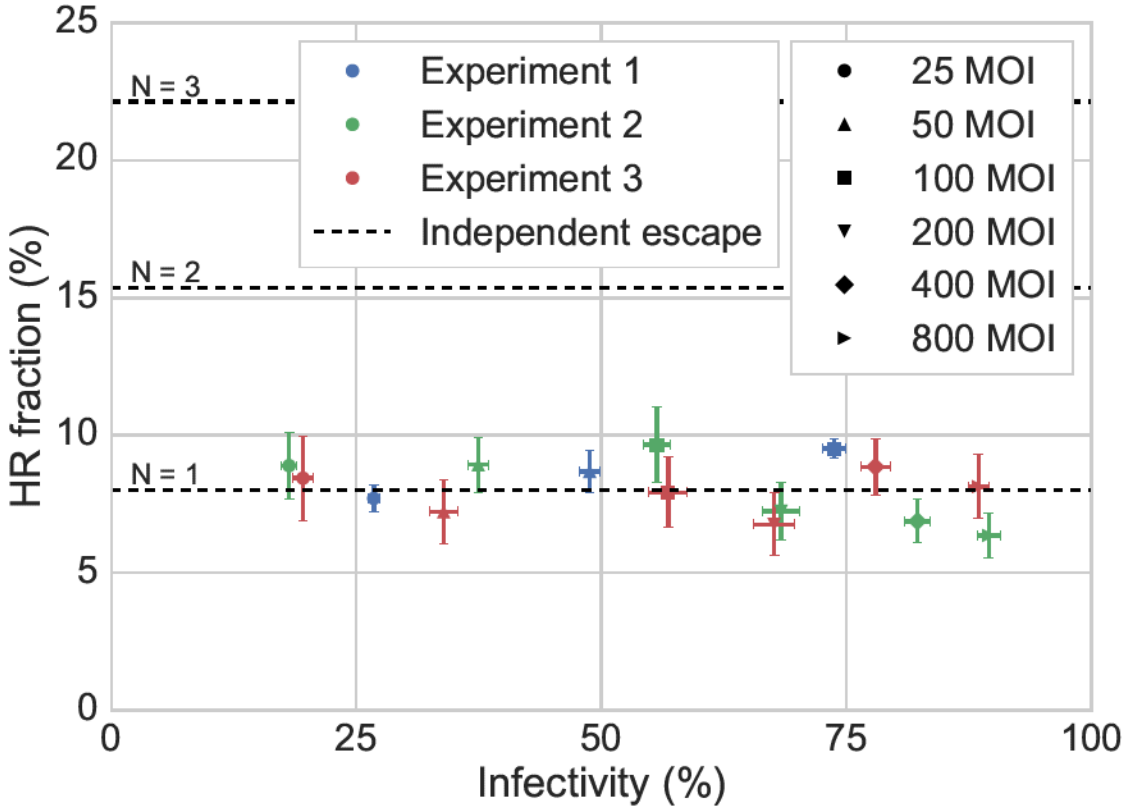

Figure 2.5: **HR fraction is independent of bacterial load.** Three independent experiments of *Salmonella* invading HeLa cells were segmented and tracked at a large range of MOI. HR fraction versus infectivity is plotted (± standard error in the mean between fields), as well as theory lines indicative of independent escape for $q = 8\%$ and initial bacterial loads of $N = 1, 2, 3$.

minority of infected hosts are unable to defend against hyper-replicative *Salmonella*.

A more complex hypothesis is that, for each infected cell, a single bacterium (likely the first to enter) attempts to escape and hyper-replicate with some mean success rate. Perhaps other bacteria receive a "stop" signal via quorum sensing [46] from that initial bacterium as a means to prevent oversaturation of the cytosol, and to hedge their bets against host immune response by occupying multiple intracellular niches. Or maybe cytosolic *Salmonella* can sense something in the host cell cytosol (such as a lack of nutrients or the presence of autophagy proteins) and enters a slow-growing persister state.

To confirm any of these speculations will require further *Salmonella*-HeLa invasion experiments with varying MOI and markers for subcellular localization. LAMP1, for instance, is a protein known to associate with SCV membranes during maturation [28]. By staining HeLa cells for this protein, it can be determined if a bacterium is contained in its SCV [65]. With this information, we can make more informed HR identification by ruling out cells with just vacuolar *Salmonella*, and also investigate growth rates of the subpopulations, instead of net population growth.

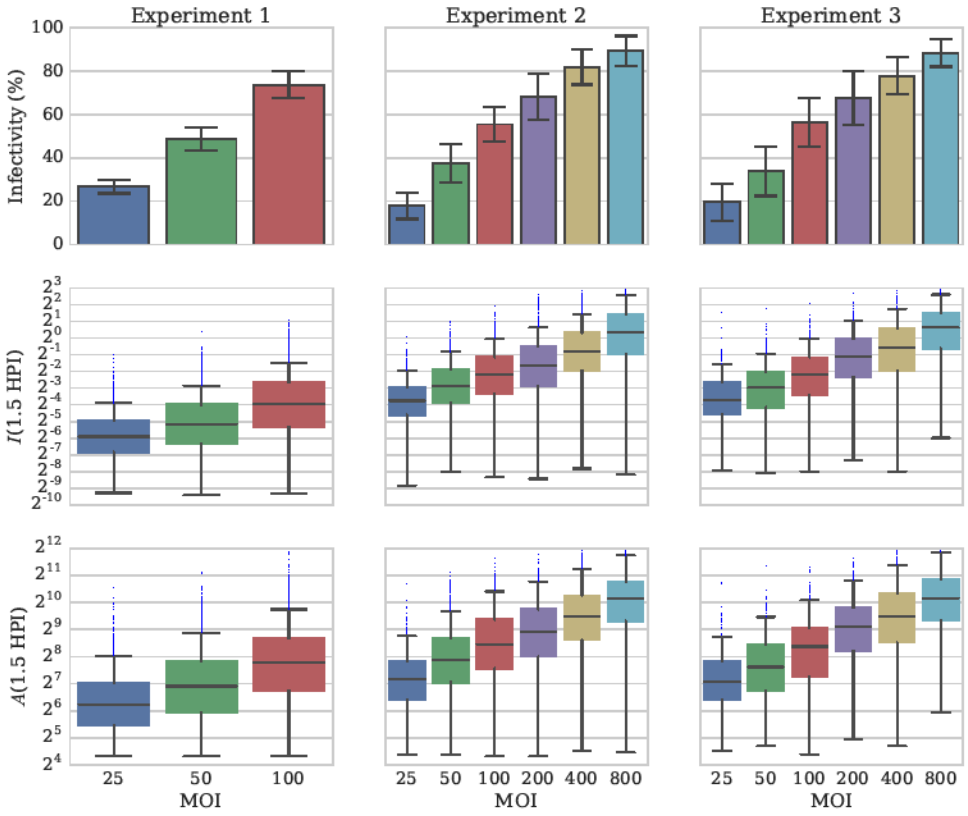

Figure 2.6: **Infectivity and bacterial load vary strongly with MOI.** Three independent experiments of *Salmonella* invading HeLa cells were segmented. Infectivity ($\pm$ standard deviation between fields) and bacterial intensity and area at 1.5 HPI are plotted (whiskers represent 1.5 times the interquartile range). The main take away here is the approximate doubling of bacterial load with doubling MOI.

2.5 Sources of Error

Despite the great care taken to achieve high quality segmentation, artifacts and errors are sometimes unavoidable. For images from experiment 1, which were taken at lower resolution and higher host confluency, it proved difficult to draw accurate boundary lines between cells. This often affected the assignment of bacteria to the appropriate host cell. In Fig 2.7, for example, there are two neighbouring infected cells (labelled 620 and 632 in the nuclear channel) which were correctly separated into unique tracks by the LAP tracking algorithm. However, the bacterial object which appears to belong to the 632 host is right on the boundary of the cells, and at 4.5 HPI the bacteria is assigned to the 620 host, leaving the 623 seemingly uninfected (indicated by the green 0 in the bacterial channel). Initially, our classification of an infected host was one that contained bacteria throughout the timelapse. Transient segmentation errors like these led to miscounting of the actual number of infected cells. It was less of a problem when counting HR cells, since the increased size of the bacterial clumps generally overtook much of the host cell and could be unambiguously assigned to the correct host. Erroneous infectivity counts is a source of systematic error in our calculation of HR fraction (= number of HR cells / number of infected cells), biasing the measurement towards higher values. In an attempt to counteract this effect, we changed our definition of infected hosts to include those infected for most (all but one time point) of the timelapse, so that cells like 632 in Fig 2.7 are correctly deemed infected. This doesn't solve the problem entirely, but certainly helped to suppress the systematics.

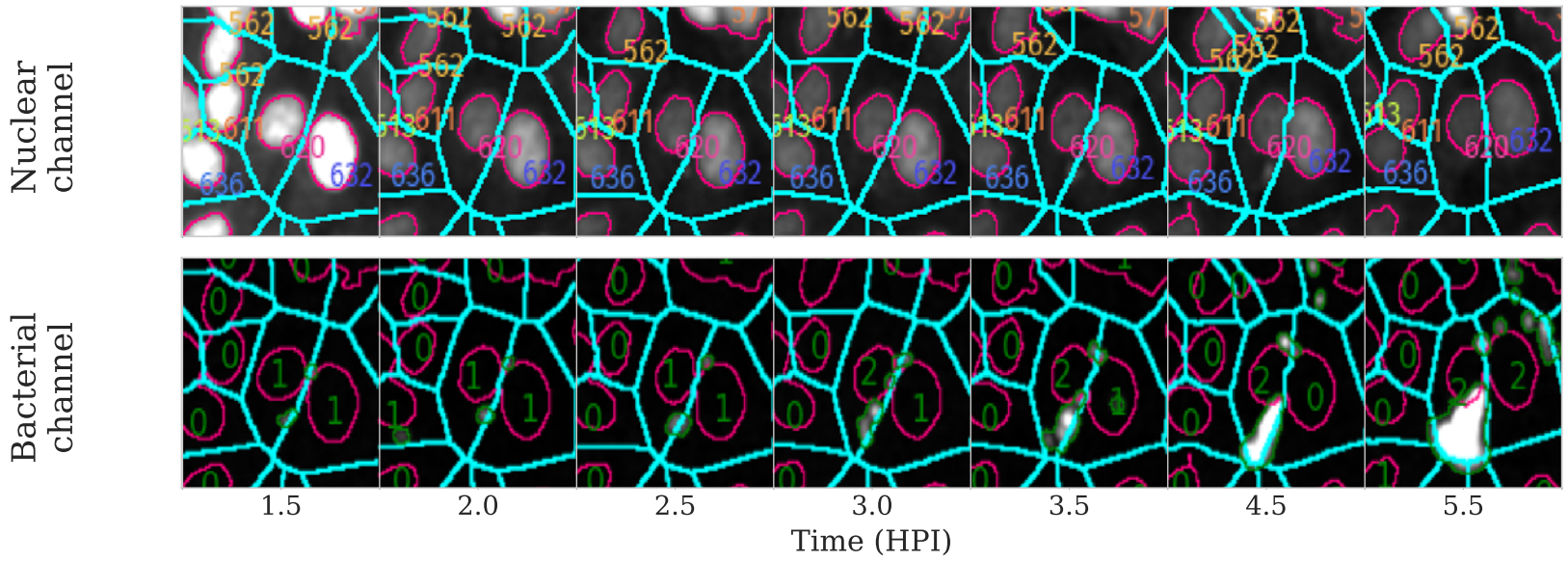

Figure 2.7: Example of a possible segmentation error in our CellProfiler analysis. The cell on the right appears to be infected throughout the timelapse, but at 4.5 HPI the bacteria is assigned to the left cell, leaving the right cell transiently uninfected.

Another source of error could come from the experimental procedure, such as determining the multiplicity of infection. Infecting at a particular MOI is an inexact process. Rarely will MOI be presented with error bars, but some studies tend to give a range in which they expect MOI to fall (Knodler et al. [37], for example). Before being added to the culture plate, bacterial populations are grown in a broth to a specific density (in relation to the host cell density). They are then added equally to each well of the culture plate. This two-step process is done manually and is therefore prone to human error. An error in measurement of the bacterial density in the first step would result in the entire culture plate receiving a lower or higher than expected MOI, on average. The HR fraction versus infectivity plot of Fig 2.5 shows the MOI average, or more specifically the average of every field of every well infected at that MOI with error bars repesenting the standard error in the mean. It is at this level we would expect a density measurement error to influence the MOI (and thereby the measured infectivity), but we have no reason to believe it would be bias the results one way or the other. If an error occurs in adding bacteria equally to each well of the culture plate, then error would propagate between wells. Fig 2.8 shows the HR fraction versus infectivity averaged per well. Comparing the scatter of these points may reveal an MOI-dependence on this well-to-well error. For instance, the left-to-right scatter of the 25 MOI points (circular markers) appears to be minor compared to those at 100 MOI (square markers). Regardless, we again have no reason to believe the error is biased one way or another, and small variations in infectivity do not change our result of a constant HR fraction. As an aside, these well HR fractions were fit to a flat line for each experiment, and were found to fall within error of each other (8.6%, 8.2% and 7.9%).

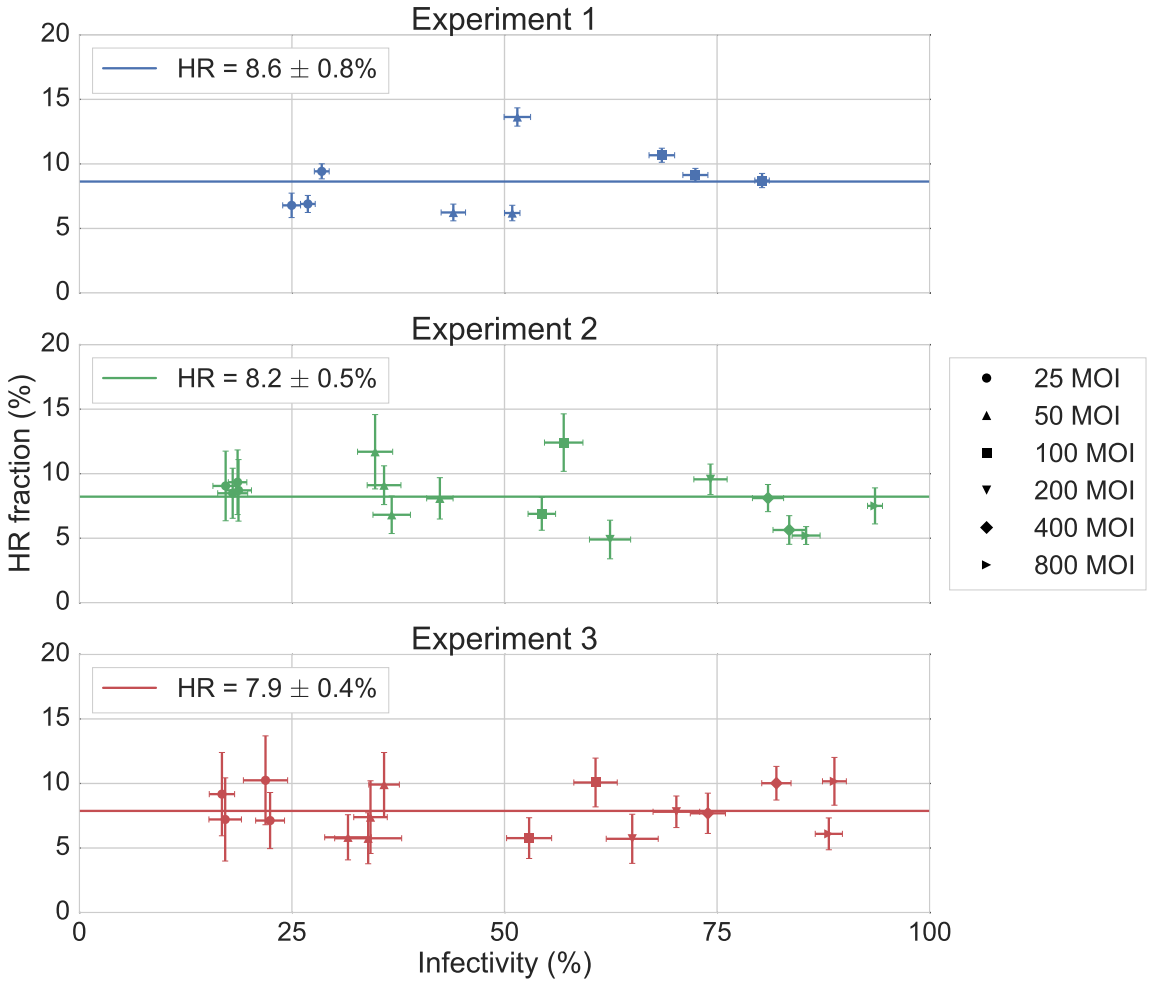

Figure 2.8: Well-averaged HR fraction versus infectivity for all three *Salmonella*-Hela fluoresence microscopy experiments. Error bars are standard error in the mean between fields. The data was also fit to a flat line to give an estimate of the constant HR fraction.

Chapter 3

Mean Field Model of Invasion

The previous chapter examined the relationship between hyper-replication fraction, bacterial load and MOI of *Salmonella*-infected HeLa cells. We found bacterial load to increase with MOI, as expected, but that HR fraction remained constant. As a supplement to our HR study, we have developed a mean field model to explore invasion dynamics and gain insight into MOI-dependence. The purpose of employing a mean field model is to simplify the host-pathogen interactions (attachment, ruffle formation, internalization, escape) into system averages. This has the advantage of being easier to compute than individually tracking and computing pairwise interactions and stochastic processes between cells and bacteria. Our mean field model should be a reasonable approximation at high cell and bacteria counts, which is typical of the laboratory invasion systems we wish to simulate. We will first present the model parameters, including counts, fractions and rates, and then go over the derivations of the dynamical equations. We then introduce data taken from the literature which was used to parameterize the model, and finally give an example of applying the model to further support our analysis of the hyper-replication data.

3.1 Model Parameters

3.1.1 Bacterial and Cellular Subpopulation Counts

Over the time course of invasion by *Salmonella*, a bacterium can be in one of 4 generalized stages: swimming (looking to attach to a host) B, attached to a host cell

B_a, invaded and vacuolar B_v, or invaded and cytosolic B_c. The counts of bacteria in these stages are constantly changing in time, but at early times, generally before the onset of bacterial replication (after 180 minutes) [37]) or host-cell death (varies, but after repication [35]) have begun, the total number will remain constant:

$$B_{\text{tot}} = B(t) + B_a(t) + B_v(t) + B_c(t). \tag{3.1}$$

We can similarly classify the relationship between host cells and bacteria in one of three ways: host cells with no bacteria H, with attached bacteria H_a, with invaded vacuolar bacteria H_v, with invaded cytosolic bacteria H_c, and with any invaded bacteria at all H_x. A host cell may have a combination of attached, vacuolar and cytosolic bacteria at one time, so these classifications are not mutually exclusive. For instance, a cell with attached bacteria may also have invaded bacteria. What we can say is that the total number of host cells is given by

$$H_{\text{tot}} = H(t) + H_a(t). \tag{3.2}$$

A well-studied feature of invasive *Salmonella* is the formation of epithelial cell membrane ruffles. We call the total number of ruffles $R(t)$, the number of host cells with at least one ruffle $H_r(t)$, and the number of bacteria attached to ruffles B_r .

When performing a typical invasion assay experiment, tunable parameters include the multiplicity of infection (MOI), m, and the confluency, c. In terms of the counts presented here, MOI can be expressed as

$$m = B_{\text{tot}}/H_{\text{tot}}. \tag{3.3}$$

The confluency (or confluence) is the proportion of area covered by cells in a culture dish. A confluent monolayer of cells would correspond to 100% confluence, with no

room left to grow. Most often, and in the case of the studies used here to parameterize the model, confluency is instead presented as the number of host cells seeded into the culture dish. For our mean field model, this number is unhelpful as the type/size of plate varies between experiments, so it does not tell us the percentage of area covered by host cells. Consequently, the percentage confluency c must instead be used as a fitting parameter rather than a known experimental value. In terms of an average host cell area A and side length of the square well L, the confluency is

$$c = H_{\text{tot}} A / L^2. \tag{3.4}$$

While this measure of host cell density does not change in time for a typical assay, the swimming bacterial density will decrease in time as *Salmonella* attach and invade. The extracellular and unattached bacterial density is

$$\rho_B(t) = B(t)/L^2. \tag{3.5}$$

3.1.2 Fractions

For a mean field model, it is more appropriate to speak in terms of fractions or ratios of the population To differentiate, we use lower case variables as listed and defined in Table 3.1. This leads to an alternate form of the bacterial density:

$$\rho_B(t) = \frac{B(t)}{L^2} = \frac{bB_{\text{tot}}}{H_{\text{tot}} A/c} = \frac{bmc}{A} = bmc \tag{3.6}$$

where in the last step, we have used the reduced units $A = 1$ (i.e. we measure area in units of average host cell area).

Variable	Definition	Equals
$h(t)$	fraction of host cells without bacteria	H/H_{tot}
$h_a(t)$	fraction of host cells with attached bacteria	H_a/H_{tot}
$h_x(t)$	fraction of host cells with invaded bacteria	H_x/H_{tot}
$h_v(t)$	fraction of host cells with vacuolar bacteria	H_v/H_{tot}
$h_c(t)$	fraction of host cells with cytosolic bacteria	H_c/H_{tot}
$h_r(t)$	fraction of host cells with ruffles	H_r/H_{tot}
$b(t)$	fraction of swimming bacteria	B/B_{tot}
$b_a(t)$	fraction of attached bacteria	B_a/B_{tot}
$b_x(t)$	fraction of invaded bacteria	$b_c + b_v$
$b_v(t)$	fraction of vacuolar bacteria	B_v/B_{tot}
$b_c(t)$	fraction of cytosolic bacteria	B_c/B_{tot}
$\tilde{r}(t)$	ruffles per host	R/H_r
$\tilde{b}_a(t)$	attached bacteria per host	$B_a/H_a = mb_a/h_a$
$\tilde{b}_x(t)$	invaded bacteria per host	$B_x/H_x = mb_x/h_x$
$\tilde{b}_r(t)$	bacteria recruited per ruffle	B_r/R

Table 3.1: List of relevant time-dependent quantities and their definitions.

3.1.3 Rates

Now that we have established what is changing in time, we can introduce the rates at which they change. For a consolidated list, refer to Table 3.2, and to see these rates labelled, refer to Fig 3.1.

A swimming bacterium can either attach to a host cell normally with a rate Γ_a, or be recruited to a previously formed membrane ruffle at Γ_b. Once attached, a bacterium can cause ruffling at Γ_r, invade the host and remain vacuolar at Γ_v, or invade the host and quickly escape into the cytosol at Γ_c. The combined invasion rate is Γ_x, but there is evidence that suggests this rate is limited by bacterial load [66]. To approximate a saturation effect, we define the limited invasion rate

$$\Gamma_x^*(t) \equiv \Gamma_x \left(1 - \frac{\tilde{b}_x(t)}{\tilde{b}_{x,\max}}\right) \tag{3.7}$$

where $\tilde{b}_{x,\max}$ is the maximum number of bacteria that can invade a single host. From

the same study, a limitation on the number of foci of infection (clusters of internalized bacteria) was found. Similarly, we approximate a saturation effect on the number of ruffles (or ruffling sites) on a single infected cell with

$$\Gamma_r^*(t) \equiv \Gamma_r \left(1 - \frac{\tilde{r}(t)}{\tilde{r}_{\max}}\right) \tag{3.8}$$

Both $\tilde{b}_{x,\max}$ and $\tilde{r}_{\max}$ are additional tunable parameters, and can be set to infinity to eliminate limits on invasion and ruffle formation.

An important feature of any host-pathogen system is host-directed killing or xenophagy. A rate of killing was omitted from our model for the simple reason of lack of data available for parameterization. Although xenophagy is not explicitly included, it can be partially reflected in the other rates. For example, Γ_c can be thought of as the rate of successful cytosolic invasion and evasion of host defense. A more detailed model would ideally have a parameter for xenophagy, as it may be dependent on variables such as confluency or bacteria per cell, but this implicit approach may suffice until more data is available.

3.1.4 Dynamical Equations

The following section presents the logic used to arrive at the dynamical equations describing the mean field model.

Variable	Definition
Γ_a	primary attachment rate per bacterial density, per host
Γ_b	ruffle recruitment rate per bacterial density, per ruffle
Γ_r	ruffle formation rate per attached bacteria, per host
Γ_r^*	effective ruffle formation rate per attached bacteria, per host
Γ_v	vacuolar invasion rate per attached bacteria, per host
Γ_c	cytosolic invasion rate per attached bacteria, per host
Γ_x	combined invasion rate per attached bacteria, per host
Γ_x^*	effective invasion rate per attached bacteria, per host

Table 3.2: List of the time-independent and effective rates governing the model.

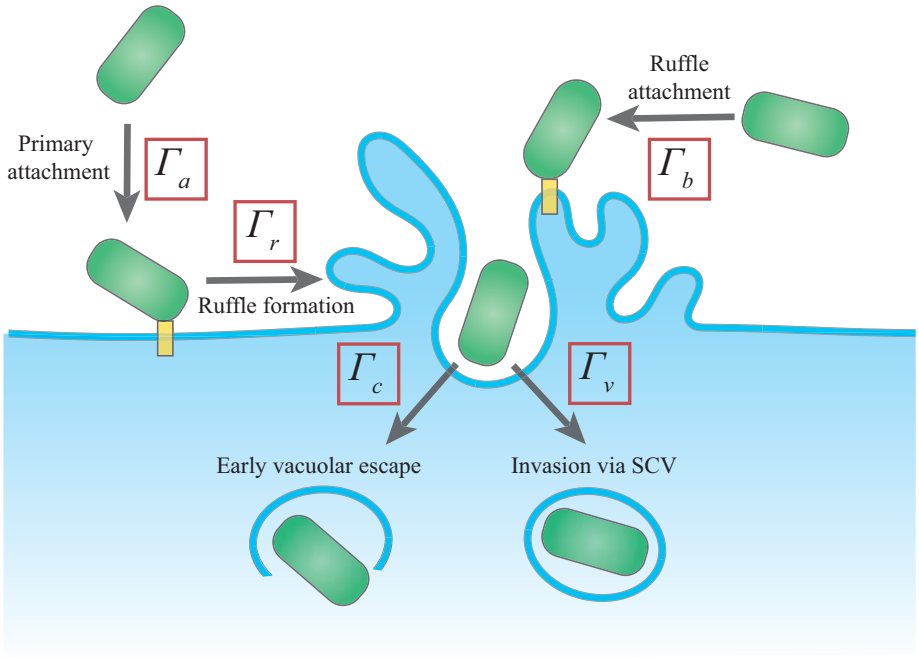

Figure 3.1: Illustration of the early stages of *Salmonella* infecting HeLa cells and the associated rates (in red) used to model the process.

The change in the number of host cells with attached bacteria H_a will only depend on the primary attachment rate Γ_a and not secondary attachment (ruffle recruitment) rate Γ_b because, for a bacterium to attach to a ruffle, another bacterium has already attached to form that ruffle. This will be given by $\dot{H}_a = \Gamma_a H \rho_B$. Using scaled variables, we obtain

$$\dot{h}_a = \frac{\dot{H}_a}{H_{\text{tot}}} = \Gamma_a \frac{H}{H_{\text{tot}}} \rho_B = \Gamma_a h(t) b(t) mc. \tag{3.9}$$

The change in the number of infected cells H_x is controlled by the invasion rates $\Gamma_x = \Gamma_c + \Gamma_v$, and is limited by the number of uninfected cells H and attached bacteria per cell $\tilde{b}_a$. From the unscaled dynamics $\dot{H}_x = \Gamma_x (H_a - H_x)\tilde{b}_a$, we obtain

$$\dot{h}_x = \frac{\dot{H}_x}{H_{\text{tot}}} = \Gamma_x \frac{H_a - H_x}{H_{\text{tot}}} \tilde{b}_a = (\Gamma_v + \Gamma_c)(h_a - h_x)\frac{b_a m}{h_a} = (\Gamma_v + \Gamma_c)\left(1 - \frac{h_x(t)}{h_a(t)}\right) b_a(t) m. \tag{3.10}$$

Similar equations can be derived for host cells with vacuolar and cytosolic bacteria:

$$\dot{h}_v = \frac{\dot{H}_v}{H_{\text{tot}}} = \Gamma_v \frac{H_a - H_v}{H_{\text{tot}}} \tilde{b}_a = \Gamma_v (h_a - h_v) \frac{b_a m}{h_a} = \Gamma_v \left(1 - \frac{h_v(t)}{h_a(t)}\right) b_a(t) m. \tag{3.11}$$

$$\dot{h}_c = \frac{\dot{H}_c}{H_{\text{tot}}} = \Gamma_c \frac{H_a - H_c}{H_{\text{tot}}} \tilde{b}_a = \Gamma_c (h_a - h_c) \frac{b_a m}{h_a} = \Gamma_c \left(1 - \frac{h_c(t)}{h_a(t)}\right) b_a(t) m. \tag{3.12}$$

Note that $\dot{H}_v + \dot{H}_c \geq \dot{H}_x$ since vacuolar and cytosolic invasion can occur independently in the same host [36].

Besides invading, attached bacteria will also induce ruffle formation at rate Γ_r:

$$\dot{h}_r = \frac{\dot{H}_r}{H_{\text{tot}}} = \Gamma_r \frac{H_a - H_r}{H_{\text{tot}}} \tilde{b}_a = \Gamma_r (h_a - h_r) \frac{b_a m}{h_a} = \Gamma_r \left(1 - \frac{h_r(t)}{h_a(t)}\right) b_a(t) m. \tag{3.13}$$

A more complicated quantity is the number of bacteria which are attached to host cells B_a, which has three means of change. The first is by regular primary attachment with rate Γ_a, the second is ruffle recruitment with rate Γ_b, and the third is a loss of attached bacteria as they invade with limited invasion rate Γ_x^*. Fractional bacterial attachment is therefore

$$\begin{aligned} \dot{b}_a = \frac{\dot{B}_a}{B_{\text{tot}}} &= \Gamma_a \frac{H_{\text{tot}} \rho_B}{B_{\text{tot}}} + \Gamma_b \frac{R \rho_B}{B_{\text{tot}}} - \Gamma_x^* \frac{B_a}{B_{\text{tot}}} \\ &= \Gamma_a b(t) c + \Gamma_b \tilde{r}(t) h_r(t) b(t) c - \Gamma_x b_a(t) \left(1 - \frac{\tilde{b}_x(t)}{\tilde{b}_{x,\max}}\right). \end{aligned} \tag{3.14}$$

Bacteria will either internalize and remain vacuolar at rate Γ_v^* or escape early and become cytosolic at rate Γ_c^*:

$$\dot{b}_v = \frac{\dot{B}_v}{B_{\text{tot}}} = \Gamma_v^* \frac{B_a}{B_{\text{tot}}} = \Gamma_v b_a(t) \left(1 - \frac{\tilde{b}_x(t)}{\tilde{b}_{x,\max}}\right) \tag{3.15}$$

$$\dot{b}_c = \frac{\dot{B}_c}{B_{\text{tot}}} = \Gamma_c^* \frac{B_a}{B_{\text{tot}}} = \Gamma_c b_a(t) \left(1 - \frac{\tilde{b}_x(t)}{\tilde{b}_{x,\max}}\right), \tag{3.16}$$

and, as expected, we have $\dot{b}_v + \dot{b}_c$ equals the negative (loss) term in Eq 3.14.

And finally, the change in ruffles R per cell H_r is dependent on the limited formation rate Γ_r^*:

$$\dot{\tilde{r}} = \frac{\dot{R}}{H_r} = \Gamma_r^* \frac{B_a}{H_r} = \Gamma_r \left(1 - \frac{\tilde{r(t)}}{\tilde{r}_{\max}}\right) \frac{m b_a(t)}{h_r(t)}. \tag{3.17}$$

Taken altogether, we now have a set of coupled differential equations that can be numerically integrated in time (for the results presented here, the fourth-order Runge Kutta method was used [67]). The source code can be seen in Appendix A.2. The system is initialized at zero infectivity $h_a(0) = h_x(0) = 0$ and with all inoculating bacteria swimming $b(0) = 1$. A sufficiently small time step Δt was determined for accuracy, and used to integrate until some set time $t_{\max}$ (the inoculation time experimentally).

3.2 Parameterizing the Model

The goal of this section is to collect data from a variety of different publications to estimate interaction rates of our mean field infection model. Before getting to actual numerics however, we can state some expectations of these parameters that help to constrain the parameter space.

The 2012 study by Misselwitz et al. [27] examined target-site selection of *Salmonella* Typhimurium. A key finding was that bacteria preferentially docked onto ruffles, an effect the authors speculated to arise from the ruffle topology. We therefore expect the ruffle attachment rate to exceed primary attachment, or $\Gamma_b > \Gamma_a$.

Another simple inference comes from Knodler et al. [37] who quantified the proportion of cytosolic *S.* Typhimurium at early times, and found it quickly rose to $\sim 20\%$ and stayed there for at least 180 minutes post infection, as plotted in Fig 3.2I.

Firstly, this shows that early vacuolar escape is not a continuous process but rather a predetermined "decision" that will either happen soon after internalization, or not at all. For this reason, we model vacuolar occupation and escape as separate internalization paths, rather than a slow progress from enclosed to escaped. Secondly, this gives us an approximate ratio of internalization rates $\Gamma_v \approx 4\Gamma_c$.

Misselwitz et al. [68] showed in 2011 how attachment of *Salmonella* to HeLa varies with MOI and time for a number of different mutants. Their data for percentage of cells with attached wildtype *Salmonella* (h_a in our model) is shown in Fig 3.2A and B. In regards to the model, this quantity is h_a and is regulated by the primary attachment rate Γ_a. The 2012 study by the same group [27] gives us ruffling fraction (percentage of cells with at least one ruffle, or h_r) versus MOI, as reproduced in Fig 3.2C.

A 1998 paper by Huang et al. [66] on entry of *Salmonella* into HeLa cells found an apparent physical limitation on internalization at large MOI. To show this, they measured invasion efficiency (percentage of internalized bacteria/starting inoculum, $b_x = b_c + b_v$ in the model) and found it to decrease past 40 MOI. Along those same lines, the number of internalized bacteria per cell ($\tilde{b}_x$) was found to saturate at around 20 bacteria, a limit we call $\tilde{b}_{x,\max}$. Also included in the study was a time course analysis of internalization. At time points between 0 and 60 minutes post infection, invasion efficiency (b_x), infectivity (h_x) and the number of infection foci per cell were measured. These foci were clusters of internalized bacteria and, interestingly, the authors found an average of 2 ± 1 foci of infection per infected cell at various times post infection and at MOIs of 40 and 400. We hypothesize that these foci of infection are sites of ruffling that enhance entry of attached bacteria, and that typically 1-3 sites form on a single HeLa cell. To incorporate this effect into the model, we use the variable $\tilde{r}_{\max}$ which will place a saturation limit on ruffles per cell $\tilde{r}$ and curb the cooperative effect at large MOI and late times. It is worth noting that our definition of "ruffle" is now misleading. More accurately, our model includes "sites of ruffling", as a single site

may have many protrusions one could call a ruffle. The relative size of these ruffling sites may change in time to enhance local attachment, but that is a detail both not available in the literature and out of the scope of this model.

Now with an abundance of extracted data, the challenge was to find parameters that fit. As a first step, a script was built for real-time manual adjustment of all parameters, which allowed us to find ballpark estimates of the values. We took these estimates as initial guesses to a least squares fitting routine which was used on each set of data separately and sequentially. To speed up the fitting, certain constraints were used on the parameter space. Since it is possible that some of the data does not experience large saturation effects due to low bacterial load, the parameters $\tilde{b}_{x,\max}$ and $\tilde{r}_{\max}$ were allowed to vary between 10-30 and 1-10, respectively. The constraint on c was between 0 and 1, since it is a percentage confluence in our model, rather than a number density. For the rates Γ_a, Γ_b, Γ_r, Γ_v and Γ_c, a very large range of variation was allowed, typically a width of three factors of ten centered around our initial guess (Γ_a between 0.0001 and 0.01, for example).

The results of our sequential fitting can be seen in Table 3.3. Each row corresponds to a different set of data from the literature in the order (top to bottom) that they were fit. Bolded values in the table indicate parameters that were allowed to vary for the particular fitting (only Γ_a and c for Fig 3.2A and B, for example). The first sets of data fit were, naturally, relevant to the first of the stage of invasion – percentage of cells with attached bacteria (h_a) versus time and MOI, courtesy of Misselwitz et al. [68]. This gave an estimate of the primary attachment rate Γ_a. The next to be fit was ruffling fraction (h_r) versus MOI from Misselwitz et al. [27] to address ruffle formation rate Γ_r, as well as tune the attachment rates Γ_a and Γ_b relative to each other. From Huang et al. [66] came various measures of bacteria internalization (invasion efficiency b_x, bacteria per cell $\tilde{b}_x$ and infectivity h_x) which were fit to give estimates of Γ_c, Γ_v and the limiting factors $\tilde{b}_{x,\max}$ and $\tilde{r}_{\max}$. And finally, the relatively simple result from

Knodler et al. [37] of cytosolic bacteria proportion (b_c / b_x) was fit to set the expected ratio between Γ_c and Γ_v.

	Γ_a	Γ_b	Γ_r	Γ_v	Γ_c	$\tilde{b}_{x,\max}$	$\tilde{r}_{\max}$	c
Initial guess	0.001	0.01	0.01	0.01	0.0025	20	2	1
Fig 3.2A	**0.000909**	0.01	0.01	0.01	0.0025	20	2	**0.989**
Fig 3.2B	**0.00214**	0.01	0.01	0.01	0.0025	20	2	**0.244**
Fig 3.2C	**0.000843**	**0.0101**	**0.0228**	0.01	0.0025	20	2	**0.528**
Fig 3.2D	0.000843	0.0101	0.0228	**0.0488**	**0.00473**	**16.4**	**1.59**	**0.962**
Fig 3.2E	0.000843	0.0101	0.0228	**0.0552**	**0.00553**	**29.4**	**1.00**	**0.522**
Fig 3.2F	0.000843	0.0101	0.0228	**0.0156**	**0.00243**	**29.8**	**2.17**	**0.159**
Fig 3.2G	0.000843	0.0101	0.0228	**0.0179**	**0.00288**	**25.1**	**2.39**	**0.216**
Fig 3.2H	0.000843	0.0101	0.0228	**0.00738**	**0.000754**	**19.3**	**3.25**	**0.159**
Fig 3.2I	0.000843	0.0101	0.0228	**0.0117**	**0.00301**	19.3	3.25	**0.959**

Table 3.3: Chronological list of the least squares fitting to each set of literature data. Bolded values indicate that, for that step, those parameters were allowed to vary. For example, Fig 3.2A and B quantify percentage of host cells with attached bacteria, so the only relevant parameters are the primary attachment rate Γ_a and the experiment-specific confluency c. For experimental parameters used in the fitting (MOI and incubation time), refer to the companion Table 3.5.

The blue lines in Fig 3.2 show the results of this step by step fitting and, in most cases, good fits by eye could be achieved. This is perhaps unsurprising, as the number of model parameters outnumbers the number of data points in most cases, which allows a great amount of numerical flexibility. Ideally, however, we would like to find a set of parameters, if they exist, that describe all *Salmonella*-HeLa infection assays. Before fitting all the data simultaneously, a fit was performed to each of the four separate studies. This modular fitting was done because, at the very least, we expect our model should be able to describe all data from a single experiment done under the same conditions (confluency, equipment, experimentalist, etc.). The results of this fitting are listed in Table 3.4 and plotted as green lines in Fig 3.2. Taking the parameters from the modular fitting as an initial guess, and keeping experiment-specific confluency c constant, a least squares fit including all of the data was performed. This final stage of fitting can be seen in the last row of Table 3.4 and as the red lines in Fig 3.2.

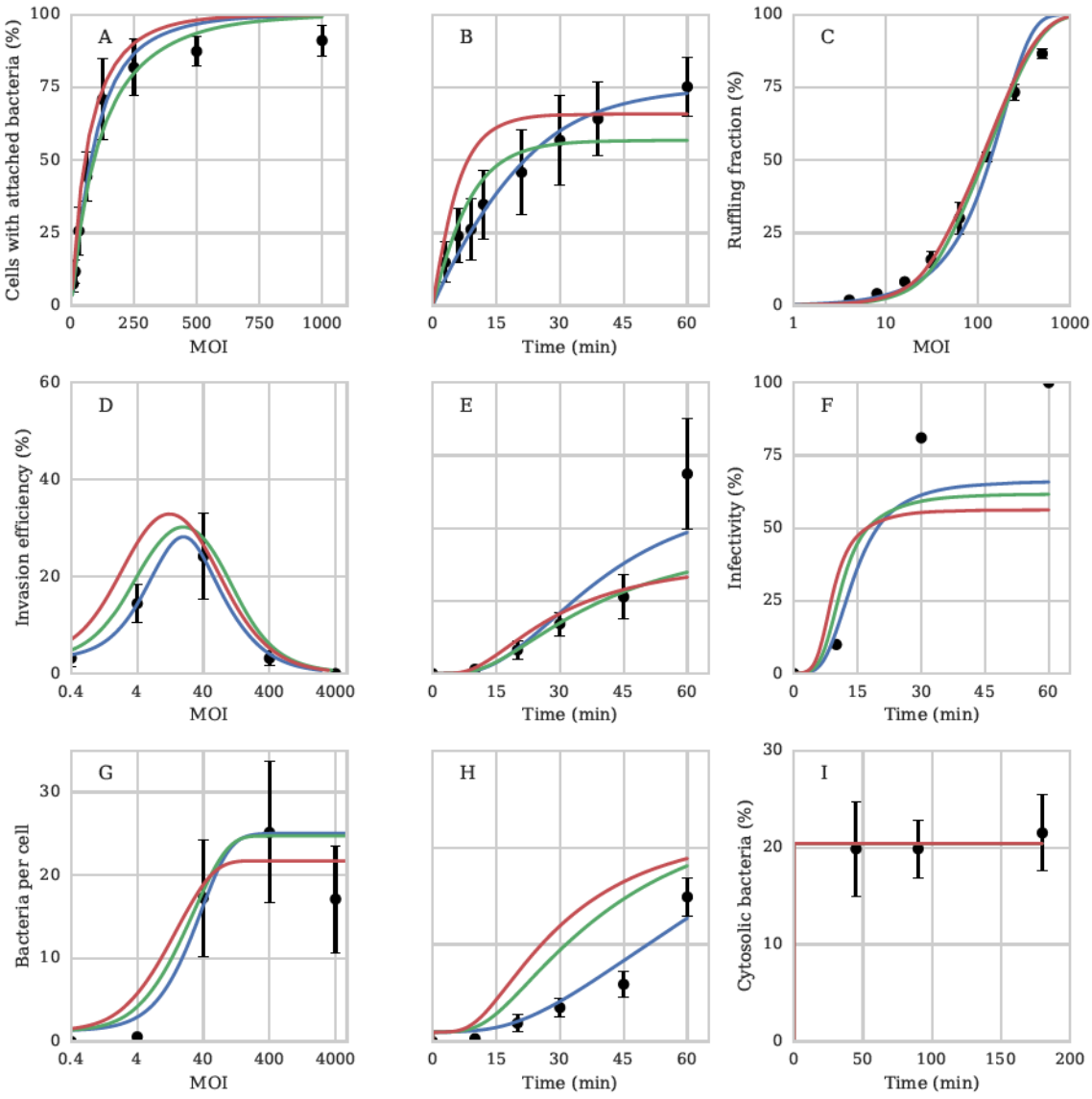

Figure 3.2: Plots of various measures of infection related to *Salmonella enterica* invasion into HeLa cells, including cells with attached bacteria (h_a), ruffling fraction (h_r), invasion efficiency (b_x), infectivity (h_x), bacteria per cell ($\tilde{b}_x$) and cytosolic bacteria (b_c/b_x). Black circles indicate data from the cited sources, with error bars if they were reported. (A,B) Misselwitz et al. [68]. (C) Misselwitz et al. [27]. (D-H) Huang et al. [66]. (I) Knodler et al. [37]. The lines show best fit estimates of the mean field model. Blue indicates the best fit to the individual data sets, green indicates the best modular fit to the individual studies, and red indicates the best simultaneous fit to all data.

Source	Fig 3.2	Γ_a	Γ_b	Γ_r	Γ_v	Γ_c	$\tilde{b}_{x,\max}$	$\tilde{r}_{\max}$	c
Misselwitz et al. [68]	A, B	**0.000891**	0.117	0.0757	0.0225	0.00346	20.8	1.95	**0.999**
Misselwitz et al. [27]	C	**0.00137**	**0.123**	**0.108**	0.0225	0.00346	20.8	1.95	**0.528**
Huang et al. [66]	D-H	0.00137	0.123	0.108	**0.0128**	**0.00195**	**24.7**	**1.55**	**0.999**
Knodler et al. [37]	I	0.00137	0.123	0.108	**0.0117**	**0.00301**	24.7	1.55	**0.139**
All data	-	**0.00149**	**0.141**	**0.0809**	**0.0133**	**0.00342**	**21.7**	**1.85**	-

Table 3.4: Chronological list of the least squares modular fitting to sets of data (from the same source) and all the data. Bolded values indicate that, for that step, those parameters were allowed to vary. When fitting all the data simultaneously, the confluency were kept constant per-experiment at modular fit estimated values. For experimental parameters used in the fitting (MOI and incubation time), refer to the companion Table 3.5.

Expectedly, the quality of the fit diminished when attempting to incorporate more sets of data from the literature. Deviations from the model can potentially be explained by variations between experiments or by incomplete/missing details, such as confluency or inexact MOI (i.e. when given a range of MOI). For illustrative purposes, some of these experimental details are listed in Table 3.5. Variables such as incubation medium and culture plate are not taken into account in our model, simply because there isn't data comparing the effects. It is entirely possible that, taken altogether, these differences in experimental procedure and equipment alter the average infection dynamics of the host-pathogen population, complicating the fitting to multiple sets of data.

Alternatively, there may of course be oversights or missing pieces in the model. It would appear the worst deviations from the model occur at late times (Fig 3.2B, E, F, H) and high MOI (Fig 3.2A, C). One theory is that there is an additional saturation effect occurring at high bacterial loads. In addition to $\tilde{b}_{x,\max}$, there may be some limitation on the number of attached bacteria per cell (call it $\tilde{b}_{a,\max}$) which restricts the attachment rates Γ_a, Γ_b at late times and high MOI. The poor fitting may also be a consequence of stochastic effects at early times and low MOI, when the sample size of infected host cells and bacteria is smallest. A more careful fitting would involve giving less weight to these data points, and may resolve disparities in the late time

and high MOI regime. Inconsistencies aside, we can say that our simple mean field model with rates can capture the approximate behaviour and essential features (such as internalization saturation) of the highly stochastic process of bacterial infection.

Fig 3.2	Source	Independent variable	Dependent variable	MOI	Incubation time (minutes)	Confluency (cells/well)	HeLa strain	Incubation temperature	Culture plate	Incubation medium	Growth medium
A	Misselwitz 2011	MOI	h_a	-	10	$6{\cdot}10^3$	Kyoto	37	96-well Greiner	DMEM, 10% FCS	DMEM
B	Misselwitz 2012	Time	h_a	62.5	60	$6{\cdot}10^3$	?	37	96-well Greiner	DMEM, 10% FCS	LB Broth
C	Misselwitz 2012	MOI	h_r	-	9	$3{\cdot}10^5$	?	37	96-well Greiner	DMEM, 10% FCS	LB Broth
D	Huang 1998	MOI	b_x	-	60	$6{\cdot}10^5$	INT407	37	24-well Sarstedt	MEM, 10% FCS	LB Miller
E	Huang 1998	Time	b_x	40	60	$6{\cdot}10^5$	INT407	37	24-well Sarstedt	MEM, 10% FCS	LB Miller
F	Huang 1998	Time	h_x	40	60	$6{\cdot}10^5$	INT407	37	24-well Sarstedt	MEM, 10% FCS	LB Miller
G	Huang 1998	MOI	$\tilde{b}_x$	-	60	$6{\cdot}10^5$	INT407	37	24-well Sarstedt	MEM, 10% FCS	LB Miller
H	Huang 1998	Time	$\tilde{b}_x$	40	60	$6{\cdot}10^5$	INT407	37	24-well Sarstedt	MEM, 10% FCS	LB Miller
I	Knodler 2014	Time	b_c/b_x	50-100	10	$5{\cdot}10^4$	ATCC	37	24-well Corning Costar	EMEM, 10% FCS	LB Gibco

Table 3.5: A list of the literature data used in this study, along with some provided experimental details. Incubation time corresponds to the $t_{\max}$ model parameter. The given confluency values are listed as HeLa seeded per well which, unfortunately, cannot be directly translated to the percentage confluency c in the model.

3.3 Applications

We have shown that a mean field model of *Salmonella* invading HeLa cells can be fit to data from different sources to find rates of the various stages of invasion. If simple time-independent rates like these can roughly describe population-level dynamics of host-pathogen invasion assays, then we believe our model could prove a useful diagnostic tool in designing experiments. For instance, if an experimentalist wishes to study some aspect of *Salmonella*-induced ruffling on HeLa cells, then by inputting confluency and incubation time, our model will output a predicted ruffling fraction of host cells for a range of MOI. From this estimation, one could determine a reasonable MOI to achieve a high sample size of ruffled cells.

Another area where we believe this model has merit is in quantifying differences between strains and species of bacteria. A common practice when studying the mechanisms of invasion is to knockout genes encoding for various pathogen virulence factors and monitoring the effects. In the absence of, say, *Salmonella* genes encoding for ruffle-related effectors, one could measure the response on the ruffle formation rate Γ_r and provide a more precise quantitative comparison.

Beyond just *Salmonella* invasion, this model is easily extendable to other host-pathogen systems. Perhaps most similar to *Salmonella* is the bacterium *Shigella flexneri*, which invades epithelial cells via T3SS-mediated trigger mechanism (with membrane ruffling as a byproduct) and then quickly ruptures its vacuole to escape into the cytosol [18, 38]. Excluding a pathogens vacuolar lifestyle can be accomplished by simply "turning off" the vacuolar invasion module (setting $\Gamma_v = 0$ in the model). The other rates Γ_a, Γ_b, Γ_r, Γ_c and saturation limits $\tilde{b}_{x,\max}$, $\tilde{r}_{\max}$ will be expectedly different from our *Salmonella* parameterization simply due to the different virulence factors and effectors used by the two bacteria. Once parameterized to

different systems, the model can provide a useful time and MOI-independent benchmark for comparing pathogens and making more specific conclusions. For example, one could find that *Salmonella* has a ruffle formation rate x times greater than that of *Shigella*, and a ruffle attachment rate y times greater, leading to the conclusion that *Salmonella* is z times more effective at cooperatively invading, independent of experimental conditions.

Moreover, we believe this first attempt at a mean field infection model for *Salmonella*-HeLa has the components to be a generic model because it incorporates the various lifestyles of pathogens: extracellular, intracellular, vacuolar and cytosolic. As with *Shigella*, the modules can be turned off and on to accommodate different mechanisms and pathways of invasion. Of course, and as mentioned previously, more detail can and should be added to more accurately model host-pathogen interactions, such as an explicit rate for xenophagy. Additionally, the model could be extended in time to include pathogen replication, host cell death and transmission to other cells, but as a first step and proof of concept, our model can potentially accommodate any host-pathogen system with data available for parameterization.

Finally, as an illustration of the model's predictive ability, the parameterization from fitting all of the literature data (Table 3.4) was applied to the infectivity data from our hyper-replication analysis (Fig 2.5). MOI and incubation time were given experimental parameters, and percentage confluency was used as the only fitting parameter. Fig 3.3 (top) shows the result of said fitting to infectivity h_x for each experiment. As with the literature data, these fits are far from perfect, especially at higher MOI, but can roughly approximate the rise in infectivity. Fig 3.3 (bottom) then shows a model-generated prediction of the number of bacteria per infected cell $\tilde{b}_x$, or N as we called it in Chapter 2 before introducing the model. From this, we can say with a high degree of confidence that, by infecting with as high as 800 MOI, we have explored a wide range of bacterial load and likely reached the physical saturation

limit of $\tilde{b}_{x,\max} \approx 20$. This further invalidates the independent-escape hypothesis as, with such high bacterial loads, we would have expected a considerable increase in HR fraction with MOI.

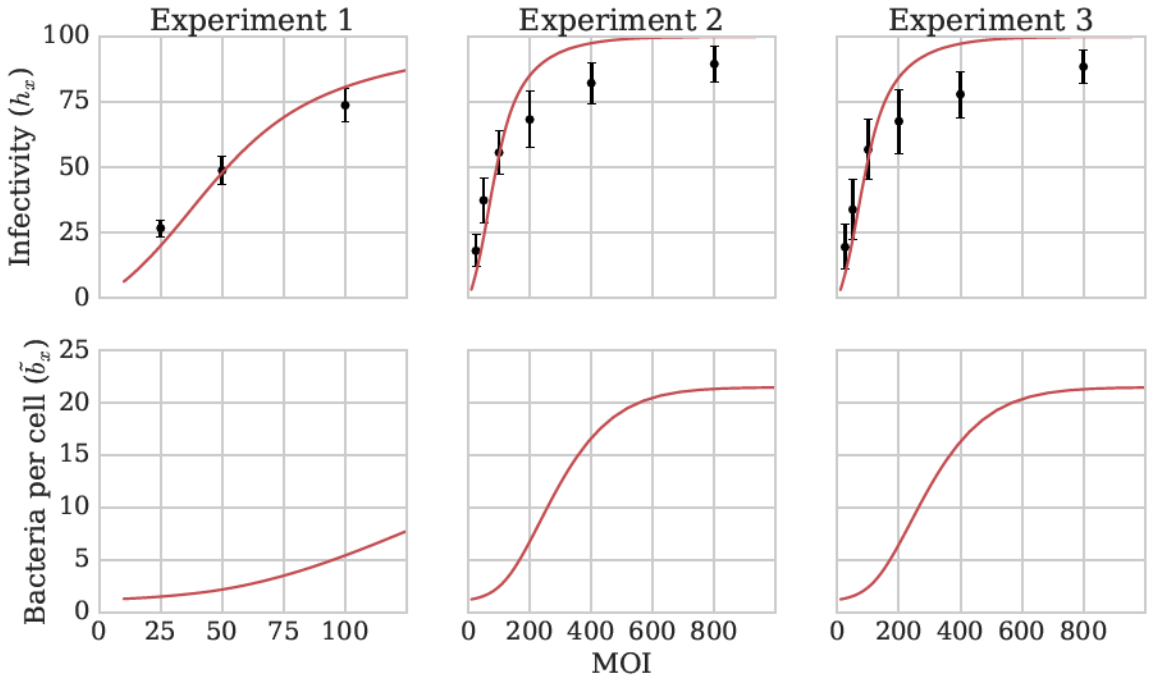

Figure 3.3: (Top) Infectivity versus MOI from the *Salmonella*-HeLa invasion experiments presented in Chapter 2. Red lines indicate fits from the literature-parameterized mean field model. The sole fitting parameter, confluency, was $c = 0.140, 0.0784, 0.0764$ for experiments 1, 2 and 3, respectively. (Bottom) Model predictions for bacteria per infected cell versus MOI.

Chapter 4

Conclusions and Future Work

In this thesis, we have studied the pathogenicity of *Salmonella enterica* Typhimurium in two distinct ways.

The first was a primarily experimental approach to investigating the hyper-replication of *Salmonella* in epithelial cell cytosol. Timelapse fluorescence microscopy images taken by our collaborators were carefully segmented and tracked to extract single cell growth curves. To better classify HR, we argued for an alternative definition of characteristic HR growth that involved modelling bimodal growth distributions as mixed Gaussians. We found that, for the large range of MOI used, this method could separate ambiguous growth into fast and slow categories, and give estimates of HR fraction. To our surprise, HR fraction was consistently around 8%, and is thus independent of initial bacterial load. This implies that vacuolar escape followed by fast cytosolic growth is not an independent choice for each bacterium. The simplest explanation is that hyper-replication is not a phenotype of *Salmonella*, but rather a susceptibility in a small subpopulation of epithelial cells. In other words, an infected cell with 1 bacterium has the same chance of harbouring HR bacteria as an infected cell with 10. Another hypothesis is that *Salmonella* are using bacterial communication to coordinate replication. Perhaps the first internalized bacterium attempts to escape (and is successful approximately 8% of the time), and the others receive a signal (via quorum sensing or some other mechanism) to stay vacuolar or enter a persister state, safe from host-directed xenophagy. This is akin to other pathogens,

like *E. Coli*, which hedge their bets through multiple intracellular niches for a better chance of population survival. Future work should involve more experiments of *Salmonella* infection for a range of MOI, but with markers for vacuolar or cytosolic localization.

The second part of this thesis was much more theoretical. A mean field model of invasion was developed to describe laboratory invasion assays of *Salmonella* Typhimurium infecting HeLa cells. This model involves rates describing stages of infection (attachment, ruffle formation, internalization) and a system of coupled differential equations to evolve the system in time. We speculated that there exists rates that could approximately describe all *Salmonella*-HeLa assays, subject to tunable experimental parameters MOI, confluency and incubation time. A collection of data was extracted from various publications to parameterize and test the model. By first hand-fitting, and then fitting data separately to get initial guesses for parameters, nine sets of data were fit simultaneously. Although not as good as the individual fits, we found the model-generated curves to reasonably approximate the data, and capture the essential features (such as the physical limitation on entry). Needless to say, there is still much that could be done to improve on the model. The first step is to include more data in the parameterization and fitting, as there is a wealth of literature available on *Salmonella* infecting HeLa. Another important addition will be a thorough error analysis to arrive at confidence intervals of model parameters. Due to the nature of the model and data (often, there were more fitting parameters than actual data points), the least-squares fitting routine could not estimate error. Ideally, the model should also include an explicit autophagy interaction rate, instead of being implicit in the other rates, but this parameterization may require data not yet available in the literature. Once better established, we believe this mean field model could be used as a predictive tool for choosing experimental parameters, and for gaining additional insight into the host-pathogen dynamics. For our HR analysis,

for example, we infected at a wide range of MOI to presumably obtain a similarly wide range of bacterial load. By fitting the mean field model to infectivity versus MOI, we were able to produce model predictions for internalized bacteria per cell. This confirmed our presumption that bacterial load varied strongly with MOI.

Appendix A

A.1 CellProfiler Pipeline

The open source software CellProfiler version 2.1.1 was used to segment images of *Salmonella* infecting HeLa cells, taken by Jennifer Fredlund. The included documentation provides additional details on each of the modules that we used, but also see [60] and [69] for overviews.

A screenshot of the CellProfiler pipeline (the set of modules to carry out the segmentation) can be seen in Fig A.1. On the left are the input modules for preprocessing and organizing the images. The Images module allows specification of input directory, and filtering of data based on filename and directory (useful for segmenting one well at a time, for example). The Metadata module extracts information about the images. As this was a timelapse experiment, each TIF file was a stack of frames in chronological order. CellProfiler automatically detects and splits the stacks (this is part of the image metadata, or embedded information) into individual frames. In our case, the images were generally renamed and placed according to the following format: `conf<Confluency>/moi<MOI>/<Well>/f<Field>_c<Channel>.tif`. This allowed simple classification and splitting into channels 1 (nuclei) and 2 (bacteria) by the NamesAndTypes module, and then further grouping by well and field through the Groups module.

Once preprocessing and grouping is finished, groups of images (the timelapses) are processed sequentially through all of the analysis modules, as pictured on the right of Fig A.1. Table A.1 lists the most important analysis modules and their settings used to segment images of the three experiments. Any modules or module options

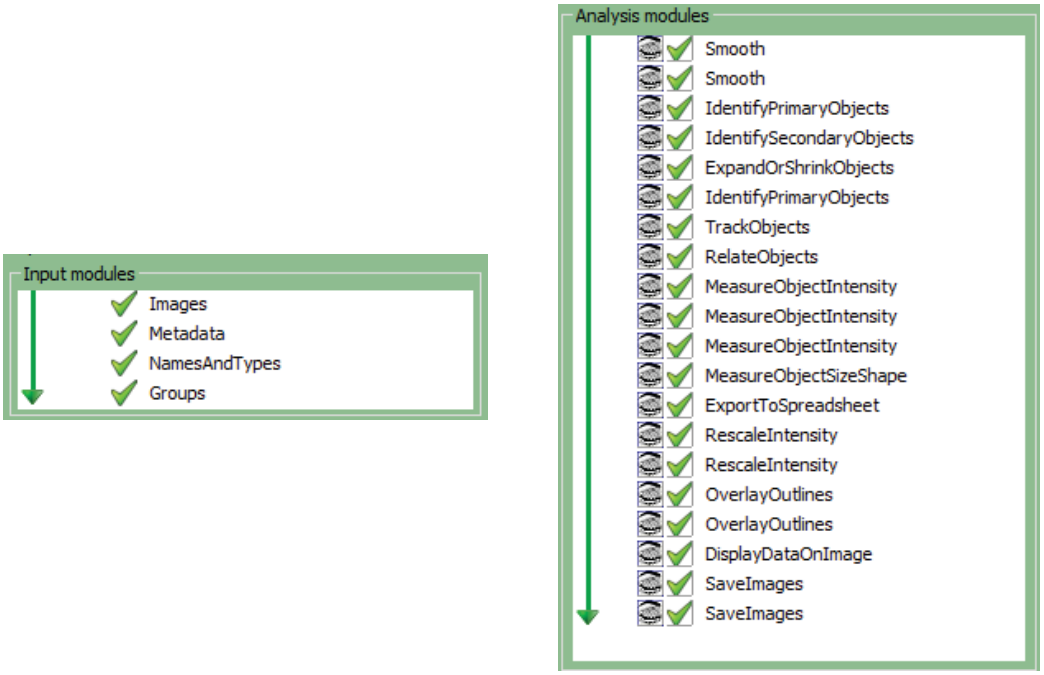

Figure A.1: Screenshot of the CellProfiler pipeline and modules used in the segmentation of the three *Salmonella*-HeLa timelapse fluorescence microscopy experiments. On the left are the input/preprocessing modules, and on the right are the analysis modules which are run sequentially on each set of images.

not listed or discussed were simply left as default or were unimportant to the actual segmentation.

The first two steps employ the Smooth module to clean up the nuclear and bacterial images. The standard method here is the simple Gaussian filter, which produced smooth images and eliminated noise for an artifact diameter set to 3 pixels. This was less important for segmentation of the nuclear images, but for single bacteria, which were seen as small as 5 pixels, small artifacts or noise can easily cause missegmentation.

The first identification module is IdentifyPrimaryObjects to segment HeLa cell nuclei. The object diameter setting was adjusted to the resolution of the data (larger diameter for the 40x images of experiments 2 and 3, compared to the 10x images of experiment 1). The upper bound on the diameter was chosen larger than actually expected in order to capture and diagnose anomalous segmentation (such as a cluster

of nuclei identified as one). The thresholding strategy for each experiment was the Otsu method, which separates pixels into two or three classes. It is the recommended initial approach by the CellProfiler documentation, and seemed to outperform other methods such as Background (which is appropriate for very low confluency images) and the Mixture of Gaussian method (which is best when the percentage confluency does not vary much from image to image). The three-class variant of the Otsu method separates the image into foreground, mid-level and background, and worked well for experiments 2 and 3 which has strong foreground signal (nuclei), relatively weak mid-level signal (cytoplasm) and strong background (extracellular). For experiment 1, however, the lower resolution, higher confluency and seemingly weaker signal made mid-level cytoplasmic detection difficult. Consequently, the two-class Otsu method was instead used to separate the image into foreground (nuclei) and background (extracellular). For highly confluent clusters of cells, it is often the case that nuclei are touching and therefore incorrectly segmented as single objects. To counteract this, the IdentifyPrimaryObjects module has methods of distinguishing clumped objects and drawing dividing lines. The shape method was found to work best with nuclei declumping, as they generally had a characteristic rounded shape. Dividing lines between clumps were also drawn by shape, which essentially meant splitting the objects by the indentations between them.

Module	Setting	Experiment 1	Experiment 2	Experiment 3
Smooth (nuclei and cells)	Method	Gaussian filter	Gaussian filter	Gaussian filter
	Artifact diameter	3	3	3
IdentifyPrimaryObjects (nuclei)	Typical diameter	10-80	30-100	30-100
	Threshold strategy	Otsu (two-class)	Otsu (three-class)	Otsu (three-class)
	Distinguish clumped objects by	Shape	Shape	Shape
	Draw dividing lines by	Shape	Shape	Shape
IdentifySecondaryObjects (cells)	Method	-	Propagation	Propagation
	Threshold strategy	-	Otsu (three-class)	Otsu (three-class)
ExpandOrShrinkObjects (cells)	Operation	Expand objects until touching	-	-
IdentifyPrimaryObjects (bacteria)	Typical diameter	5-500	5-500	5-500
	Threshold strategy	Manual	Manual	Manual
	Manual threshold	0.00007	0.000085	0.000085
TrackObjects (nuclei)	Tracking method	LAP	LAP	LAP
	Motion model	Both	Both	Both
	Number of standard deviations for search radius	3.0	3.0	3.0
	Search radius limit	30-60	80-100	80-100
	Gap cost	190	180	180
	Split cost	200	180	180
	Merge cost	220	200	200
	Maximum gap displacement	60	100	100
	Maximum split score	150	200	200
	Maximum merge score	150	200	200
	Maximum gap	5	5	5
RelateObjects	Child objects	bacteria	bacteria	bacteria
	Parent objects	cells	cells	cells

Table A.1: List of CellProfiler segmentation pipeline modules and settings for the different *Salmonella*-Hela fluorescence microscopy experiments.

IdentifySecondaryObjects was used to identify HeLa cell edges for experiments 2 and 3. This module identifies objects using other objects identified upstream (nuclei from the IdentifyPrimaryObjects module) as a starting point. The most sophisticated and recommended approach is the Propagation method, which expands outward from the nuclei and draws cell boundaries based on distance to other objects and intensity gradients. It is thought to be an improvement over the traditional Watershed method and, indeed, was found to produce better cellular outlines than the other available methods. This step was complicated, however, by an apparent decrease in secondary cytoplasmic staining over time. At late times, when the signal was weakest, the default automatic thresholding had trouble separating the cell from background. The three-class Otsu method was again found to be the winner in automatically determining the threshold and identifying the decaying mid-level signal. This was an unfortunate necessity as it greatly increased the computational time required, but a necessity nonetheless.

A simpler approach was taken for identifying cell edges for experiment 1. As previously mentioned, the high confluency and low resolution made mid-level cytoplasmic identification difficult. Instead, nuclei objects were simply expanded evenly and simultaneously until touching one another (through the ExpandOrShrinkObjects module), assigning every pixel of the image to an object (this is commonly called Voronoi expansion). The borders of contact between nuclei are then simply the estimated cellular outlines. This seemed a reasonable compromise as very little of these highly confluent monolayer images was background, and most nuclei were close to touching one another so the expansion was often minimal.

With the nuclear channel fully segmented, the next step is to identify bacteria through the IdentifyPrimaryObjects module. The lower bound for object diameter was set to 5 pixels (slightly above the artifact diameter setting in the Smooth module) to capture small single bacteria, and the upper bound was set arbitrarily large (500

pixels) to capture any sized clump of bacteria. It was found that, for all three experiments, the fluorescent DsRed signal was consistently above a manually set threshold (around 0.00008) regardless of time. This was a welcome finding, as the automatic thresholding methods often failed to to find the appropriate value to separate bacteria from background. Generally, these failures seemed to occur at late times and high MOI where large bright clumps of *Salmonella* resulted in a wide distribution of pixel intensities that likely threw off the threshold calculation.

Perhaps the most important, and certainly the most carefully tuned module of this pipeline was the TrackObjects module. Early on, especially with the highly confluent experiment 1, it was found that the simple tracking methods (Overlap and Distance) were mistracking nuclei throughout time. When nuclei were close together or nearly touching in one frame, despite being correctly declumped by the IdentifyPrimaryObjects module, CellProfiler would erroneously assign them the same track label. Thankfully, a more sophisticated method called the LAP (linear assignment problem) [62] is available. This approach involves two passes over the images. The first is a basic frame-to-frame object linking (based on object overlap and size) to form track segments. The second pass links the resulting partial trajectories into complete trajectories by solving a global combinatorial optimization problem whose solution identifies the most likely set of object trajectories. This step, which takes place at the end of the CellProfiler analysis run, tends to resolve issues of object disappearance (gaps in the trajectory due to objects going off frame), splitting events (when touching nuclei are split into distinct tracks) and merging events (less common, but sometimes strangely shaped nuclei, due to segmentation error or because the cell is dying, are split into pieces and must be merged to the same track). There are many parameters involved with the LAP two-pass method, but the documentation is unfortunately lacking, often giving advice such as "set the split cost lower if objects are being split that should not be split." The more intuitive parameters include motion

model ("both" was used as cells sometimes moved predictably with a velocity, and sometimes seemingly randomly), and search radius limit (a smaller radius was used for experiment 1 due to the high confluency and constricted cellular movement). For the other parameters, trial and error was necessary to achieve high quality (by eye) tracking of HeLa cell nuclei, and the parameters used in the end are listed in Table A.1.

To give an example of the issues resolved by this careful tracking of cellular trajectories, Fig A.2 shows a segmented infected cell (indicated by the yellow 159 label) with a nuclei touching a neighbouring nuclei (indicated by the magenta 161 label) at early times. First attempts at tracking cells like these resulted in assignment of touching cells to the same trajectory. The second pass optimization step of the LAP algorithm (once appropriately tuned) identified the splitting event here, and gave each their own track label.

After tracking, the RelateObjects module combines the two channels by associating cells to overlapping bacteria in each frame. The next 4 modules take measurements of object brightness (MeasureObjectIntensity) and size (MeasureObjectSizeShape). The most important measurements to our HR analysis were the intensity and size of bacterial objects, as they act as our proxies for bacterial load per cell. This data, and much more, is output by the ExportToSpreadsheet module which produces a CSV file to be later parsed for analysis by our Python scripts. The last 7 modules (RescaleIntensity twice, OverlayOutlines twice, DisplayDataOnImages and SaveImages twice) are simply extra steps to visually examine the segmentation and produce images like in Fig A.2

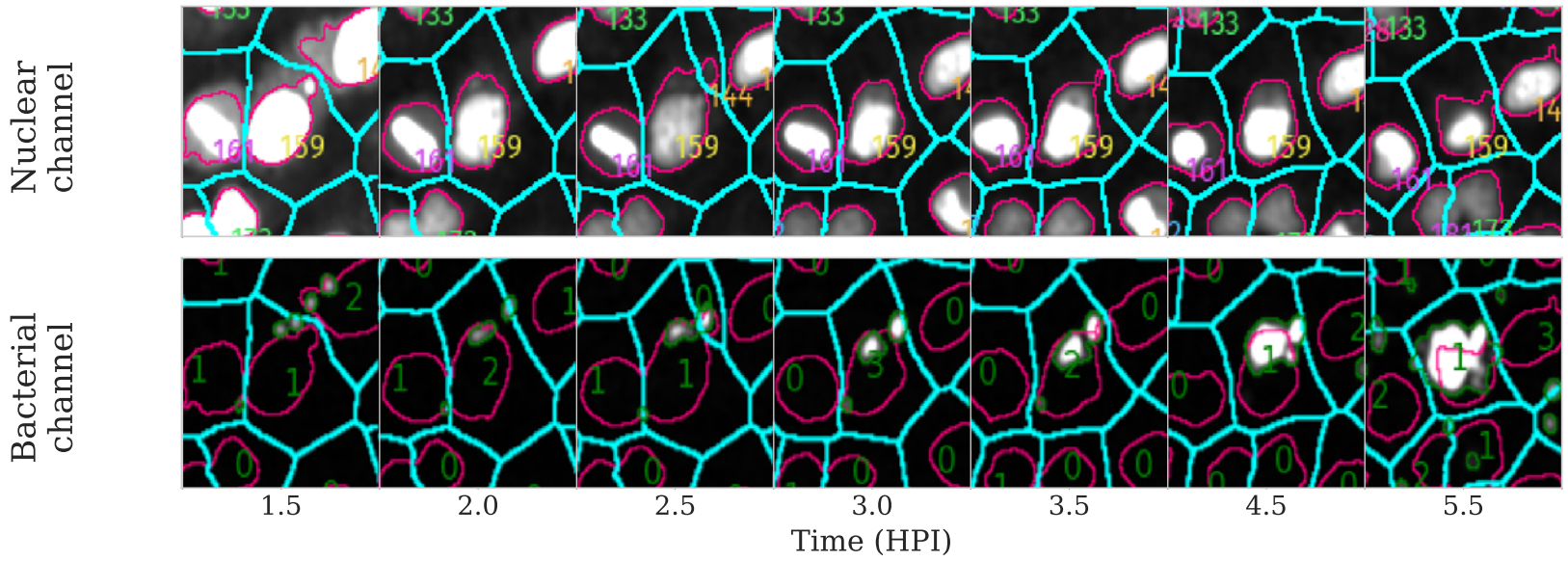

Figure A.2: Segmentation example of a difficult-to-track infected HeLa cell. This particular cell came from a well of experiment 1 which was infected at 25 MOI. Nuclei are outlined in red, cell edges in blue, and bacteria clumps in green. The numbers in nuclear channel are the nuclei-specific track labels and the numbers in the bacterial channel are the bacterial object counts per cell. The LAP (two-pass) tracking algorithm was able to split these nuclei (labelled above as 159 and 161) into two separate tracks. Images courtesy of Jennifer Fredlund.

A.2 Mean Field Model Code

The following Python script used the differential equations derived in Section 3.1.4 to integrate the dynamical variables listed in Table 3.1.

```
import numpy as np

def dv(v, m, ga, gb, gr, gv, gc, bx_max, r_max, c):
    """
    This function, given the current state of the system, the model
    parameters and experimental parameters, will calculate the
    incremental change in the time-dependent variables.

    Model parameters:
    ga = primary attachment rate
    gb = ruffle attachment rate
    gr = ruffle formation rate
    gv = vacuolar invasion rate
    gc = cytosolic invasion rate
    bx_max = saturation limit on invaded bacteria per cell
    r_max = saturation limit on ruffles per cell

    Experimental parameters:
    m = multiplicity of infection (ratio of bacteria to host)
    c = confluency (fraction of culture dish area occupied by host cells)

    Time-dependent variables:
    ha = host cells with attached bacteria
    hx = host cells with invaded bacteria
    hv = host cells with vacuolar bacteria
    hc = host cells with cytosolic bacteria
    hr = host cells with ruffles
    ba = bacteria attached to a host cell
    bv = bacteria that are vacuolar
    bc = bacteria that are cytosolic
    bx = bacteria that have invaded a host cell
    bx_max = maximum bacteria that can invaded a host cell
    r = ruffles per host
    """
    ha, hx, hv, hc, hr, ba, bv, bc, r = v

    h = 1 - ha
    bx = bv + bc
```

```
    b = 1 - (ba + bx)
    if hx > 0.0: bx_hx = m*bx / hx
    else: bx_hx = 0.0

    dha = h*ga*b*m*c
    if ha > 0:
        dhx = (1-hx/ha)*(gc+gv)*ba*m
        dhv = (1-hv/ha)*gv*ba*m
        dhc = (1-hc/ha)*gc*ba*m
        dhr = (1-hr/ha)*gr*ba*m
    else:
        dhx = 0
        dhv = 0
        dhc = 0
        dhr = 0

    scale_fact = 1 - bx_hx/bx_max
    dbc = ba*gc*scale_fact
    dbv = ba*gv*scale_fact
    dba = ga*b*c + r*hr*gb*b*c - (dbc+dbv)
    if hr > 0: dr = gr*(1-r/r_max)*ba*m/hr
    else: dr = 0

    dv_array = np.array([dha, dhx, dhv, dhc, dhr, dba, dbv, dbc, dr])
    return dv_array

def rk4(v, dt, m, ga, gb, gr, gv, gc, bx_max, r_max, c):
    """
    Fourth order Runge-Kutta integration.
    """
    k1 = dt * dv(v, m, ga, gb, gr, gv, gc, bx_max, r_max, c)
    k2 = dt * dv(v+0.5*k1, m, ga, gb, gr, gv, gc, bx_max, r_max, c)
    k3 = dt * dv(v+0.5*k2, m, ga, gb, gr, gv, gc, bx_max, r_max, c)
    k4 = dt * dv(v+k3, m, ga, gb, gr, gv, gc, bx_max, r_max, c)
    return (k1 + 2*k2  + 2*k3 + k4) / 6

def integrate(m, ga, gb, gr, gv, gc, bx_max, r_max, c, dt, tmax):
    """
    This function will initialize a host-pathogen system and
    integrate in time by step size dt until tmax is reached.

    If the given multiplicity of infection (m) is a single value,
    then a single integration will be performed, and the time
    evolution of the system will be returned.
```

```
    If a list of MOI is given, then an integration for each MOI
    will be performed, and only the values at the end of the
    integration (at tmax) will be returned.
    """

    # Initial conditions
    ha = 0.0 # fraction of cells with bacteria attached
    hx = 0.0 # fraction of cells infected
    hv = 0.0 # fraction of cells infected with vacuolar bacteria
    hc = 0.0 # fraction of cells infected with cytosolic bacteria
    hr = 0.0 # fraction of cells with ruffles
    ba = 0.0 # fraction of bacteria attached
    bv = 0.0 # fraction of vacuolar bacteria
    bc = 0.0 # fraction of cytosolic bacteria
    r = 1.0 # ruffles per ruffled cell

    t_points = np.arange(0, tmax+dt, dt)
    # If the given MOI is a list, integrate for each MOI
    if type(m) == list or type(m) == np.ndarray:
        v_points = np.empty(shape=(len(m), 9))
        for i,moi in enumerate(m):
            v = np.array([ha, hx, hv, hc, hr, ba, bv, bc, r])
            for t in t_points:
                v += rk4(v, dt, moi, ga, gb, gr, gv, gc, \
                         bx_max, r_max, c)
            v_points[i,:] = v
        return v_points
    # Else the given MOI is a single value, so integrate once
    else:
        v = np.array([ha, hx, hv, hc, hr, ba, bv, bc, r])
        v_points = np.empty(shape=(len(t_points), len(v)))
        for i,t in enumerate(t_points):
            v_points[i,:] = v
            v += rk4(v, dt, m, ga, gb, gr, gv, gc, bx_max, r_max, c)
        return v_points, t_points
```

To give an example of how this code is used in practice, the following is a simple script to calculate internalized bacteria per infected cell $\tilde{b}_x$ versus time at a specific MOI.

```
import numpy as np
from mean_field import integrate

# First, set the experimental parameters.
```

```
m = 40 # multiplicity of infection
c = 0.5 # percentage confluency
tmax = 60 # incubation time in minutes
dt = 0.01 # time step in minutes


# Second, set the model parameters (rates and limits).
ga = 1e-3
gb = 1e-1
gr = 1e-2
gv = 1e-2
gc = 2.5e-3
bx_max = 20
r_max = 3

# Compute the integration.
v_points, t_points = integrate(m, ga, gb, gr, gv, gc, \
                               bx_max, r_max, c, dt, tmax)

# The 2d v_points array now contains the time evolution
#  of the dynamical variables.
# Each column corresponds to a different variable and
#  can be unpackaged like so:
ha = v_points[:,0]
hx = v_points[:,1]
hv = v_points[:,2]
hc = v_points[:,3]
hr = v_points[:,4]
ba = v_points[:,5]
bv = v_points[:,6]
bc = v_points[:,7]
r = v_points[:,8]

# Some quantities of interest are not given directly by
#  the integration, and must be calculated afterwards.
# Invaded bacteria per infected cell, for instance:
bx = bc + bc
# To avoid dividing by zero, initialize an array
#  of zeros, then do appropriate slicing to divide.
bx_hx = np.zeros(len(bx))
bx_hx[hx>0] = m*bx[hx>0] / hx[hx>0]

# We now have bacteria per cell (bx_hx) versus time
#  (t_points) and can compare/fit to real data.
```

Bibliography

[1] World health statistics 2016: monitoring health for the SDGs, sustainable development goals. Technical report, World Health Organization, 2016.

[2] C. Janeway, P. Travers, M. Walport, and M. Shlomchik. *Immunobiology: The Immune System in Health and Disease*. Garland Science, New York, 5 edition, 2001.

[3] S. Akira, S. Uematsu, and O. Takeuchi. Pathogen recognition and innate immunity. *Cell*, 124:783–801, 2006.

[4] N. A. Campbell, J. B. Reece, L. A. Urry, M. L. Cain, S. A. Wasserman, P. V. Minorsky, and R. B. Jackson. *Biology (8th Edition)*. Pearson, San Francisco, 8 edition, 2008.

[5] A. Iwasaki and R. Medzhitov. Regulation of adaptive immunity by the innate immune system. *Science*, 327(January):291–296, 2010.

[6] A. Garcia-Sastre and P. J. Sansonetti. Host–pathogen interactions. *Current Opinion in Immunology*, 22:425–427, 2010.

[7] J. H. Powers. Antimicrobial drug development – the past, the present, and the future. *Clin Microbiol Infect*, 10:23–31, 2004.

[8] S. Reardon. Antibiotic resistance sweeping developing world. *Nature*, 509:141–2, 2014.

[9] X. Jiang, T. Yu, X. Jiang, W. Zhang, L. Zhang, and J. Ma. Emergence of plasmid-mediated quinolone resistance genes in clinical isolates of *Acinetobacter baumannii* and *Pseudomonas aeruginosa* in Henan, China. *Diagnostic Microbiology and Infectious Disease*, 79(3):381–383, 2014.

[10] A. Bryce, A. D. Hay, I. F. Lane, H. V. Thornton, M. Wootton, and C. Costelloe. Global prevalence of antibiotic resistance in paediatric urinary tract infections caused by *Escherichia coli* and association with routine use of antibiotics in primary care: systematic review and meta-analysis. *BMJ*, 352:i939, 2016.

[11] K. Lewis. Recover the lost art of drug discovery. *Nature*, 485:439–440, 2012.

[12] H. Ham, A. Sreelatha, and K. Orth. Manipulation of host membranes by bacterial effectors. *Nature Reviews*, 9(September), 2011.

[13] J. Reidl and K. E. Klose. *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiology Reviews*, 26, 2002.

[14] A. Casadevall. Evolution of intracellular pathogens. *Annu Rev Microbiol*, 62: 19–33, 2008.

[15] J. Gruenberg and F. G. van der Goot. Mechanisms of pathogen entry through the endosomal compartments. *Nature Reviews Molecular Cell Biology*, 7(July): 495–505, 2006.

[16] P. Cossart and A. Helenius. Endocytosis of viruses and bacteria. *Cold Spring Harbor perspectives in biology*, 6(8), 2014.

[17] Y. Kumar and R. H. Valdivia. Leading a sheltered life: intracellular pathogens and maintenance of vacuolar compartments. *Cell Host and Microbe*, 5(6):593–601, 2009.

[18] K. Ray, B. Marteyn, P. J. Sansonetti, and C. M. Tang. Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nature reviews. Microbiology*, 7(5): 333–40, 2009.

[19] J. Fierer and D. G. Guiney. Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. *The Journal of Clinical Investigation*, 107(7): 775–780, 2001.

[20] B. Coburn, G. A. Grassl, and B. B. Finlay. *Salmonella*, the host and disease: a brief review. *Immunology and Cell Biology*, 85:112–118, 2007.

[21] H. K. de Jong, C. M. Parry, T. van der Poll, and W. J. Wiersinga. Host-pathogen interaction in invasive Salmonellosis. *PLoS Pathogens*, 8(10):1–9, 2012.

[22] S. E. Majowicz, J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O. Brien, T. F. Jones, A. Fazil, and R. M. Hoekstra. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clinical Infectious Diseases*, 50:882–889, 2010.

[23] J. A. Crump and E. D. Mintz. Global trends in typhoid and paratyphoid fever. *Clinical Infectious Diseases*, 50:241–6, 2010.

[24] F. Ramos-Morales. Impact of *Salmonella* enterica type III secretion system effectors on the eukaryotic host cell. *ISRN Cell Biology*, 2012:1–36, 2012.

[25] R. L. Santos, M. Raffatellu, C. L. Bevins, L. G. Adams, C. Tukel, R. M. Tsolis, and A. J. Baumler. Life in the inflamed intestine, *Salmonella* style. *Trends in Microbiology*, 17:498–506, 2009.

[26] B. B. Finlay, S. Ruschkowski, and S. Dedhar. Cytoskeletal rearrangements accompanying *Salmonella* entry into epithelial cells. *Journal of Cell Science*, 99 (Pt 2):283–296, 1991.

[27] B. Misselwitz, N. Barrett, S. Kreibich, P. Vonaesch, D. Andritschke, S. Rout, K. Weidner, M. Sormaz, P. Songhet, P. Horvath, M. Chabria, V. Vogel, D. M. Spori, P. Jenny, and W.-D. Hardt. Near surface swimming of *Salmonella* Typhimurium explains target-site selection and cooperative invasion. *PLoS pathogens*, 8(7):e1002810, 2012.

[28] M. A. Bakowski, V. Braun, and J. H. Brumell. *Salmonella*-containing vacuoles: directing traffic and nesting to grow. *Traffic*, 9(12):2022–2031, 2008.

[29] O. Steele-Mortimer. The *Salmonella*-containing vacuole – moving with the times. *Current Opinion in Microbiology*, 11(1):38–45, 2008.

[30] T. A. Agbor and B. A. Mccormick. *Salmonella* effectors: important players modulating host cell function during infection. *Cellular Microbiology*, 13(12): 1858–1869, 2011.

[31] C. Beuzón, S. Salcedo, and D. Holden. Growth and killing of a *Salmonella* enterica serovar Typhimurium sifA mutant strain in the cytosol of different host cell lines. *Microbiology*, 148:2705–2715, 2002.

[32] C. L. Birmingham, A. C. Smith, M. a. Bakowski, T. Yoshimori, and J. H. Brumell. Autophagy controls *Salmonella* infection in response to damage to the *Salmonella*-containing vacuole. *The Journal of Biological Chemistry*, 281 (16):11374–83, 2006.

[33] L. Knodler. *Salmonella* enterica: living a double life in epithelial cells. *Current Opinion in Microbiology*, 23:23–31, 2015.

[34] T. L. M. Thurston, M. P. Wandel, N. von Muhlinen, Á. Foeglein, and F. Randow. Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. *Nature*, 482(7385):414–418, 2012.

[35] L. A. Knodler, B. A. Vallance, J. Celli, S. Winfree, B. Hansen, M. Montero, O. Steele-mortimer, B. A. Vallanceb, S. Winf, and M. Monterob. Dissemination of invasive *Salmonella* via extrusion of mucosal epithelia. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41):17733–17738, 2010.

[36] P. Malik-Kale, S. Winfree, and O. Steele-Mortimer. The bimodal lifestyle of intracellular *Salmonella* in epithelial cells: replication in the cytosol obscures defects in vacuolar replication. *PloS ONE*, 7(6):e38732, 2012.

[37] L. A. Knodler, V. Nair, and O. Steele-mortimer. Quantitative Assessment of Cytosolic *Salmonella* in Epithelial Cells. *PLoS ONE*, 9(1), 2014.

[38] J. Fredlund and J. Enninga. Cytoplasmic access by intracellular bacterial pathogens. *Trends in Microbiology*, 22(3):128–37, 2014.

[39] H. B. Yu, M. A. Croxen, A. M. Marchiando, R. B. R. Ferreira, K. Cadwell, and L. J. Foster. Autophagy facilitates *Salmonella* replication in HeLa cells. *mBio*, 5:e00865–14, 2014.

[40] S. M. Eswarappa, V. D. Negi, S. Chakraborty, B. K. C. Sagar, and D. Chakravortty. Division of the *Salmonella*-containing vacuole and depletion of acidic lysosomes in *Salmonella*-infected host cells are novel strategies of *Salmonella* enterica to avoid lysosomes. *Infection and Immunity*, 78(1):68–79, 2010.

[41] G. Balazsi, A. Van Oudenaarden, and J. J. Collins. Cellular decision making and biological noise: From microbes to mammals. *Cell*, 144(6):910–925, 2011.

[42] H. Maamar and D. Dubnau. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Molecular Microbiology*, 56: 615–624, 2005.

[43] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305:1622–1626, 2004.

[44] E. Maisonneuve and K. Gerdes. Molecular mechanisms underlying bacterial persisters. *Cell*, 157(3):539–548, 2014.

[45] M. Ayrapetyan, T. C. Williams, and J. D. Oliver. Bridging the gap between viable but non-culturable and antibiotic persistent bacteria. *Trends in Microbiology*, 23(1):7–13, 2015.

[46] J. B. Kaper and V. Sperandio. Bacterial cell-to-cell signaling in the gastrointestinal tract. *Infection and Immunity*, 73(6):3197–3209, 2005.

[47] J. W. Hastings and K. H. Nealson. Bacterial bioluminescence. *Ann Rev Microbiol*, 31:549–95, 1977.

[48] M. E. Taga and B. L. Bassler. Chemical communication among bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, 100:14549–14554, 2003.

[49] J. Choi, D. Shin, and S. Ryu. Implication of quorum sensing in *Salmonella enterica* serovar typhimurium virulence : the luxS gene Is necessary for expression of genes in pathogenicity island 1. *Infection and Immunity*, 75(10):4885–4890, 2007.

[50] A. Mogilner, R. Wollman, and W. F. Marshall. Quantitative modeling in cell biology: what is it good for? *Developmental Cell*, 11:279–287, 2006.

[51] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson. Integrative analysis of cell cycle control in budding yeast. *Molecular Biology of the Cell*, 15(August):3841–3862, 2004.

[52] P. S. Shah, J. A. Wojcechowskyj, M. Eckhardt, and N. J. Krogan. Comparative mapping of host–pathogen protein–protein interactions. *Current Opinion in Microbiology*, 27:62–68, 2015.

[53] A. B. Doeschl-Wilson. The role of mathematical models of host-pathogen interactions for livestock health and production. *Animal*, 5:895–910, 2011.

[54] F. P. Davis, D. T. Barkan, N. Eswar, J. H. Mckerrow, and A. Sali. Host-pathogen protein interactions predicted by comparative modeling. *Protein Science*, 16: 2585–2596, 2007.

[55] A. A. Alonso, I. Molina, and C. Theodoropoulos. Modeling bacterial population growth from stochastic single-cell. *Applied and Environmental Microbiology*, 80 (17):5241–5253, 2014.

[56] R. M. Anderson and R. M. May. The population dynamics of microparasites and their invertebrate hosts. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 291:451–524, 1981.

[57] M. A. M. D. Aguiar, E. M. Rauch, and Y. Bar-Yam. Mean-field approximation to a spatial host-pathogen model. *Physical Review E*, 67:1–4, 2003.

[58] D. N. Macklin, N. A. Ruggero, and M. W. Covert. The future of whole-cell modeling. *Current Opinion in Biotechnology*, 28:111–115, 2014.

[59] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-garcia, J. I. Glass, and M. W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150:389–401, 2012.

[60] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. A. Guertin, J. H. Chang, R. a. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10):R100, 2006.

[61] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(January): 146–165, 2004.

[62] K. Jaqaman, D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser. Robust single-particle tracking in live-cell time-lapse sequences. *Nature Methods*, 5(8):695–702, 2008.

[63] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[64] S. Kreibich, M. Emmenlauer, J. Fredlund, P. Rämö, C. Münz, C. Dehio, J. Enninga, and W.-D. Hardt. Autophagy proteins promote repair of endosomal membranes damaged by the *Salmonella* type three secretion system 1. *Cell Host & Microbe*, 18(5):527–537, 2015.

[65] D. Drecktrah, S. Levine-Wilkinson, T. Dam, S. Winfree, L. A. Knodler, T. A. Schroer, and O. Steele-Mortimer. Dynamic behavior of *Salmonella*-induced membrane tubules in epithelial cells. *Traffic*, 9(12):2117–29, 2008.

[66] X. Z. Huang, B. Tall, W. R. Schwan, and D. J. Kopecko. Physical limitations on *Salmonella* typhi entry into cultured human intestinal epithelial cells. *Infection and Immunity*, 66(6):2928–2937, 1998.

[67] M. Newman. *Computational Physics*. Createspace Independent Publishing, 2012.

[68] B. Misselwitz, S. Dilling, P. Vonaesch, R. Sacher, B. Snijder, M. Schlumberger, S. Rout, M. Stark, C. von Mering, L. Pelkmans, and W.-D. Hardt. RNAi screen of *Salmonella* invasion shows role of COPI in membrane targeting of cholesterol and Cdc42. *Molecular Systems Biology*, 7(474):474, 2011.

[69] L. Kamentsky, T. R. Jones, A. Fraser, M.-a. Bray, D. J. Logan, K. L. Madden, V. Ljosa, C. Rueden, K. W. Eliceiri, and A. E. Carpenter. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. *Bioinformatics*, 27(8):1179–1180, 2011. doi: 10.1093/bioinformatics/btr095.