

COMPUTATIONAL APPROACHES TO COMPLEX BIOLOGICAL PHENOMENA

By

Joshua Luke Franklin

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Microbiology and Molecular Genetics – Doctor of Philosophy

2021

# ABSTRACT

## COMPUTATIONAL APPROACHES TO COMPLEX BIOLOGICAL PHENOMENA

By

Joshua Luke Franklin

Biological systems can be difficult to understand due to a vast array of interacting phenomena. The result is that some seemingly "easy" questions go unanswered. For example, we have long known that bacteria utilize many distinct flagellar configurations, but in most cases it remains unclear why they do so. We know that cell differentiation is critical to many biological processes, yet we still do not fully understand how such spatiotemporal patterning occurs. Despite mutation being one of the driving forces of evolution, we still have a hazy understanding of how organisms respond and adapt to high mutation rates. However, advances in technology, modeling, and experimental techniques have enabled us to investigate the small and nuanced effects that can answer these questions.

In this dissertation, I have used modern computational tools and statistical techniques to investigate evolutionary and behavioral processes. Agent-based models of evolution and flagellar inheritance have allowed me to investigate the evolution of mutational robustness and trade-offs associated with flagellar motility, respectively. By writing Bayesian mixed-effect models, I have been able to precisely quantify the metabolic cost of producing flagella and describe spatiotemporal patterns of cell differentiation within fruiting bodies of *Myxococcus xanthus*. Careful quantitative modeling of biological phenomena can help cut through the complexity of these systems. As computational power continues to grow and software continues to become more sophisticated, these computational approaches will both become more powerful and easier to use. Computational approaches will not replace experiments in most cases; instead, computational models can direct experiments, which can themselves direct new modeling efforts, in an iterative loop of progressing knowledge.

Copyright by  
JOSHUA LUKE FRANKLIN  
2021

For Heather, with whom  
laughs come freely as raindrops  
in a thunderstorm.



## ACKNOWLEDGEMENTS

It may be cliché to say that it takes a village to raise a Ph.D., but it is grounded in truth. Were it not for the support, friendship, and mentorship of dozens of people, I can safely say I would not have completed graduate school. Unfortunately, I do not have room to acknowledge everyone who has supported me, since I would have to list almost everyone I have run into during my stay at Michigan State University (MSU). Nevertheless, there are several people who have been particularly important to me, and I would like to thank them here.

I would like to thank my committee members, Dr. Christoph Adami, Dr. Daniel Ducat, and Dr. Jonathan Hardy, for their guidance. They have provided me with valuable scientific feedback and helpfully voiced their support of my work and my career trajectory.

I would like to thank my mentor, Dr. Yann Dufour, for having countless meetings with me. The hours we spent pair programming (read: debugging [read: struggling with]) models in brms and STAN were well worth the effort. Dr. Dufour has incredible intensity and clarity of thought, and I am a much better scientist for having joined his lab.

I am nowhere near skilled enough with words to convey the care my graduate student friends have shown me. Dr. Kyle Card was the first student I met at MSU and has been a close friend for my entire graduate career. I always look forward to our next conversation about evolution, consciousness, and the newest Vsauce video. I also appreciate Dr. Card's genuinely uncanny ability to make me keel over with laughter. Dr. Nkrumah Grant is possibly the most positive and supportive person I have ever met. Despite the distance now separating us, I know we will always be friends. Dr. Thomas LaBar is a great mentor, friend, and inspiration, as his path from mathematics to biology has in some ways mirrored my own trajectory. Dr. Paula Roszczenko and Dr. Maciej Jasiński were fantastic friends during their time at MSU. I fondly remember all our activities together, from eating pastries at Chapelure to evenings spent working on Pudding's suspension. Dr. Nhu Nguyen was

the best lab mate anyone could ask for, and her undying support and friendship has been invaluable to surviving my graduate career. Her perseverance and work ethic in spite of the daily ups and downs of graduate student life were, and continue to be, inspiring to me.

Of course, my family has made me who I am today. I feel my mother's unwavering support and love every time I speak with her. Her path has not been easy, but nevertheless she has climbed the nursing ladder from CNA to RN to DNP, all while raising me and supporting my endeavors. My stepfather has been a deeply important figure in my life, and I can only aspire to have as many interesting stories when I reach his age. I appreciate the unyielding confidence of my father, who never, ever, fails to tell me how proud he is of me. I only hope he knows how proud I am of him.

While William Halle is not related to me by blood, he may as well be my brother. From coding, to gaming, to sending absurd messages, we have formed a connection that cannot be destroyed by time, space, or lack thereof.

Finally, my fiancée Heather has supplied an endless stream of patience, encouragement, and laughs as I needed them. She is the only person I could have been with *literally* every hour of every day for months of quarantine, and still find myself growing closer to every day. I look forward to spending the rest of our lives laughing together on our walks beneath the stars.

# TABLE OF CONTENTS

|  |      |
|--|------|
| LIST OF TABLES . . . . .   | viii |
| LIST OF FIGURES . . . . .  | ix   |
| INTRODUCTION . . . . .   | 1    |
| CHAPTER 1 - MAPPING THE PEAKS: FITNESS LANDSCAPES OF THE FITTEST<br>AND THE FLATTEST . . . . .   | 6    |
| Abstract . . . . .   | 7    |
| Background . . . . .   | 7    |
| Materials and Methods . . . . .  | 9    |
| Results . . . . .  | 14   |
| Discussion . . . . .   | 17   |
| Conclusions . . . . .  | 21   |
| Supporting Information . . . . .   | 23   |
| CHAPTER 2 - CELL DENSITY, ALIGNMENT, AND ORIENTATION CORRE-<br>LATE WITH C-SIGNAL-DEPENDENT GENE EXPRESSION DURING <i>MYXO-<br/>COCCUS XANTHUS</i> DEVELOPMENT . . . . . | 25   |
| Abstract . . . . .   | 26   |
| Introduction . . . . .   | 27   |
| Results . . . . .  | 29   |
| Discussion . . . . .   | 39   |
| Materials and Methods . . . . .  | 43   |
| Supporting Information . . . . .   | 48   |
| CHAPTER 3 - COST-BENEFIT ANALYSIS OF PERITRICHIOUS FLAGELLATION<br>IN LOW VISCOSITY ENVIRONMENTS . . . . .   | 71   |
| Abstract . . . . .   | 72   |
| Introduction . . . . .   | 72   |
| Results . . . . .  | 74   |
| Discussion . . . . .   | 83   |
| Materials and Methods . . . . .  | 85   |
| Supporting Information . . . . .   | 89   |
| CHAPTER 4 - CONCLUSIONS AND FUTURE DIRECTIONS . . . . .  | 92   |
| Introduction . . . . .   | 93   |
| Chapter 1 . . . . .  | 93   |
| Chapter 2 . . . . .  | 96   |
| Chapter 3 . . . . .  | 97   |
| Closing recommendations . . . . .  | 101  |
| BIBLIOGRAPHY . . . . .   | 106  |

## LIST OF TABLES

|           |                                     |    |
|-----------|-------------------------------------|----|
| Table 2.1 | <b>Strains used in this study.</b>  | 52 |
| Table 2.2 | <b>Plasmids used in this study.</b> | 52 |
| Table 2.3 | <b>Primers used in this study.</b>  | 53 |

## LIST OF FIGURES

|            |  |    |
|------------|--|----|
| Figure 1.1 | <b>Populations adapt to different fitness peaks in high mutation rate environments.</b> a) Fitness values for the most abundant genotype from each population after the <i>Initial Adaptation</i> experiment (denoted by “Initial”) and after adaptation to low (“Fit”) and high (“Flat”) mutation rate environments. For boxplots, lines are the median values, boxes show the interquartile range, and whiskers are 1.5 times the interquartile range or the minimum/maximum value; points are outliers. Overlapping notches suggest that the difference in medians is not significant. b) Relative average fitness values over the course of the <i>Mutation Rate Adaptation</i> experiment. Lines are the mean value across all populations and the shaded regions are standard deviations. Blue represents fit genotypes and red represents flat genotypes. . . . . | 15 |
| Figure 1.2 | <b>Competition outcome between fit and flat genotypes is determined by the mutation rate.</b> Outcome from competition experiments between one fit/flat genotype pair. Black line is the mean fraction of the population consisting of the fit genotype as a function of time (measured in generations). Green shading represents one standard deviation. Each subplot shows the results for a given genomic mutation rate for five replicates. . . .  | 16 |
| Figure 1.3 | <b>Description of local fitness landscapes for fit and flat genotypes.</b> a-e) Distribution of fitness effects with mutations grouped by their broad effect. Boxplots as previously described. Lineage refers to whether the data is for fit genotypes (blue) or flat genotypes (red). f) Average relative fitness of all one-step mutants for each fit and flat genotype. . . . .  | 17 |
| Figure 1.4 | <b>Summarized distribution of fitness effects for mutants with multiple mutations.</b> Proportion of mutations with a mutational effect classification for mutants up to four mutations away from the wild-type sequence. See text for definitions of different classes of mutations. Distance is the number of mutations away from the wild-type sequence. Data for mutants with distance of one is the same as in Figure 3. Blue (red) boxplots are the proportions from the fit (flat) genotypes. Boxplots as previously described. . . . .   | 18 |
| Figure 1.5 | <b>Distribution of <math>\alpha</math> and <math>\beta</math> values for fit and flat genotypes.</b> Distributions of $\alpha$ and $\beta$ for each treatment. $\alpha$ is a measure of single-mutant fitness, while $\beta$ is a measure of epistasis (see Methods for further details). Boxplots as previously described. . . . .  | 19 |
| Figure 1.6 | <b>Histogram of the residual standard error (RSE) of the fits of Equation 1 to all fit and flat peaks.</b> . . . . .   | 23 |

Figure 1.7 **Boxplot of the residuals of the fit to Equation 1 for the fit peaks (blue) and flat peaks (red) separately.** As in the boxplots of the main text, lines are the median values, boxes show the interquartile range, and whiskers are 1.5 times the interquartile range of the minimum or maximum value; points are outliers. . . . . 24

Figure 2.1 **Visualization of transitioning cells near the bottom of a nascent fruiting body.** *M. xanthus* wild-type strain DK1622 was starved to initiate development under submerged culture conditions. FM 4-64 (2  $\mu\text{g/mL}$ ) was added at the start of starvation. Confocal images were acquired at the indicated times poststarvation (PS) to show FM 4-64 staining of cell membranes (red fluorescence shown as grayscale). Images are representative of three biological replicates. (A) Image of an optical section near the bottom of the same nascent fruiting body (NFB) during development. Red arrow, rod-shaped cell. Green arrows, transitioning cells (TCs). Blue arrow, spore. Bar, 20  $\mu\text{m}$ . (B) Cartoon depiction of tangential orientation. Upper left quadrant depicts one tangentially oriented rod. Arrow indicates radius of NFB (gray). Other quadrants depict many smaller tangentially oriented rods. (C) Enlarged images of representative cells. Images from panel A were enlarged to show cells with a particular shape (a rod at 18 h, TCs at 24 and 30 h, and a spore at 42 h). Colored arrows point to the same cell classes as in panel A. Bar, 2  $\mu\text{m}$ . (D) Tracking the fate of TCs. Images from Movie S2. Orange arrows, TC that becomes a spore. Magenta arrows, TC that changes shape. White arrows, TC whose fate is uncertain. Bar, 2  $\mu\text{m}$ . . . . . 54

**Figure 2.2 Classification of cellular morphologies and the temporal patterns of rods, transitioning cells, and spores.** Labeled cells that produce a fluorescent protein under control of a vanillate-inducible promoter (strain YH7, YH8, YH14, or Y15) were mixed with unlabeled wild-type cells (strain DK1622) at a ratio of 1 to 5 and starved under submerged culture conditions. Vanillate (0.5 mM) was added at the start of starvation. Z-stacks of optical sections of the same nascent fruiting body (NFB) for each of three biological replicates were acquired at the indicated times poststarvation (PS) to show fluorescence from tdTomato or mNeonGreen. Segmented cells from all z-stacks were classified as rods, transitioning cells (TCs), or spores (5256, 1459, and 3213 cells, respectively). (A) Confocal images of the same NFB. Images show an optical section near the bottom of a representative NFB formed by the mixture of strains YH8 and DK1622. Red fluorescence from YH8 cells is shown as grayscale. Red arrow, rod. Green arrows, TCs. Blue arrow, spore. Bar, 20  $\mu\text{m}$ . (B) Enlarged images of representative cells. Images from panel A were enlarged to show cells with a particular shape (a rod at 27 h, TCs at 30 and 36 h, a spore at 48 h). Colored arrows point to the same cell classes as in panel A. Bar, 2  $\mu\text{m}$ . (C) Representative cellular morphologies. Three-dimensional renderings reconstructed from the image stacks are shown for three cells of each class. Apparent features  $<0.5 \mu\text{m}$  are artifacts created by shading during the rendering process. Bar, 1  $\mu\text{m}$ . (D) Proportion of each cell class at the indicated times PS. Dot, median. Shaded region, 90% credible interval. . . . . 55

**Figure 2.3 Radial patterns of cell density, alignment, and orientation early correlate with spore proportion later as nascent fruiting bodies mature.** In the experiments described in the legend of Figure 2.2 and in the text, segmented cells from all z-stacks were classified as rods, transitioning cells, or spores. (A) Radial distribution of cell classes and density. The proportion of each cell class at different times poststarvation is shown radially from the center (0  $\mu\text{m}$ ) to the edge (60  $\mu\text{m}$ ) of nascent fruiting bodies (NFBs). Total cell density is normalized to the maximum between the center and the edge of NFBs. Line, median. Shaded region, 90% credible interval. (B) Radial patterns of rod neighbor alignment and tangential orientation. Alignment is the cosine of the angle of a rod to its neighbors averaged using a Gaussian kernel ( $\sigma = 2.5 \mu\text{m}$ ) (1, perfect alignment; 0, orthogonal). Orientation is the cosine of the angle of a rod with the circumference of the NFB (1, tangent to the circumference; 0, orthogonal). (C) Cross-correlations between the radial patterns of total cell density, rod neighbor alignment, and rod tangential orientation early, and that of the spore proportion later. Values are Pearson correlation coefficients and color shade indicates strength of positive (purple) or negative (yellow) correlation (\*  $p < 0.05$ , \*\*  $p < 0.01$ ). . . . . 56

**Figure 2.4 C-signal-dependent gene expression increases in transitioning cells and spores as nascent fruiting bodies mature.** As described in the Figure 2.2 legend, strains YH14 bearing  $P_{dev}$ -tdTomato (designated *dev*) or YH15 bearing  $P_{fmgE}$ -tdTomato (designated *fmgE*) were mixed with unlabeled wild-type strain DK1622 at a ratio of 1 to 5 and co-developed under submerged culture conditions. Both the *dev* and *fmgE* strains also contain Pvan-mNeonGreen. Vanillate (0.5 mM) was added at the start of starvation. Z-stacks of confocal images of the same nascent fruiting body (NFB) were acquired at the indicated times post starvation (PS), measuring both red and green fluorescence. Images are representative of three biological replicates. (A) Confocal images of the same NFB. Images show merged (M) red and green fluorescence. Red arrows, rods. Green arrows, transitioning cells (TCs). Blue arrows, spores. Bar, 20  $\mu$ m. (B) Enlarged images of representative cells. Images from panel A were enlarged to show green (G), red (R), or M fluorescence of cells with a particular shape (rods at 27 and 30 h, TCs at 30 and 36 h, spores at 36 and 48 h). Colored arrows point to the same cell classes as in panel A. Bar, 2  $\mu$ m. (C) Quantification of gene expression. Segmented cells from z-stacks were classified as rods, TCs, or spores based on green fluorescence. For each cell class, the  $\log_{10}$ (relative fluorescence intensity) (i.e., red/green) at each time PS is plotted for the *dev* and *fmgE* strains as indicated. Dot, median. Vertical line, 90% credible interval. (D) Gene expression in TCs and spores. The relative fluorescence intensity of TCs and spores was normalized to that of rods at 27 h (dashed line at  $y = 0$ ). Dot, median. Vertical line, 90% credible interval. . . . . 57

**Figure 2.5 Radial patterns of cell density, alignment, and orientation early correlate with C-signal-dependent gene expression later as nascent fruiting bodies mature.** In the experiment described in the legend of Figure 2.4 and in the text, both red and green fluorescence was quantified for all segmented cells from the z-stacks. The results from three biological replicates were combined. (A) Radial distribution of C-signal-dependent gene expression. For transitioning cells and spores, the  $\log_{10}$ (relative fluorescence intensity) (i.e., red/green) for the *dev* and *fmgE* strains at each time post-starvation is shown radially from the center (0  $\mu$ m) to the edge (60  $\mu$ m) of nascent fruiting bodies. Line, median. Shaded region, 90% credible interval. (B) Cross-correlations between the radial patterns of total cell density, rod alignment, and rod tangential orientation early, and that of the *dev* strain relative fluorescence intensity (RFI) later. Values are Pearson correlation coefficients and color shade indicates strength of positive (purple) or negative (yellow) correlation. (C) Cross-correlations as in panel B except with the *fmgE* strain RFI (\*  $p < 0.05$ , \*\*  $p < 0.01$ ). . . . . 59



Figure 2.6 **Model showing the inferred radial pattern of C-signaling based on observed patterns of cellular arrangement, morphology, and gene expression as nascent fruiting bodies mature.** The radial distributions of cell classes, cell density, and the neighbor alignment and tangential orientation of rods are depicted at three time points based on Figure 2.3A and 2.3B. The inferred radial pattern of C-signaling at 30 h poststarvation is based on correlations with later C-signal-dependent *dev* and *fmgE* expression in transitioning cells and spores (Fig. 2.5). . . . . 60

Figure 2.7 **Cartoon and scanning electron micrographs of *M. xanthus* development.** The cartoon at the top depicts starving rod-shaped cells, which coordinate their movements to build a mound, with some cells lysing (dashed outline) during the process. The mound matures into a fruiting body as some rods differentiate into round spores, while other rods remain outside the fruiting body as peripheral rods. The scanning electron micrographs at the bottom show the mound and fruiting body stages. The cartoon is adapted from [89] and the micrographs are from [94] . . . . . 61

Figure 2.8 **FM 4-64 does not interfere with *M. xanthus* development.** Wild-type strain DK1622 was subjected to starvation under submerged culture conditions. FM 4-64 (2  $\mu$ g/mL) was added at the start of starvation or withheld. (A) Bright-field images of the bottom of nascent fruiting bodies (NFBs). The same NFB was examined at the indicated times poststarvation (PS). Darkening of mounds typically reflects the formation of spores. Images are representative of three biological replicates. Bar, 20  $\mu$ m. (B) Quantification of sonication-resistant spores. Samples were collected at 42 h PS, sonicated, and the number of spores/mL was determined. In the main graph, each symbol represents one biological replicate. The insert shows the posterior probability distribution of the difference in the number of sonication-resistant spores/mL in response to 2 vs. 0  $\mu$ g/mL of FM 4-64. Vertical line, mean. Thick and thin horizontal lines, 90% and 98% credible intervals, respectively. . . . . 62

Figure 2.9 **Temporal pattern of cellular shape near the bottom of nascent fruiting bodies.** Based on our observations of cells undergoing submerged culture development and visualized by membrane staining with FM 4-64, the proportion of cells exhibiting the indicated shapes at different times poststarvation was approximated. Rods undergo lysis or become transitioning cells or peripheral rods. Transitioning cells may undergo lysis or persist, but some become spores. . . . . 63

Figure 2.10 **Cell classification using Gaussian mixture modeling.** Each point on the scatter plots represents a cell from the aggregated data across time and fruiting bodies, collected in experiments described in the legend of Figure 2 and in the text. Clusters were determined to be rods (red), transitioning cells (green), or spores (blue) according to their morphology (equivalent diameter, surface area, length, and mean width). The Pearson correlation coefficients are indicated for each pair of properties for all cells (black font) and for each cell class as indicated (\*\* indicates p-value < 0.01, \*\*\* indicates p-value < 0.001). Histograms on the left are the marginal distributions for each cell class for the corresponding morphological property. . . . . 64

Figure 2.11 **Morphological property distributions of cells classified as rods, transitioning cells, or spores.** In the experiments described in the legend of Figure 2 and in the text, segmented cells from all z-stacks were classified as rods, transitioning cells, or spores. The graph on the left in each panel shows the distribution of the indicated property for each cell class. Gray dot, individual cell. Black dot, median. The graph on the right in each panel shows the posterior probability intervals of the property difference for the indicated pair of cell classes. Dot, median. Thick and thin horizontal lines, 90% and 98% credible intervals, respectively. (A) Equivalent diameter. (B) Cell length. (C) Mean cell width. (D) Sphericity. . . . . 65

Figure 2.12 **Spatiotemporal distribution of cell classes and density in individual nascent fruiting bodies.** In the experiments described in the legend of Figure 2 and in the text, segmented cells from z-stacks were classified as rods, transitioning cells, or spores. The proportion of each cell class at the indicated time poststarvation is shown radially from the center (0  $\mu\text{m}$ ) to the edge (60  $\mu\text{m}$ ) of one representative nascent fruiting body (NFB) formed by mixing the strain indicated on the right with unlabeled strain DK1622. Total cell density is normalized to the maximum between the center and the edge of the NFB. Line, median. Shaded region, 90% credible interval. . . . . 66

Figure 2.13 **A *csgA* mutant fails to form nascent fruiting bodies and fails to express C-signal-dependent genes.** *M. xanthus csgA* mutant strains YH46 bearing  $P_{dev}$ -*tdTomato* (designated *dev*) or YH45 bearing  $P_{emphfmgE}$ -*tdTomato* (designated *fmgE*) were starved under submerged culture conditions. Both the *dev* and *fmgE* strains also contain  $P_{van}$ -*mNeonGreen*. Vanillate (0.5 mM) was added at the start of starvation. Wells were examined at 6-h intervals from 18 to 48 h poststarvation (PS) using confocal microscopy to visualize both red and green fluorescence. Nascent fruiting bodies failed to form, cells remained rod shaped, and the rods exhibited green fluorescence but showed very little red fluorescence. Confocal images near the bottom of the biofilm were acquired at the indicated times PS. Images show merged (M) red and green fluorescence. Bar, 2  $\mu\text{m}$ . . . . . 67

Figure 2.14 **C-signal-dependent gene expression in individual cells of nascent fruiting bodies.** In the experiment described in the legend of Figure 4 and in the text, segmented cells from z-stacks were classified as rods, transitioning cells, or spores based on green fluorescence. The results from three biological replicates were combined. (A) Distributions of red and green fluorescence intensities. For each cell (dot), the  $\log_{10}(\text{red fluorescence intensity})$  and the  $\log_{10}(\text{green fluorescence intensity})$  at each time poststarvation (PS) is plotted for the *dev* and *fmgE* strains as indicated. Cells in the lower left quadrant of each plot (i.e., exhibiting low red and green fluorescence) were eliminated from analyses in panels B-D. (B) Distributions of red fluorescence intensity. For each cell class, the  $\log_{10}(\text{red fluorescence intensity})$  at each time PS is plotted for the *dev* and *fmgE* strains as indicated. Dot, individual cell. Bar, median. (C) Distributions of green fluorescence intensity. Same as panel B except the  $\log_{10}(\text{green fluorescence intensity})$  is plotted. (D) Distributions of relative fluorescence intensity. Same as panel B except the  $\log_{10}(\text{relative fluorescence intensity})$  (i.e., red/green) is plotted. . . . . 68

Figure 2.15 **Vanillate-inducible gene expression in individual cells of nascent fruiting bodies.** As described in the Figure 2 legend, strain YH7 bearing  $P_{van}-mNeonGreen$  or YH8 bearing  $P_{van}-tdTomato$  were mixed with unlabeled wild-type strain DK1622 cells at a ratio of 1 to 5 and co-developed under submerged culture conditions. Vanillate (0.5 mM) was added at the start of starvation. Z-stacks of confocal images of the same nascent fruiting body were acquired at the indicated times poststarvation (PS) without changing the microscope settings. Segmented cells from z-stacks were classified as rods, transitioning cells, or spores. The results from three biological replicates were combined. For each cell class, the distribution of  $\log_{10}(\text{fluorescence intensity})$  (i.e., green or red) at each time PS is plotted. Dot, individual cell. Bar, median. . . . . 70

Figure 3.1 **Ectopic control of flagellar expression and flagellar hook detection in *S. enterica*.** A) Composite image of fluorescently labeled EM8242 after induction of flagellar expression ( $[\text{AnTc}] = 2,000 \text{ pg/mL}$ ). DNA was stained with DAPI (blue), flagellar hook with a maleimide conjugated fluorescent dye (red), and flagellar filament with anti-flagellin antibodies conjugated to a fluorescent dye (green). B) Number of flagellar hook foci as a function of inducer concentration in single cells (750 cells per condition). Each color dot represents a cell (red: EM8242, blue: TH9677). The estimated mean for each population is indicated in black (dot: mean, line: 95% CI). C) Mean lengths of flagella as a function of the number of foci in each cell (symbols represent different AnTc concentrations). The relationship was fitted using a linear model with a gamma response (black line: mean, light and dark blue: 50% and 95% CI). . . . . 76

Figure 3.2 **The relative metabolic cost of flagellar synthesis.** A) Calculating growth rate by monitoring change in culture optical density at 600 nm ( $OD_{600}$ ) over time. Representative measurements from non-induced EM8242 (red circle) and TH9677 (blue circle) are overlaid. Growth rates ( $r$ ), background OD, and OD of the inoculum were estimated by least-square fitting of an exponential growth model (solid lines) while cells were at low density. B) Calculated growth rates as a function of inducer concentration for EM8242 (red) and TH9677 (blue). Each dot represents a single batch culture (between 30 and 200 per condition). The estimated growth rate for each condition is indicated in black (dot: mean, line: 95% CrI). C) Regression of growth rate as a function of number of foci. Each dot represents a single inducer concentration (EM8242 in red) or wild type FlhDC activity (TH9677 in blue). The crossed lines represent the 95% credible interval from Figure 3.1B and 3.2B. The black line represents the mean of the regression model (dark and light blue areas: 50% and 95% CrI). Inset: Posterior probability distribution of the relative cost (slope/intercept) of one flagellum of the growth rate (dot: mean, line: 95% CrI). . . . . 78

Figure 3.3 **Swimming performance of TH9677 with varying number of flagella.** A) Illustrative examples of single cell tracking followed by detection of fluorescently labeled flagellar hooks after immobilizing cells with a UV-activated hydrogel. B-F) Linear regressions of behavioral parameters calculated from cell trajectories against number of foci from labeled flagellar hooks. Varying number of foci does not explain the variability in swimming performance. Each dot represents a single cell. The black line represents the mean of the regression model (dark and light blue areas: 50% and 95% CrI). . . . 79

Figure 3.4 ***In silico* evolution of the optimal number of flagella in different conditions.** A) Schematic representation of the agent-based model. Flagellated agents swim up with a speed inversely proportional to the chemical gradient (blue) and are pulled down by gravity. Position along the gradient and determines the resources available for growth and flagellar synthesis. The rate of flagellar synthesis is allowed to evolve by natural selection as the population grows until steady state is reached. B) New flagella (red dots) are inserted randomly on the agent body. Cell elongation follows one of three alternative models for cell wall synthesis (blue) to determine the allocation of flagella to daughter cells. C) Results from repeated *in silico* evolutionary experiments with different combinations of flagellar cost and growth reward from chemotaxis. Each dot represents the mean number of flagella per cell in an independent population after its evolutionary trajectory reached steady state. The solid lines represent the best fit from our mathematical model (with  $P(\neg flg)$  as a free parameter). Inset: Mean and 95% CrI for the parameter  $P(\neg flg)$  fitted to simulation results. . . . . 81

|            |  |     |
|------------|--|-----|
| Figure 3.5 | <b>Mathematical predictions of the optimal number of flagella.</b> A) Optimal number of flagella ( $n$ ) as a function of the reward to cost ratio for chemotaxis ( $\delta/c$ ) for different probability of not inheriting a flagellum after cell division ( $p$ ) from equation 3.2. B) Expected growth reward ( $d$ ) as a function of number of flagella ( $n$ ) for different probability of not inheriting a flagellum ( $p$ ) and a fixed cost of 1.17% of the growth rate per flagellum ( $c$ ) from equation 3.3. C) Expected number of flagella ( $n$ ) as a function of the probability of not inheriting a flagellum ( $p$ ) for different reward to cost ratios ( $\delta/c$ ) from equation 3.2. The parameter values estimated in this study for <i>S. enterica</i> are indicated with dotted lines. . . . . | 83  |
| Figure 3.6 | <b>Effect of AnTc on the growth rate of TH9677.</b> Each dot represents a single batch culture, and each color represents an independent experiment. The estimated growth rate for each condition is indicated in black (dot: mean, line: 95% CrI). The addition of AnTc had a very small positive effect on growth ( $5.6e-4 \text{ hr}^{-1}(\text{ng/ml})^{-1}$ [ $2.3e-4, 8.9e-4$ ] 95% CrI). . . . .   | 89  |
| Figure 3.7 | <b>Relative distribution of foci along the cell midline.</b> Most foci are located between the cell poles, with a nearly even distribution from 0.2 to 0.8, and a slight dip at the cell midpoint. The pole and midpoint may be crowded with other cellular machinery (eg. FtsZ complex). . . . .  | 90  |
| Figure 3.8 | <b>Optimal number of flagella for different combinations of flagellum costs and rewards.</b> Results from repeated <i>in silico</i> evolutionary experiments with different combinations of flagellar cost (color scale) and growth reward from chemotaxis. Each dot represents the mean number of flagella per cell in an independent population after its evolutionary trajectory reached steady state. Simulations were repeated using three growth models for new cell wall synthesis and flagella allocation between daughter cells (end: growth near cell pole, mid: growth from middle of cell length, random: diffusive growth along cell length). . . . .   | 91  |
| Figure 4.1 | <b>A bacterial “racetrack”.</b> The width and depth of the channel is about 150 $\mu\text{m}$ . The total length of the track is approximately two meters long. . . .  | 101 |

# INTRODUCTION

Author: Joshua L. Franklin

As our understanding of biology grows, so do the questions we are trying to answer. Behavioral studies of bacteria are becoming more nuanced and are considering the myriad of interactions between microbes in the environment. We are trying to predict the outcome of evolutionary processes with greater precision than ever before. We are also determined to answer the big questions in biology, the holy grail of which is the origin of life. As our ambition grows, so do the obstacles we face. Nuanced understanding of small, but important, effects is difficult in the face of biological variability and noise. The amount of data flowing from our experiments is becoming difficult to store, let alone process and comprehend. Some of our questions are so difficult that they cannot be easily tested in biological systems, as they would require an impossible degree of control or exploration of an enormous parameter space.

Nonetheless, I am optimistic about the future. Raw computational power has exploded over the past couple of decades, and the software required to utilize such power has been in rapid development. Statistical tools that were curiosities for want of computational power in the 1900's are now daily tools of many scientists. For example, Bayesian statistics was mostly infeasible prior to the computational revolution, but that is no longer the case. Tools like STAN and brms allow researchers to write and run complex multi-level Bayesian models with relative ease [22, 162]. Biologists can now easily write complex agent-based models to investigate phenomena with an unprecedented degree of control and precision [184]. Even the tools of our trade are improving, with microscopy and sequencing becoming more advanced by the day.

My research touches on a number of distinct topics, such as the evolution of mutational robustness, collective behavior of cells, and evolutionary trade-offs associated with motility. What unifies these topics are the computational approaches I have used to ask and answer questions that would otherwise be difficult to approach. In particular, agent-based models and mixed-effect Bayesian models are central to my work, because they allow me to write fairly simple code describing agent interactions or the structure of an experiment. Yet they

also provide rich outputs that can be nearly as deep as biological datasets. I hope that the following chapters of my dissertation provide interesting examples to inspire other biologists, and encourages them to try a new computational analysis or write a new statistical model.

The first chapter expands on the “Survival of the Flattest” (SOF) effect, which relates to how organisms adapt to high mutation rates [189]. Briefly, the authors of the original paper used the artificial life model Avida to evolve pairs of digital organisms in high and low mutation rate environments, then performed competitions between the resulting organisms. They found that slowly reproducing but mutationally robust organisms (“flat”) could outcompete more quickly reproducing but mutationally fragile organisms (“fit”) in a high mutation rate environment [189]. The SOF effect was later supported in biological experiments [33, 101, 150], but a key question remained: how were organisms adapting their genomes to high mutation rates? To investigate, we first replicated the original study, then characterized the local fitness landscapes by evaluating tens of millions of mutants for each pair of fit and flat organisms. We then binned the fitness effect of each mutation as lethal, deleterious, neutral, or beneficial relative to the reference genotype. We also investigated changes in epistasis by fitting a model to average fitness as a function of mutational distance. Both lines of evidence indicated that mutational robustness limits the number of deleterious mutations *that can be passed to offspring*. A surprising consequence is that robustness often results in both increased rates of neutral and lethal mutations, since lethal mutations cannot be inherited.

The second chapter relates to fruiting body development in *Myxococcus xanthus*. Starving *M. xanthus* cells cooperate to form a fruiting body that spreads environmentally resistant spores in search of fresh resources [196]. Fruiting body formation requires both coordination of cells to form the fruiting body structure and differentiation of a subset of cells into spores [93]. While fruiting body formation has long served as a model of cooperation and differentiation, the exact temporal and spatial dynamics of differentiation were not well understood. We wanted to better characterize these transitions by performing image analysis



of micrographs of fruiting bodies over time. However, the high cell density within a fruiting body makes segmentation and characterization of individual cells difficult. To overcome this hurdle, we seeded a small number of fluorescently labeled cells into developing fruiting bodies of wildtype (i.e., non-fluorescent) *M. xanthus*, allowing us to easily segment and characterize fluorescent micrographs despite the high cell density. We then used a Gaussian mixture model to cluster and classify individual cells as rods, transitioning cells, or spores. To interpolate the proportion of these categories across space, we used Bayesian generalized additive models. We also investigated cell orientation, relative to both the fruiting body and neighboring cells. Our results indicate the spatial patterning of cell-cell interactions may enhance contact-dependent C-signaling, a critical signal for sporulation and fruiting body development [93].

The third chapter relates to peritrichously flagellated bacteria. We know that flagella are metabolically expensive due to their protein content and propensity to be lost during evolution experiments in shaken flasks [138]. As such, we expect that natural selection will optimize the cost-benefit trade-off to produce an optimal number of peritrichous flagella for a given environment. To better understand the trade-offs involved, we quantified the cost of flagella using growth curve assays. Because growth curves are noisy and the cost of flagella is relatively small, we used a Bayesian mixed-effect model to increase both our accuracy and precision. We determined that each flagellum decreases the growth rate by about 1%, in line with previous estimates [138]. However, when we performed single-cell tracking under the microscope, we found that flagella number and swimming performance are not well correlated in low-viscosity (i.e. water-like) liquids. Given the high metabolic costs and lack of swimming benefits, why should cells make more than one flagellum? To investigate, we wrote an agent-based model of flagellar inheritance, which showed a logarithmic relationship between benefit/cost ratio and optimal flagella number. This model is consistent with bacteria that make a small number of peritrichous flagella (e.g. *Salmonella enterica*). However, cells producing many peritrichous flagella (e.g. *Bacillus subtilis*) appear to be subject to a different

evolutionary trade-off.

In the concluding chapter, I emphasize the importance of the computational and statistical methods used in each study. I also elaborate on more exploratory future directions for each project than would be present in a typical scientific article. I then recommend biologists embrace agent-based and mixed-effect models as key tools for expanding their research capabilities. I end with some suggestions for model-writers to make their models easier to re-use for educational purposes. Taken together, I hope this dissertation explores computational tools useful for biologists in an approachable manner and encourages them to write their own models for both research and educational purposes.

## CHAPTER 1 - MAPPING THE PEAKS: FITNESS LANDSCAPES OF THE FITTEST AND THE FLATTEST

Authors: Joshua L. Franklin\*, Thomas LaBar\*, and Christoph Adami

\*\*Co-first authors

This chapter was originally published as an article in *Artificial Life* 25: 250-262

Reproduced with permission from the publisher:

<https://direct.mit.edu/journals/pages/rights-permissions>

Figures have been renumbered from the original publication.

## Abstract

Populations exposed to a high mutation rate harbor abundant deleterious genetic variation, leading to depressed mean fitness. This reduction in mean fitness presents an opportunity for selection to restore adaptation through the evolution of mutational robustness. In extreme cases, selection for mutational robustness can lead to “flat” genotypes (with low fitness but high robustness) out-competing “fit” genotypes with high fitness but low robustness—a phenomenon known as “survival of the flattest”. While this effect was previously explored using the digital evolution system Avida, a complete analysis of the local fitness landscapes of “fit” and “flat” genotypes has been lacking, leading to uncertainty about the genetic basis of the survival of the flattest effect. Here, we repeated the survival of the flattest study and analyzed the mutational neighborhoods of fit and flat genotypes. We found that flat genotypes, compared to the fit genotypes, had a reduced likelihood of deleterious mutations as well as an increased likelihood of neutral and, surprisingly, of lethal mutations. This trend holds for mutants one to four substitutions away from the wild-type sequence. We also found that flat genotypes have, on average, no epistasis between mutations, while fit genotypes have, on average, positive epistasis. Our results demonstrate that the genetic causes of mutational robustness on complex fitness landscapes are multifaceted. While the traditional idea of the survival of the flattest effect emphasized the evolution of increased neutrality, others have argued for increased mutational sensitivity in response to strong mutational loads. Our results show that both increased neutrality and increased lethality can lead to the evolution of mutational robustness. Furthermore, strong negative epistasis is not required for mutational sensitivity to lead to mutational robustness. Overall, these results suggest that mutational robustness is achieved by minimizing *heritable* deleterious variation.

## Background

All well-adapted populations receive an influx of deleterious variation every generation due to *de-novo* mutations. This influx reduces the population’s average fitness and causes

a mutational load [2, 39]. For populations with high mutation rates, deleterious variation can alter the dynamics of natural selection and lead to selection acting not just on individual genotypes, but on genotypes and their likely mutant genotypes [13, 19, 124, 186]. This combination of a wild-type sequence and its mutants is known as a “quasispecies” and is thought to be a unit of selection in organisms such as the first RNA replicators [47, 48] and modern-day RNA viruses [100].

Since the quasispecies concept was first proposed [47, 48], many studies have analyzed the conditions and consequences of evolutionary dynamics in this strong mutational regime (e.g., [7, 87, 186]). One consequence of evolution under high mutation rates is that populations are expected to evolve mutational robustness by fixing mutations that decrease the likelihood (and/or decrease the average effect) of deleterious mutations [42, 188]. In the extreme case, this trend can lead to genotypes with lower fitness out-competing genotypes with higher fitness if the lower-fitness genotypes are more mutationally-robust [152]. This effect has been termed the “survival of the flattest” effect, where “flat” genotypes, that is, those with lower fitness but greater mutational robustness, out-compete “fit” genotypes, meaning those with higher fitness but decreased mutational robustness [189]. While this effect was first observed in the digital evolution system Avida, it was later demonstrated with experimental evolution of RNA viruses [33, 101, 150].

Even though the evolutionary dynamics of the survival of the flattest effect are well-understood, the exact genetic causes of robustness are unknown. Indeed, the differences in the distribution of fitness effects [52] (that is, the local fitness landscapes) between fit genotypes and flat genotypes have not been explored. While it is generally assumed that mutational robustness evolves due to the evolution of an increased number of neutral mutational neighbors [53, 173, 185], other authors have argued that increased mutational sensitivity may evolve [4, 87, 99, 126, 190]. Additionally, recent work has indicated that populations evolving under strong genetic drift are expected to evolve “drift-robust” genetic architectures, giving rise to genotypes with a decreased likelihood of small-effect deleterious mutations and an

increased likelihood of neutral and/or strongly-deleterious mutations [98]. A mapping of the local fitness landscapes of fit genotypes and flat genotypes is needed to compare the survival of the flattest effect to these other proposed evolutionary mechanisms.

To explore the local fitness landscapes of fit genotypes and flat genotypes, we first repeated the original survival of the flattest experiment. We then took genotypes adapted to either low or high mutation rates and obtained the distribution of fitness effects for these genotypes by examining mutants up to four substitutions away from the wild-type sequence. We found that flat genotypes had a greater likelihood of beneficial and neutral mutations (as expected for mutationally-robust genotypes) while fit genotypes had a greater likelihood of deleterious mutations, as expected from mutationally-sensitive genotypes. Contrary to expectations, we also found that flat genotypes had a greater likelihood of lethal mutations, suggesting that the survival of the flattest effect relies on a reduction of heritable deleterious mutation, which can be achieved by both an increase in neutral variations, as well as lethal mutations. These results illustrate the complexity of mutationally-robust genome architectures on non-trivial fitness landscapes.

## Materials and Methods

### Avida

Here, we describe the relevant details of Avida (version 2.14; our version, with added code for data analysis, is available at <https://github.com/joshf93/MappingThePeaks>) for the current experiments (see [128] for a full overview of the Avida software). In Avida, a finite population of computer programs ("avidians") compete for the resources (memory, space, and processing time) required for reproduction. Each program consists of a circular genome of computer instructions that encode the ability to self-replicate. During this replication process, instructions may be copied inaccurately, resulting in mutations that are passed onto an avidian's offspring. As avidians with faster replication speeds enter the population, they reproduce faster than slower-replicating avidians, leading to the spread of faster replicators and the extinction of slower replicators. In other words, there is selection for faster repli-

cators. Therefore, because there is heritable variation and selection in Avida, populations of avidians undergo evolution by natural selection [133]. Avida has been used to test many concepts difficult to test in biological systems [1, 37, 59, 60, 97, 103, 163].

The Avida world consists of a toroidal grid of  $N$  cells; each cell can be occupied by at most one avidian and thus  $N$  is the maximum population size. During reproduction, offspring avidians are placed into one of the nine adjacent cells of its parent, including the cell that contains the parent. If some of these cells are unoccupied, a random empty cell is chosen. If all neighboring cells are occupied, a random cell is chosen, its occupant is removed, and the new avidian then occupies the cell. This random selection of replacement adds an element of genetic drift to Avida.

The limited number of cells in the Avida environment is one of the limiting resources over which avidians compete. The other limiting resource is the opportunity to execute the instructions in an avidian's genome. The resource required to execute one instruction is called a "Single Instruction Processing" unit (SIP). During each unit of time in Avida (called an update)  $30N$  SIPs are available to the population for execution. These SIPs are probabilistically distributed to avidians according to a figure of merit that is related to their ability to perform certain computations. These computations (binary logic operation) are the "traits" that avidians display. In a monoclonal population, all avidians express the same traits and thus have the same merit and thus receive on average 30 SIPs (i.e., they execute 30 genome instructions during that update). However, avidian lineages can evolve the ability to increase merit, and thus increase the number of instructions that can be executed per update.

Fitness in Avida is implicit and can only be estimated by executing the computer instructions within an avidian's genome. In this regard, Avida differs from classical evolutionary simulations where genotypes are assigned a fitness and their frequency increases or decreases proportionally to this assigned value. Here, faster replicators out-compete slower replicators and fitness emerges from this dynamic. There are two means by which an avidian lineage

can evolved increased replication speed. First, genotypes that require fewer instruction executions to undergo reproduction will reproduce faster than avidians that require a greater number of instruction executions. In environments where genome size can vary, this is often achieved by decreasing genome size. In fixed genome size environments, this trajectory of fitness gain still occurs, often through the fixation of additional instructions that copy genome information from parent to offspring. The second methods of fitness increase in Avida is through the evolution of complex adaptations: the ability to perform one- and two-input Boolean logic calculations (the traits discussed earlier). In the setup used here, nine of these functions are potentially rewarded with SIPs; the amount of SIPs awarded is proportional to the complexity of the evolved trait [104]. The performance of these calculations increases an avidian’s merit. In other words, the more calculations an avidian can perform, the more SIPs it receives per update and it can then execute a larger portion of its genome per unit time. This increases the avidian’s replication speed and leads to selection for the ability to perform these calculations. When Avida estimates a genotype’s fitness, it is estimated as the genotype’s merit divided by the number of instruction executions required for replication (otherwise known as its gestation time).

Avidian populations evolve these complex adaptations by fixing a sequence of mutations. This genetic variation is introduced into a population when an avidian copies one instruction inaccurately into its daughter’s genome. The rate at which these mutations occur is set by the experimenter and is usually set as a rate “per instruction copied” (a copy-mutation rate). When mutations occur during this copying process, multiple mutations can occur per replication event. Additionally, in experiments where genome size can evolve, insertion and deletion mutations occur upon division (not during the replication process) at a pre-specified rate. These indel mutations insert or delete one instruction. Finally, large-scale genome changes can occur, but these changes are caused by the specific replication algorithm encoded by an avidian’s genome; these mutations are inherent to the Avida genetic code (implicit mutations) and are not under the control of the experimenter.



## Experimental Design

Here, we repeated the experimental protocol of the original survival of the flattest study [189]. We differed from the original study in one aspect: the placement of offspring. Here, we used Avida’s current default setting, where a new offspring randomly replaces one of the nine neighbors of its parent (including possibly the parent). The original paper used mass-action reproduction, where new offspring could randomly replace any offspring in the population. We first generated a collection of genotypes during the *Initial Adaptation* step. We evolved 201 populations of 3,600 individuals for  $5 \times 10^4$  updates. All populations initially started with the default Avida ancestor with an one-hundred instruction genome. Point mutations occurred at a rate of 0.0075 mutations per instruction copied and insertion/deletion mutations occurred at a rate of 0.05 mutations per division each. At the end of these populations’ evolution, we extracted the most abundant genotype.

For the *Mutation Rate Adaptation* phase of the experiment, we duplicated each of the 201 genomes and used one as progenitor of an experiment at a fixed *low* genomic mutation rate of 0.5 mutations per genome per generation, and one at a fixed high genomic rate of 2.0 mutations genome per generation. To achieve this, we adjusted the per-site mutation rate of each genome so that the targeted genomic rate was achieved, and disallowed sequence length changes. In this manner, the *Initial Adaptation* created genetic variation used as input to the adaptation step at a fixed genomic rate. From this stage onward each genotype was limited to the tasks it could complete at the end of the *Initial Adaptation* phase, i.e., they were not allowed to evolve novel traits, though they could lose traits they already possessed. Each of the populations adapting to high or low mutation rate consisted of 3600 individuals and evolved for  $10^4$  generations. After evolution we again selected the most abundant genotype from each population and denoted genotypes from the low mutation environment “fit” genotypes, and genotypes from the high mutation environment “flat” genotypes.

Next, we performed competition experiments between each pair of fit and flat genotypes (each pair has the same ancestor). We only performed these competitions with genotype pairs

where the fit genotype was 1.5 times more fit than the flat genotype. For the competitions, we seeded populations of 3600 individuals with 1800 individuals of each genotype and ran the simulation for 200 generations. We repeated these competitions across a range of genomic mutation rates (0.5 to 3.0 mutations/genome/generation in increments of 0.5) and performed five replicates per mutation rate. We tracked the abundance of the fit genotypes over time to determine the outcome of the competition. We classified the flat genotype as the winner if it reached at least 95% of the population in three of five replicates at a given mutation rate.

## Data Analysis

We performed mutational analyses to determine the fitness landscapes of the fit and flat genotypes used in the competitions using Avida’s “analyze” mode. In analyze mode, avdians run through their life-cycle in isolation (as opposed to in a population) and characteristics such as their fitness can be estimated. We first estimated the base fitness of all avidian genotypes used in the experiment. Then, we took the genotypes used in the competition and generated all possible single mutants. For each genotype, there are  $25L$  single mutants, where  $L$  is the genome size and 25 is the number of possible alternative alleles at each position in the genome, as there are 26 instructions in the Avida instruction set. We then repeated this procedure for all double mutants. For triple- and quadruple-mutants, we randomly sampled  $10^8$  mutants as the number of all possible mutants made generating every mutant computationally prohibitive<sup>1</sup>.

After generating these mutants, we calculated their fitness and compared them to their ancestor genotype. Mutants with greater fitness were designated as beneficial and mutants with the same fitness were designated as neutral. Mutants that were viable but with decreased fitness were designated as deleterious. Mutations resulting in genomes that could not reproduce were designated as lethal. To estimate differences in epistasis between fit and flat genotypes, we fit the mutational analysis data to the following equation:

---

<sup>1</sup>The number of possible  $n$ -mutants of a length  $L$  genome encoded in an alphabet of size  $D$  is  $\binom{L}{n}(D-1)^n$ , which translates to over  $1.5 \times 10^{12}$  possible mutants for  $L = 100$ ,  $n = 4$ , and  $D = 26$ .

$$\log\left(\frac{w(n)}{w_0}\right) = -\alpha n^\beta \quad (1.1)$$

Here,  $n$  is the genetic distance from the initial genotype,  $w(n)$  is the mean fitness of all mutants at a given genetic distance, and  $w_0$  is the fitness of the initial genotype. This is a similar equation to that fit in the original survival of the flattest paper [189], with the mean fitness estimated at a given mutation rate substituted by the mean fitness of all  $n$ -mutants (as is often used [103]).  $\alpha$  and  $\beta$  can be thought of as parameters that measure a genotype’s mutational robustness and epistasis, respectively. A greater  $\alpha$  indicates that a genotype is less mutationally-robust, while a greater  $\beta$  indicates epistasis is more negative (i.e., the fitness effect of a double-mutant is less than expected from the fitnesses of the two single-mutants).

All statistical analyses were performed using R 3.4.4 [137] and figures were generated with either the ggplot2 R package [183] or the Matplotlib Python package [69].

## Results

We first evolved 201 populations and isolated the most abundant genotype from each population. We then evolved these genotypes in both a high mutation rate environment and a low mutation rate environment and isolated the most abundant genotype from each environment. The genotypes from the low mutation rate environment did not significantly change in fitness compared to their ancestor (Fig. 1.1a; Wilcoxon Rank Sum test,  $p = 0.411$ ). The genotypes from the high mutation rate environment significantly decreased in fitness compared to their ancestors (Fig. 1.1a;  $p = 2.216 \times 10^{-7}$ ). By tracking the fitness of the high-mutation-rate populations through time, we observe that these genotypes first severely decrease in fitness upon transfer, but then recover fitness as they search out new, and presumably more mutationally robust, fitness peaks (Fig. 1.1b).

Of our 201 pairs of genotypes adapted to low mutation rates and high mutation rates, 166 pairs evolved such that the low mutation rate genotype (the fit genotype) had a fitness

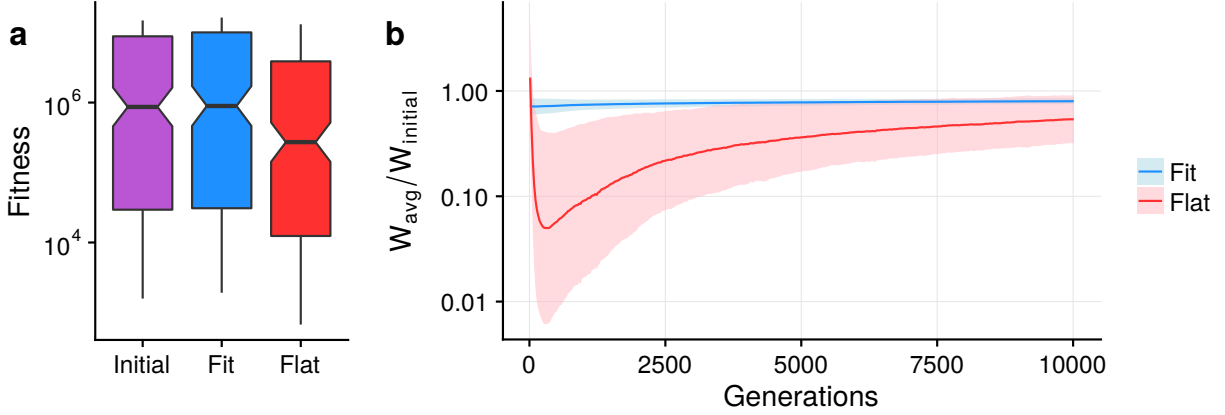


Figure 1.1: **Populations adapt to different fitness peaks in high mutation rate environments.** a) Fitness values for the most abundant genotype from each population after the *Initial Adaptation* experiment (denoted by “Initial”) and after adaptation to low (“Fit”) and high (“Flat”) mutation rate environments. For boxplots, lines are the median values, boxes show the interquartile range, and whiskers are 1.5 times the interquartile range or the minimum/maximum value; points are outliers. Overlapping notches suggest that the difference in medians is not significant. b) Relative average fitness values over the course of the *Mutation Rate Adaptation* experiment. Lines are the mean value across all populations and the shaded regions are standard deviations. Blue represents fit genotypes and red represents flat genotypes.

greater than 1.5 times the high mutation rate genotype (the flat genotype), which is the threshold used in the original study [189]. We took these pairs and performed competition experiments between the two genotypes across a range of mutation rates. In 76 (45.8%) of these competition experiments we noted that the eventual winner of the competition depended on the mutation rate, and that a switch between winners occurred at a critical mutation rate, just as was observed in the initial study [189]. Fit genotypes win the competition at low mutation rates, but flat genotypes win the competition at high mutation rates even with the large fitness difference (Fig. 1.2). It is worth noting that our definition of winning excluded many genotype pairs that showed the effect to a lesser degree.

Next, we analyzed the mutational neighborhood (the local fitness landscape), of all 166 pairs of fit and flat genotypes. We first generated all genotypes containing one point mutation and measured their fitness. There are significant differences in particular classes of mutations. Flat genotypes have a greater likelihood of beneficial (Fig. 1.3a; Wilcoxon Rank Sum test,

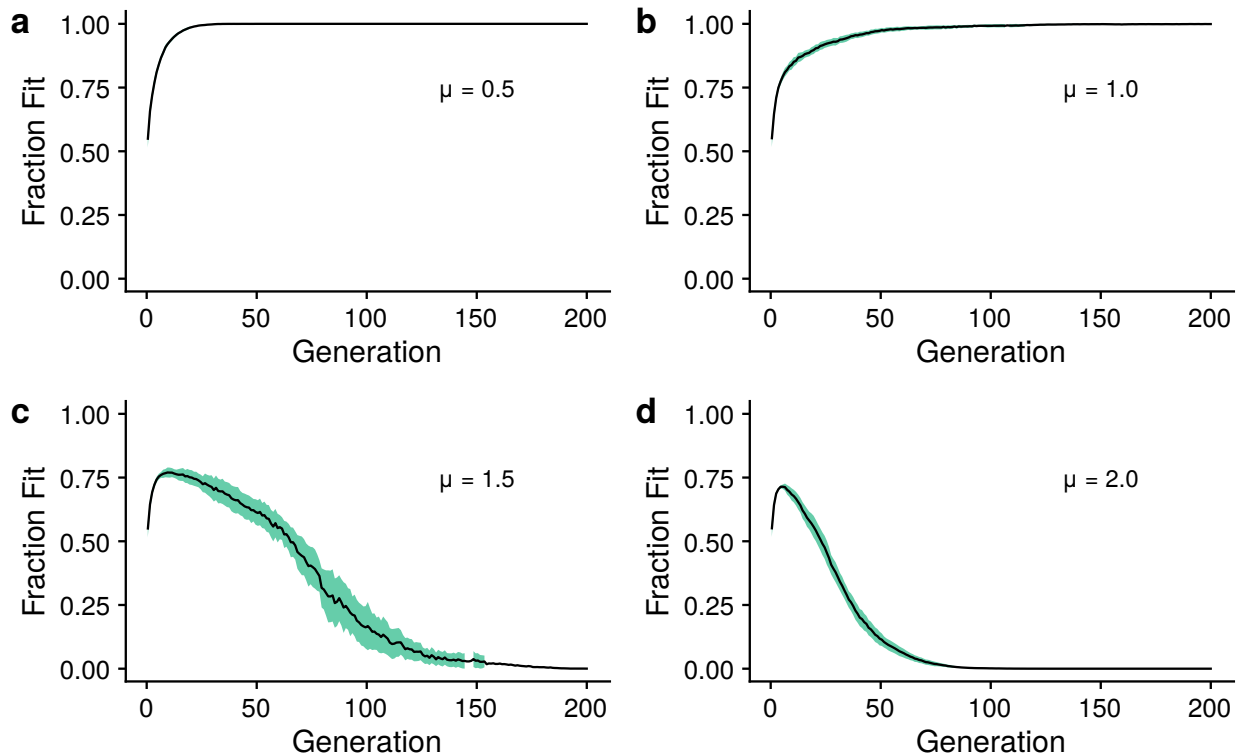


Figure 1.2: **Competition outcome between fit and flat genotypes is determined by the mutation rate.** Outcome from competition experiments between one fit/flat genotype pair. Black line is the mean fraction of the population consisting of the fit genotype as a function of time (measured in generations). Green shading represents one standard deviation. Each subplot shows the results for a given genomic mutation rate for five replicates.

$p = 1.35 \times 10^{-20}$ ), neutral (Fig. 1.3b;  $p = 2.66 \times 10^{-35}$ ), and lethal (Fig. 1.3d;  $p = 0.0168$ ) mutations. Fit organisms have a greater likelihood of deleterious mutations (Fig. 1.3c;  $p = 1.31 \times 10^{-44}$ ). Additionally, fit genotypes have lower mean relative fitness than flat genotypes (Fig 1.3d;  $p = 2.24 \times 10^{-27}$ ). We also analyzed the mutational neighborhood for sequences that are two, three, or four mutations away from the wild-type sequence and found similar results to the one-mutant mutational neighborhood (Fig. 1.4).

Finally, we used the mutational neighborhood data for sequences up to four mutations away from the wild-type sequence to estimate differences in epistasis between fit and flat genotypes. We fit the relative average fitness values as a function of mutational distance to a curve containing two parameters (see Methods for further details):  $\alpha$  (as a measure of mutational robustness) and  $\beta$  (as a measure of epistasis). As expected from the data on the

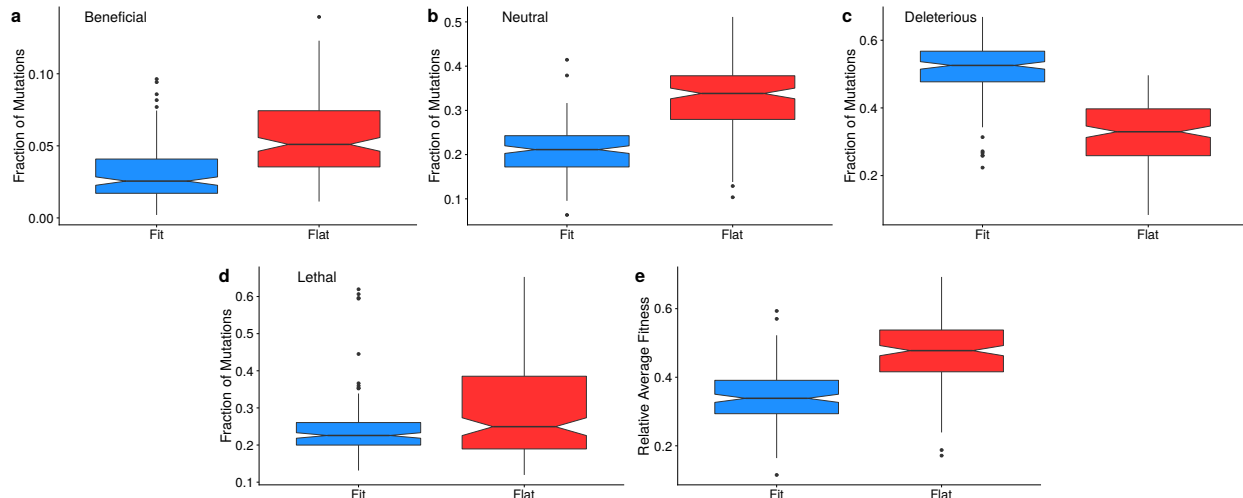


Figure 1.3: **Description of local fitness landscapes for fit and flat genotypes.** a-e) Distribution of fitness effects with mutations grouped by their broad effect. Boxplots as previously described. Lineage refers to whether the data is for fit genotypes (blue) or flat genotypes (red). f) Average relative fitness of all one-step mutants for each fit and flat genotype.

relative fitness of one-step mutants (Fig. 1.3e), fit genotypes have greater  $\alpha$  values (Fig. 1.5a;  $p = 1.06 \times 10^{-27}$ ). Flat genotypes have greater  $\beta$  values (Fig. 1.5b;  $p = 5.44 \times 10^{-34}$ ), indicating that flat genotypes have greater negative epistasis than fit genotypes. However, it should be noted that flat genotypes have, on average, almost no epistasis, while fit genotypes have positive epistasis. These data demonstrate how the selection pressure for mutational robustness drives populations to alternative areas of the fitness landscape.

## Discussion

We used digital experimental evolution to map the local fitness landscape of high fitness, but mutationally-fragile, “fit” genotypes adapted under low mutation rates, and low fitness, but mutationally-robust, “flat” genotypes adapted under high mutation rates. After repeating the original survival of the flattest effect [189], we calculated the distribution of fitness effects for *de-novo* point mutations up to four substitutions from the wild-type genotypes. We found that, as expected, flat genotypes were more mutationally robust than fit genotypes. Flat genotypes had a lower likelihood of deleterious mutations and a greater likelihood of

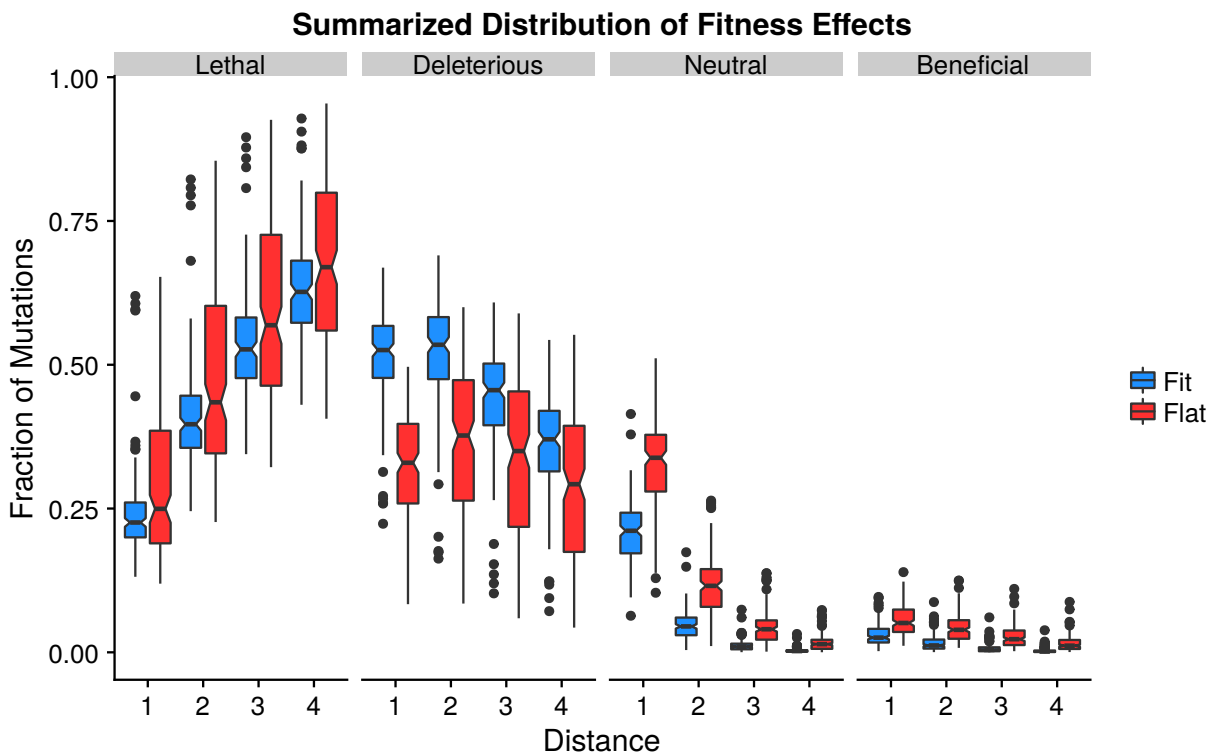


Figure 1.4: **Summarized distribution of fitness effects for mutants with multiple mutations.** Proportion of mutations with a mutational effect classification for mutants up to four mutations away from the wild-type sequence. See text for definitions of different classes of mutations. Distance is the number of mutations away from the wild-type sequence. Data for mutants with distance of one is the same as in Figure 3. Blue (red) boxplots are the proportions from the fit (flat) genotypes. Boxplots as previously described.

neutral and beneficial mutations than fit genotypes. Surprisingly, flat genotypes also had a greater likelihood of lethal mutations and more negative epistasis than fit genotypes, suggesting that the factors responsible for mutational robustness involve only limiting *heritable* deleterious variation.

Around the time when the original survival of the flattest paper first appeared, mutational robustness was largely conceptualized as a decrease in the likelihood of a deleterious mutation [173,189]. And while these changes in genetic architecture can lead to mutational robustness, work since that original publication has suggested that *increases* in the effect of deleterious mutations, not decreases, can also lead to robustness [4,126]. Increased severity of deleterious mutations can lead to population-level robustness as purifying selection

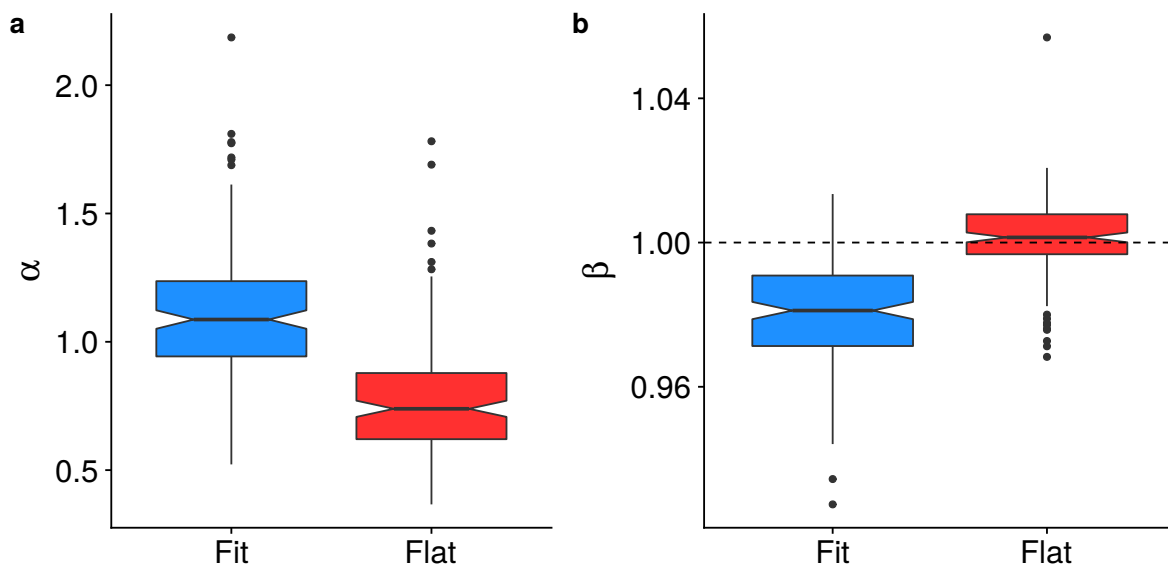


Figure 1.5: **Distribution of  $\alpha$  and  $\beta$  values for fit and flat genotypes.** Distributions of  $\alpha$  and  $\beta$  for each treatment.  $\alpha$  is a measure of single-mutant fitness, while  $\beta$  is a measure of epistasis (see Methods for further details). Boxplots as previously described.

is more effective against strongly-deleterious mutations. It has also been proposed that increased negative (or synergistic) epistasis can evolve as a mechanism to increase the strength of purifying selection and hence robustness [99, 190].

Our results suggest that all of these genetic mechanisms (increased neutrality, increased severity of deleterious mutations, and increased negative epistasis) can contribute to mutational robustness. These data support a previously-proposed relationship between increased neutrality (low  $\alpha$ ) and increased negative epistasis (high  $\beta$ ), based on data from Avida, RNA-folding, and neutral network fitness landscapes [187]. In other words, a population can only maximize immediate mutational robustness at the cost of increased deleterious severity of mutants farther away in the fitness landscape. We should also note that mutations on a “flat” genetic background are not, on average, negatively epistatic. Flat genotypes evolve to have, on average, no epistasis between mutations, as previously shown to occur in high mutation rate Avida populations [46]. However, they are more negatively-epistatic than fit genotypes, which have positive epistasis between mutations. The lack of negative epistatic



interactions in flat genotypes is likely related to their abundance of lethal mutations. Additional mutations on a genetic background with a lethal mutations cannot, by definition, lead to a greater decline in fitness (negative epistasis).

Theoretical arguments suggest that increasing the severity of mutations not only protects against high mutation rates, but also can protect against fitness loss due to genetic drift [85], although this depends on the specific distribution of fitness effects [21]. Recently, LaBar and Adami proposed the concept of “drift robustness,” or the idea that small populations tend to populate drift-robust fitness peaks with a low likelihood of slightly-deleterious mutations (and a high likelihood of both neutral and strongly-deleterious mutations), while large populations would adapt to drift-fragile fitness peaks with a high likelihood of slightly-deleterious mutations [98]. Both flat genotypes and drift-robust genotypes have some noticeable similarities in their distribution of fitness effects (increased neutrality, in particular), so it is worth asking whether the evolution of drift robustness and survival of the flattest is really the same phenomenon.

The main difference between drift robustness and survival of the flattest is the role of natural selection in the two phenomena. The evolution of drift robustness does not require competition between drift-robust and drift-fragile genotypes [98]. Instead, small populations tend to evolve towards drift robustness because they can only maintain fitness on drift-robust genetic backgrounds. Survival of the flattest, however, occurs when a flat genotype out-competes a fit genotype, and is driven by selection for mutational robustness. It has also been established that the critical high mutation rate above which this selective advantage occurs is independent of population size [35], while drift robustness, almost by definition, is strongly dependent on population size. Furthermore, our previous work demonstrated that drift-robust genotypes are not more mutationally-robust than drift-fragile genotypes, if one measures mutational robustness as the mean relative fitness of all single mutants [98]. Single mutants of flat genotypes are more fit, on average, than those of fit genotypes, again illustrating a difference between drift robustness and survival of the flattest. Thus, while

similarities exist between drift-robust genetic architecture and "flat" genetic architecture, it appears that they represent distinct evolutionary responses to different population-genetic environments.

It would be worthwhile to determine if the results observed held here can be replicated with RNA viruses, the biological model for the survival of the flattest effect [33, 101, 150]. It would be illuminating to test if the likelihood of lethal mutations does increase in these evolved viruses, especially in light of studies arguing that RNA viruses have high mutational *fragility*, not high mutational robustness [49]. Such a study would lead to a better understanding of the mutational mechanisms that lead to robust organisms in nature.

## Conclusions

In high mutation rate environments, populations will evolve alternative genetic architectures that increase their robustness to deleterious mutations. Classically, this is viewed as a survival of the flattest effect [189], where populations adapt to fitness peaks with increased neutrality [173]. Here, we repeated the original experiments to finely map the fitness landscapes of fit and flat genotypes. Flat, mutationally-robust genotypes adapt to fitness peaks with an increased likelihood of both neutral genotypes and lethal genotypes, a decreased likelihood of deleterious mutations, and greater negative epistasis compared to fit genotypes. These results demonstrate the genetic complexity of mutationally robustness genomes that minimize the impact of heritable deleterious genetic variation.

## Availability of data and material

All Avida configuration files and data analysis scripts to recreate these experiments, analyze the data, and recreate the figures are available at:

<https://github.com/joshf93/MappingThePeaks>

## Competing interests

The authors declare that they have no competing interests.

## **Funding**

This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## **Author’s contributions**

JF performed the experiments and analyzed the data. JF, TL, and CA designed the study and wrote the manuscript.

## **Acknowledgements**

T.L. acknowledges a Michigan State University Distinguished Fellowship, a BEACON fellowship, and the Russell B. DuVall award for support. This work was supported in part by Michigan State University through computational resources provided by the Institute for Cyber-Enabled Research.

## Supporting Information

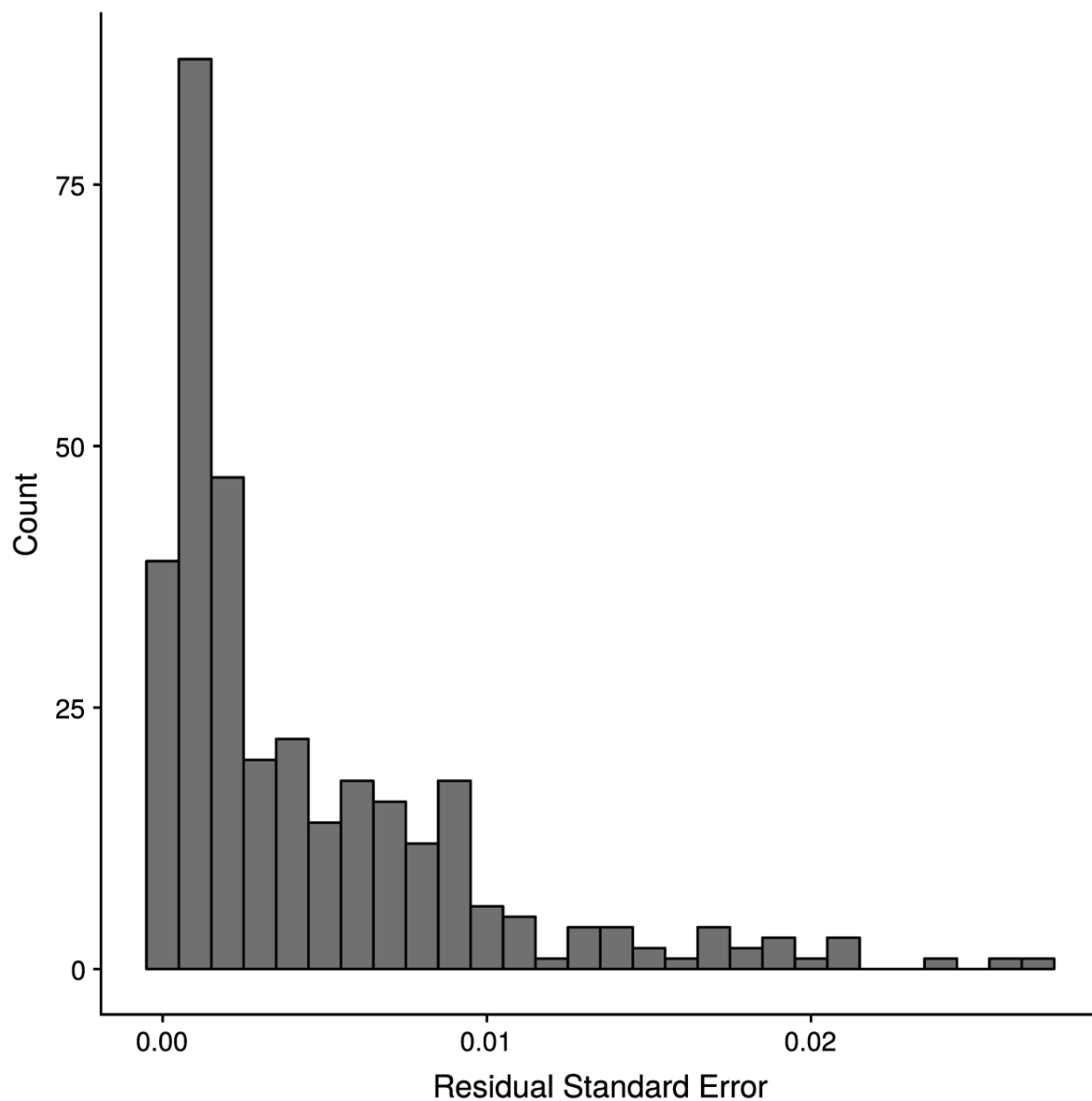


Figure 1.6: Histogram of the residual standard error (RSE) of the fits of Equation 1 to all fit and flat peaks.

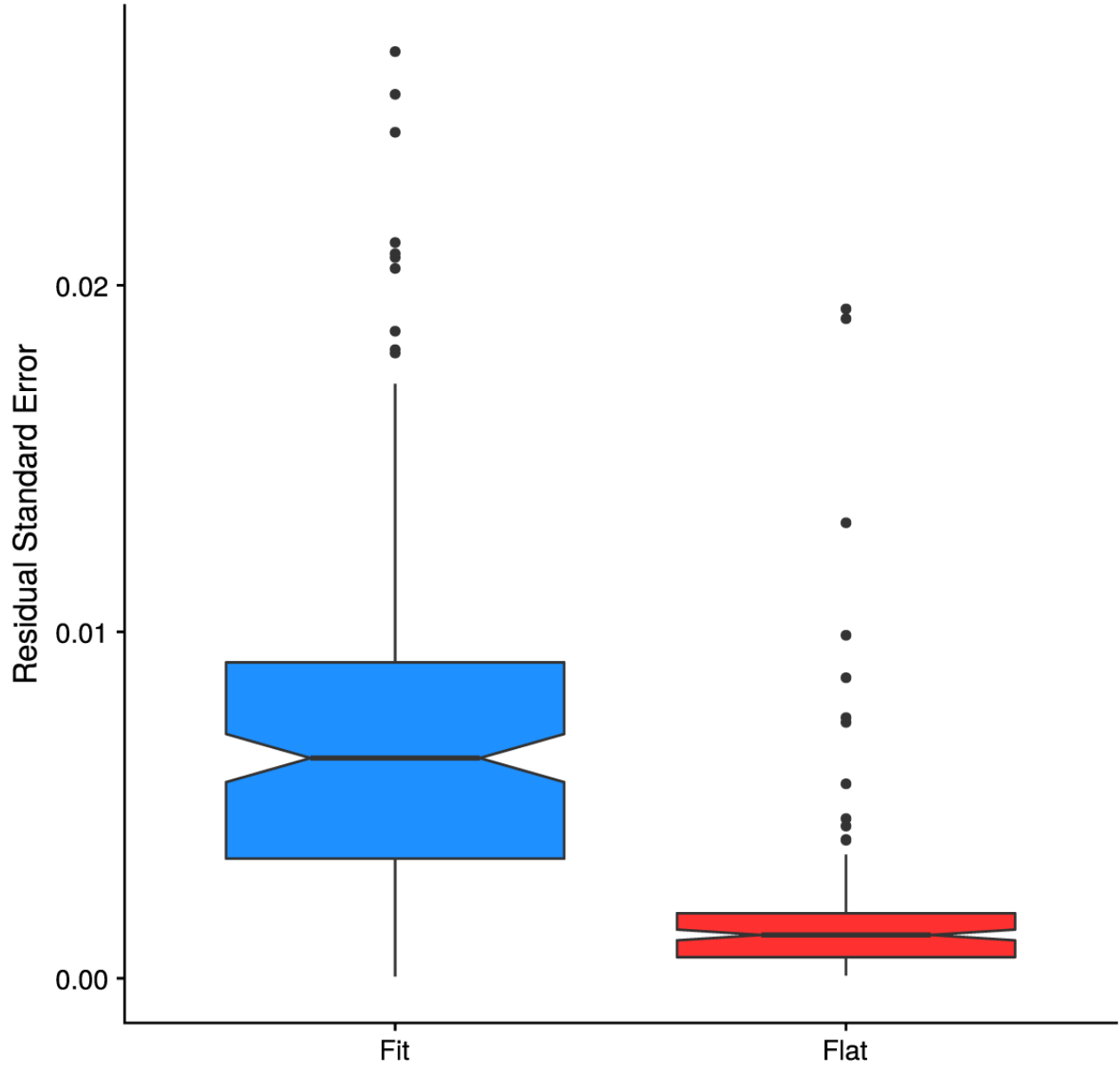


Figure 1.7: **Boxplot of the residuals of the fit to Equation 1 for the fit peaks (blue) and flat peaks (red) separately.** As in the boxplots of the main text, lines are the median values, boxes show the interquartile range, and whiskers are 1.5 times the interquartile range of the minimum or maximum value; points are outliers.

**CHAPTER 2 - CELL DENSITY, ALIGNMENT, AND ORIENTATION COR-  
RELATE WITH C-SIGNAL-DEPENDENT GENE EXPRESSION DURING  
*MYXOCOCCUS XANTHUS* DEVELOPMENT**

Authors: Y Hoang\*, Joshua L. Franklin\*, Yann S. Dufour, and Lee Kroos

\*\*Co-first authors

Notice: My contribution to this article focused on the data analysis, modeling, and figure construction, as well as some edits to the manuscript.

## Abstract

Starving *Myxococcus xanthus* bacteria use short-range C-signaling to coordinate their movements and construct multicellular mounds, which mature into fruiting bodies as rods differentiate into spherical spores. Differentiation requires efficient C-signaling to drive the expression of developmental genes, but how the arrangement of cells within nascent fruiting bodies (NFBs) affects C-signaling is not fully understood. Here, we used confocal microscopy and cell segmentation to visualize and quantify the arrangement, morphology, and gene expression of cells near the bottom of NFBs at much higher resolution than previously achieved. We discovered that “transitioning cells” (TCs), intermediate in morphology between rods and spores, comprised 10 to 15% of the total population. Spores appeared midway between the center and the edge of NFBs early in their development and near the center as maturation progressed. The developmental pattern as well as C-signal-dependent gene expression in TCs and spores were correlated with cell density, the alignment of neighboring rods, and the tangential orientation of rods early in the development of NFBs. These dynamic radial patterns support a model in which the arrangement of cells within the NFBs affects C-signaling efficiency to regulate precisely the expression of developmental genes and cellular differentiation in space and time. Developmental patterns in other bacterial biofilms may likewise rely on short-range signaling to communicate multiple aspects of cellular arrangement, analogous to juxtacrine and paracrine signaling during animal development

## Significance Statement

Significance Statement Pattern formation is foundational to development of multicellular organisms. To investigate patterning in bacteria during formation of *Myxococcus xanthus* fruiting bodies, cellular arrangement, morphology, and gene expression driven by C-signaling were quantified using high-resolution microscopy. How short-range C-signaling controls gene expression governing rod-to-spore differentiation within multicellular mounds has been unclear. The results support a new model in which cell density, rod neighbor alignment, and

rod tangential orientation affect C-signaling efficiency to pattern gene expression and spore maturation. Other bacterial biofilms may likewise be patterned by short-range signaling that captures local cellular arrangement and dynamical interactions, which could be exploited to manipulate biofilm development. Our results provide fundamental insights into spatiotemporal patterning by short-range signaling, which is relevant to animal development.

## Introduction

How cells build multicellular structures and adopt alternative fates are fundamental questions in developmental biology. The mechanisms of pattern formation can be lineage-dependent or signal-dependent, and formation of complex patterns of differentiated cell types typically involves both mechanisms [41,68]. Many bacteria form multicellular structures such as filaments, aggregates, and biofilms, in some cases with differentiated cells, although the patterns tend to be relatively simple (i.e., fewer cell types and less precise spatial organization than in many multicellular eukaryotes) [30,108]. Single-species bacterial multicellular structures have proven to be useful models to discover fundamental principles of pattern formation (e.g., *Anabaena* heterocyst differentiation [65], *Streptomyces* development [20], biofilm formation by *Bacillus subtilis* [179], *Vibrio cholerae* [169], and *Pseudomonas aeruginosa* [182]). Compared to other bacteria, *Myxococcus xanthus* biofilms form relatively complex patterns when starved [196]. Cells coordinate their movements to build regularly-spaced mounds of nearly uniform size and hemispherical shape, each containing approximately 100,000 cells. Some of the rod-shaped cells within mounds differentiate into round spores (Fig. 2.7). The mature spore-filled mounds are called “fruiting bodies”, which promote survival [175]. However, the majority of cells lyse during the developmental process and 10-20% of the cells remain outside of fruiting bodies as “peripheral rods” [102,125,191] (Fig. 2.7).

The dynamic nature of mound building by *M. xanthus* was first visualized by low-resolution time-lapse microscopy [76,142]. Increased resolution and tracking of individual rods have revealed changes in motility behavior that lead to mound formation. Early flocking of rods into groups is followed by alignment in streams facilitated by directional reversals



and slime-trail-following [6, 72, 170]. Intersecting streams appear to create “traffic jams” of rods with reduced motility [76, 156], which lead to raised mounds as monolayers of rods add vertically in tiers [40]. Some mounds mature into fruiting bodies, while others disperse, merge, or split in two [194]. Reduced motility in traffic jams is insufficient to explain all the observed dynamics of mound building [199]. Rods in addition perform a biased random walk toward mound centroids, suggestive of chemotaxis, and neighboring rods align their trajectories in a direction radial to the nearest mound [36].

Nascent fruiting bodies (NFBs) are stable mounds in the process of maturing into spore-filled fruiting bodies. Two NFB domains have been described, an outer hemispherical domain of densely packed rods surrounding an inner hemisphere of less ordered rods at three-fold lower cell density [145]. Rods in the outer domain were reported to rarely reverse their direction of movement and to orbit the inner domain in either direction. Differential gene expression in the two domains and temporal changes in refractile patches possibly due to spore precursors led to a model in which spore differentiation begins in the outer domain and fills inward over time, with the orbiting movement of rods in the outer domain transporting nonmotile spore precursors to the inner domain [146].

C-signaling is key among several factors known to coordinate mound building with spore differentiation, ensuring that spores form only within mounds [88, 167]. C-signaling involves a proteolytic fragment of CgsA [82, 106] and/or lipids produced by CsgA cardiolipin phospholipase activity [14]. A *csgA* null mutant fails to build stable mounds and fails to form spores [155]. Overexpression of *csgA* results in small mounds forming prematurely and spores forming prematurely outside of mounds [93]. Although the mechanism of C-signal transmission is unknown, cell movement enhances C-signaling, possibly due to end-to-end and/or side-to-side contact between cells [81, 84, 90, 147, 200]. Expression of *csgA* increases in response to starvation [38, 61, 105]. It has been proposed that alignment of rods in streams during mound building and in the outer domain of NFBs increase short-range C-signaling [72, 73, 145]. Both an increasing level of intracellular CsgA and increasing intercellular C-signaling may allow

thresholds to be reached that regulate differential gene expression required to coordinate mound building with spore differentiation [83, 105].

C-signaling regulates the expression of many genes during development [91]. In recipient cells, C-signaling has been proposed to activate the transcription factor FruA post-translationally, but the mechanism is unknown [50, 148]. FruA binds cooperatively with a second transcription factor, MrpC, to promoter regions of several C-signal-dependent genes [25, 116, 117, 143, 154, 159]. The cooperative binding sites exhibit different affinities and locations that may explain how increasing C-signaling could differentially regulate gene expression during development [89]. However, the relationships between C-signal-dependent gene expression, and the arrangement and morphology of individual cells, have not been visualized and quantified within NFBs.

We discovered that *M. xanthus* undergoes development in submerged culture on thin plastic slides amenable to high-resolution confocal microscopy. Using this approach in combination with fluorescent protein reporters, we found that individual cells could be visualized through several layers near the slide surface. We have quantified the spatiotemporal patterns of cellular arrangement, morphology, and C-signal-dependent gene expression in NFBs. We did not observe an inner domain of threefold lower cell density or evidence of spore precursor transport by the movement of rods. Rather, we observed rods aligned at high density near the radial center of early NFBs and our results support a new model in which cell density, and the neighbor alignment and tangential orientation of rods, affects C-signaling efficiency to pattern gene expression and spore maturation radially.

## Results

Nascent fruiting bodies contain cells intermediate in shape between rods and spores. To examine the temporal pattern of cellular shape near the bottom of NFBs, wild-type strain DK1622 was starved to initiate development under submerged culture conditions in wells of thin-bottom plastic slides. The lipophilic dye FM 4-64 was added at the beginning of starvation in order to stain membranes. In a control experiment, mounds formed at the

normal time, 18 h poststarvation (PS), and darkened by 42 h (as visualized by bright-field microscopy from the bottom), in the absence or presence of FM 4-64 (Fig. 2.8). The number of sonication-resistant spores at 42 h was greater on average in the presence of FM 4-64 than in its absence ( $p = 0.042$ ). Importantly, FM 4-64 did not interfere with development.

The FM 4-64 membrane stain revealed the shape of individual cells in images of optical sections obtained using confocal laser scanning microscopy. Figure 2.1A shows a series of images obtained near the bottom of the same NFB at the indicated times PS. At 18 h, cells were rod-shaped. Strikingly, by 24 h, some rods had thickened and shortened, especially near the center of the NFB. These cells were intermediate in shape between rods and round spores. Herein, we refer to these cells as “transitioning cells” (TCs), since they are transitioning in shape. Some TCs become spores, but we have not been able to determine the percentage that do so (see below). At 20 to 40  $\mu\text{m}$  from the center of the NFB, many rods appeared to be aligned and oriented tangentially (i.e., nearly perpendicular to a radius of the NFB drawn through the center of the rod) (Fig. 2.1B), perhaps corresponding to the outer domain described previously [146] (see below). At 30 h, TCs were still abundant near the center of the NFB and there appeared to be less rods and less alignment farther from the center. By 42 h, many round spores had formed. TCs were also observed within the NFB and rods were present outside the NFB (i.e., becoming peripheral rods if they remain outside). Figure 2.1C shows enlarged images of rods at 18 h, lanceolate-shaped TCs at 24 h, elliptical TCs at 30 h, and round spores at 42 h.

We tried to determine the fate of TCs by capturing images of an NFB stained with FM4-64 (as described above) every 15 min from 30 to 42 h PS. The images were assembled into a time-lapse movie. In most cases, the fate of individual TCs was unclear. Many TCs left the plane of focus, possibly due to lysis or movement (driven by shape change and/or movement of other TCs or rods). Importantly though, some TCs become round spores. Others changed shape but did not become round. To illustrate the different outcomes, a small section of Movie 1 was enlarged. Figure 2.1D shows frames from movie 2. Arrows point to three TCs

at 30 h. The TC indicated by the orange arrow became a spore. The TC indicated by the magenta arrow changed shape, transiently elongating and possibly fusing with another TC or vesicle, and remaining irregularly shaped. The TC indicated by the white arrow changed shape abruptly, as seen comparing 32 and 34 h PS in Figure 2.1D, but after 38 h we could not be certain what happened to this TC. Although we could not determine the fate of all TCs, we conclude that NFBs contain TCs from 24 h onward, and that some TCs become spores by 42 h. Figure 2.9 summarizes our observations of the temporal pattern of cellular shape near the bottom of NFBs.

Several morphological characteristics distinguish transitioning cells from rods and spores. The membrane stain FM 4-64 allowed visualization of cells in images of optical sections, but quantification of three-dimensional cellular morphology was not possible because cells in NFBs were not sufficiently well-separated. Thus, we mixed wild-type strain DK1622 with one of four strains (YH7, YH8, YH14, YH15) derived from DK1622 that produce a fluorescent protein (*tdTomato* or *mNeonGreen*) under control of a vanillate-inducible promoter ( $P_{van}$ ). The two strains were mixed at a ratio of five unlabeled DK1622 to one labeled derivative (based on cell density measurements) and the mixture was starved as described above except vanillate (not FM 4-64) was added at the beginning of starvation. In each case, mound formation by the mixture was delayed by 6 to 9 h compared to that of unmixed strain DK1622 and each unmixed derivative, in the presence or absence of vanillate. We do not understand why mound formation by the mixtures was delayed, but it was reproducible. For each of three biological replicates, a z-stack of optical sections (with z-resolution of 0.5  $\mu\text{m}$  and a step size of 0.2  $\mu\text{m}$ ) was collected at 27, 30, 36, and 48 h PS from near the bottom of one NFB (i.e., 0.25 to 0.5  $\mu\text{m}$  above the bottom of the well) to 5  $\mu\text{m}$  up (10% of the height of NFBs). For three additional biological replicates of the mixture of strains DK1622 and YH8, z-stacks were collected to 10  $\mu\text{m}$  up. Light scattering prevented high-resolution visualization of cellular morphology farther up. Excessive light exposure has been reported to inhibit sporulation [105]. We did not observe inhibition at the laser intensities we used,

but we limited z-stack collection to 5  $\mu\text{m}$  up for the other 12 NFBs examined (3 NFBs  $\times$  4 mixtures) in order to avoid excessive light exposure, especially since laser excitation at two wavelengths was used to visualize both cellular morphology and gene expression in some NFBs.

Figure 2.2A shows representative images of optical sections near the bottom of one NFB formed by the mixture of strains DK1622 and YH8. At 27 and 30 h, labeled YH8 cells were mostly rod-shaped, although there appeared to be a few TCs and/or spores (keeping in mind that an optical section may show only a portion of a cell). By 36 h, there appeared to be less rods and more TCs and/or spores. At 48 h, cells appeared to be mostly spores, although a few rods and apparent TCs were observed. Figure 2.2B shows enlarged images of rods and apparent TCs and spores. To quantify cellular morphology, we developed a computational pipeline involving three-dimensional segmentation of cells in image stacks and classification of individual cells. Four morphological properties (equivalent diameter, surface area, length, width) were used to classify cells as rods, TCs, or spores by fitting a Gaussian mixture model. Although the morphological transition between rods and spores is continuous, a mixture of three multivariate Gaussian distributions successfully captured the distribution of the four morphological properties of the whole population and quantitatively defined TCs. Scatter plots show clustering of cell classes based on each combination of two properties (Figure 2.10). Three-dimensional renderings of randomly sampled cells of each class were reconstructed from the image stacks and visually inspected to verify the classification method. Figure 2.2C shows renderings of representative cellular morphologies and Figure 2.11 shows the distributions of morphological properties of cells classified as rods, TCs, and spores. Rods were either straight or bent, and had the greatest median equivalent diameter (i.e., the diameter of a sphere with the same volume as the object) and length, and the least median width and sphericity (i.e., a measure of how closely the shape of an object resembles that of a sphere). Spores were nearly spherical, with the greatest median width and the least median length and equivalent diameter, whereas TCs had median morphological

properties intermediate between spores and rods.

Proportions of rods and spores change as nascent fruiting bodies mature. To determine the temporal patterns of rods, TCs, and spores in NFBs, we calculated the proportion of each cell class for each z-stack collected as described above. Figure 2.2D summarizes the temporal distribution for the 15 NFBs. At 27 and 30 h PS, 75% of the cells were rods and small percentages of TCs and spores were observed. By 36 h, the proportion of rods had decreased more than twofold, that of TCs had changed very little, and that of spores had increased considerably to >50%. At 48 h, the proportions of rods and TCs were low, and that of spores was 75%. The temporal changes in proportions were reproducible among NFBs and the most dramatic change occurred between 30 and 36 h, with the proportions of rods and spores decreasing and increasing, respectively, while the proportion of TCs was 10-15% at all times examined, indicating a steady flux of cells transitioning from rods to spores (see Supplementary Discussion for more about TCs).

Radial patterns of cell density, alignment, and orientation early correlate with spore proportion later as nascent fruiting bodies mature. In terms of the spatial pattern of each cell class for the 15 NFBs, we focused on the radial distributions, since all but three of the z-stacks went to only 5  $\mu\text{m}$  up from near the bottom, representing only 10% of the height of the NFBs, and no vertical patterns were discernible. Figure 2.3A summarizes the radial distribution of normalized cell density and the proportions of rods, TCs, and spores for the 15 NFBs. At 27 h PS, cell density was greatest at the radial center of NFBs and decreased steadily to the edge of NFBs at 60  $\mu\text{m}$ . Later, NFBs became more compact radially, with most cells within 40  $\mu\text{m}$  from the center. By 48 h, the maximum cell density had shifted to 15  $\mu\text{m}$  from the center. Previously, an inner domain of three-fold lower cell density was described for NFBs formed on starvation agar at 24 h PS [146]. Under our submerged culture conditions, we observed only a slight decrease in cell density near the center of NFBs at 48 h (Fig. 2.3A). The radial distribution of normalized cell density was very similar for each of the 15 NFBs (Fig. 2.12 shows a representative NFB formed by strain DK1622 mixed

with each of the four labeled DK1622-derived strains). The proportions of rods, TCs, and spores varied over time along the radii of NFBs, with very little variation between individual NFBs (Fig. 2.12). At 27 and 30 h PS, the proportion of rods is slightly lower at the radial center of NFBs than at the edge ( $p = 0.049$  and  $p = 0.053$ , respectively) (Fig. 2.3A). The difference increases dramatically by 36 h ( $p < 0.001$ ), and persists at 48 h, when rods are nearly absent within 20  $\mu\text{m}$  from the center. The proportion of TCs was maximal 17  $\mu\text{m}$  from the center at 30 h ( $p = 0.003$  comparing the maximum with the edge), and increased closer to the center at later times. The proportion of spores was maximal at 24  $\mu\text{m}$  from the center at 36 h ( $p = 0.003$  and  $p < 0.001$  in comparison with the center and the edge, respectively), and increased both closer to the center and closer to the edge by 48 h ( $p < 0.001$  comparing the center with the edge).

It is interesting that the proportions of TCs and spores were maximal at 20  $\mu\text{m}$  from the center and increased closer to the center as NFBs mature. Previously, it was proposed that spore differentiation begins in the outer domain of NFBs and fills inward over time, with the orbiting movement of rods in the outer domain transporting nonmotile spore precursors to an inner domain of less ordered rods at three-fold lower cell density [145]. However, as noted above, NFBs formed under submerged culture conditions do not have an inner domain of three-fold lower cell density (Fig. 2.3A). Moreover, our time-lapse movie of membrane-stained cells near the bottom of an NFB shows no strong evidence of orbiting movements of rods transporting TCs or spores toward the center. Therefore, it is unlikely that the transport model [145] accounts for the radial distributions of the proportions of TCs and spores (Fig. 2.3A). C-signaling requires cells to move into alignment [81, 84, 90, 147] and C-signal-dependent gene expression is necessary for sporulation [83, 89, 91, 93, 105, 167]. We noted above that many rods appeared to be aligned with each other and oriented tangentially to the circumference of the NFBs in a region possibly corresponding to the outer domain described previously [146], in NFBs at 24 h PS (Fig. 2.1A and 2.1B). Since mound formation by the mixtures of unlabeled strain DK1622 with labeled derivatives was delayed by 6 to 9

h (Fig. 2.2A), we expected the neighbor alignment and tangential orientation of rods to be greatest at 30 to 33 h, and at 20 to 40  $\mu\text{m}$  from the radial center of NFBs. Surprisingly, we discovered that rods are most aligned with each other near the center at 27 and 30 h (Fig. 2.3B). In agreement with our expectation, the tangential orientation of rods was greatest at 20 to 40  $\mu\text{m}$  from the center at 27 and 30 h.

To determine if the spatial arrangement of rods early in the development of NFBs is predictive of spore maturation, we performed a cross-correlation analysis of the posterior probabilities summarizing the radial patterns across NFBs and time points (Fig. 2.3A and 2.3B). Since the proportions of rods and spores change dramatically between 30 and 36 h (Fig. 2.2D), we calculated the cross-correlation coefficients of cell density, neighbor alignment, and rod tangential orientation at 27, 30, and 36 h with spore proportions at 36 and 48 h (Fig. 2.3C to 2.3E). Strikingly, the best predictor of spore proportion at 36 h is the rod tangential orientation at 30 h, followed by rod neighbor alignment. Cell density is a weaker predictor for early spore formation. However, rod tangential orientation becomes a weak predictor for spore formation at 48 h, whereas rod neighbor alignment and cell density become strong predictors across all time points. These results support a new model in which the initial increase in the proportion of spores 20 to 40  $\mu\text{m}$  from the center of the NFBs at 36 h can be explained by more efficient C-signaling between tangentially oriented rods relative to the center ( $<20 \mu\text{m}$ ) where cells are at higher density and aligned to their neighbors. The spatial arrangement of cells at the center may not be as conducive to C-signaling, resulting in a lag during spore development. Therefore, our model predicts that genes differentially sensitive to C-signaling will exhibit distinct radial patterns of expression. To test the model, we devised methods to quantify gene expression of individual cells, as described below.

C-signal-dependent gene expression increases in transitioning cells and spores as nascent fruiting bodies mature. To measure the spatiotemporal pattern of C-signal-dependent gene expression, two of the strains (YH14, YH15) mixed with strain DK1622 in experiments described above contained a C-signal-dependent promoter fused to *tdTomato*, in addition



to  $P_{van}$  fused to  $P_{fmgE}$ . For NFBs formed by those mixtures, two z-stacks measuring red and green fluorescence intensity were collected in the experiments described above. The green fluorescence from  $P_{van}$  allowed visualization of all labeled cells, whether or not red fluorescence was observed from the C-signal-dependent promoter. The green fluorescence also allowed cellular morphology measurements as described above and served as an internal control for normalization of red fluorescence as described below. Red fluorescence intensity from fusions of *tdTomato* to two different C-signal-dependent promoters,  $P_{dev}$  (strain YH14) and  $P_{fmgE}$  (strain YH15) was measured.  $P_{dev}$  was expected to be more sensitive than  $P_{fmgE}$  to a low level of C-signaling, based on in nitro and in vitro analyses of cooperative binding sites for FruA (proposed to be activated by C-signaling) and MrpC. The *tdTomato* fusions were integrated ectopically in the chromosome so as not to interfere with the native *dev* and *fmgE* loci, which contribute to the timing and level of sporulation, respectively [140, 159]. Figure 2.4A shows representative images of optical sections near the bottom of one NFB each formed by strain YH14 (designated *dev*) or YH15 (designated *fmgE*) mixed with strain DK1622. The images show merged red and green fluorescence. As expected, rods (red arrows), TCs (green arrows), and spores (blue arrows) in both NFBs exhibited green fluorescence from  $P_{van}$ -*mNeonGreen*. Strikingly though, rods of the *dev* strain (bearing  $P_{dev}$ -*tdTomato*) showed much greater red fluorescence than rods of the *fmgE* strain (bearing  $P_{fmgE}$ -*tdTomato*), which were barely visible. Figure 2.4B shows enlarged images of green and red fluorescence separately, as well as merged. Some TCs of the *fmgE* strain also showed very little red fluorescence at 30 h (e.g., Fig. 2.4B), but in general TCs and spores of both strains exhibited readily visible red fluorescence. The results suggest that  $P_{dev}$  is active at 27 h, earlier than  $P_{fmgE}$ , and that  $P_{fmgE}$  activity differs more than  $P_{dev}$  activity comparing rods with TCs and spores. These observations are consistent with the expectation that  $P_{dev}$  would be more sensitive to low-level C-signaling than  $P_{fmgE}$  [25, 159, 176, 177], since C-signaling appears to increase during development [83, 93, 105]. In the absence of C-signaling,  $P_{dev}$  and  $P_{fmgE}$  were inactive, since a *csgA* null mutant bearing  $P_{van}$ -*mNeonGreen* and either

$P_{dev}$ -*tdTomato* (YH46) or  $P_{fmgE}$ -*tdTomato* (YH45) failed to form NFBs and showed very little red fluorescence at 18 to 48 h, although both strains exhibited green fluorescence (Fig. 2.13).

To quantify C-signal-dependent gene expression of individual cells of the *dev* and *fmgE* strains mixed with strain DK1622, cells were classified as rods, TCs, or spores based on green fluorescence from  $P_{van}$ . The distributions of red and green fluorescence intensities were plotted (Fig. 2.14). At 36 and 48 h PS, some cells exhibited very low levels of both red and green fluorescence (lower left quadrants), suggesting these cells were compromised, so we eliminated them from further analyses. For the remaining cells, the distributions of red and green fluorescence intensities were plotted separately for rods, TCs, and spores (Fig. 2.14B and 2.14C). The laser and microscope settings were predetermined based on preliminary experiments (Supplementary Results). For each cell, the red fluorescence intensity was divided by the green fluorescence intensity to yield the relative fluorescence intensity (RFI) (Fig. 2.14D). Obviously, the distributions of RFI are broad for each cell class and both strains. Nevertheless, the median RFI of the *dev* strain exceeded that of the *fmgE* strain by 4-fold for all cell classes at 27 h and for rods and TCs at 30 h, consistent with the expectation that  $P_{dev}$  would be more sensitive than  $P_{fmgE}$  to a low level of C-signaling [25, 159, 176, 177]. Figure 2.4C shows the median (dot) and 90% credible interval (vertical line) of each RFI distribution (Fig. 2.14D). For both the *dev* and *fmgE* strains, the median RFI of rods changed little as NFBs matured, while the RFI of TCs and spores increased for *dev* ( $p = 0.046$  and  $0.025$ ) and *fmgE* ( $p = 0.0063$  and  $0.016$ ). The RFI of TCs and spores for each strain was then normalized to the RFI of rods at 27 h to quantify the increase in C-signal-dependent gene expression during development of NFBs (Fig. 2.4D). By 48 h, C-signal-dependent gene expression was 10 to 30-fold greater in both TCs ( $p = 0.046$  and  $0.0065$ ) and spores ( $p = 0.013$  and  $0.0088$ ) relative to early rods. This analysis supports that C-signal-dependent gene expression is cell class-specific. We conclude that C-signaling increases gene expression in TCs and spores as NFBs mature.

Distinct radial patterns of C-signal-dependent gene expression correlate differently with cell density, rod neighbor alignment, and rod tangential orientation early in development of nascent fruiting bodies. Our new model to explain the radial distribution of the spore proportion in maturing NFBs (Fig. 2.3A) predicts that genes differentially sensitive to C-signaling will exhibit distinct radial patterns of expression. As expected,  $P_{dev}$  appeared to be more sensitive than  $P_{fmgE}$  to a low level of C-signaling (Fig. 2.4) [25,159,176,177]. Therefore, our model predicts that  $P_{dev}$  activity may be elevated throughout the NFBs regardless of cell arrangement and C-signaling efficiency, whereas  $P_{fmgE}$  activity may be more sensitive to efficient C-signaling between tangentially oriented rods at 20 to 40  $\mu\text{m}$  from the center during maturation of the NFBs.

To determine if cell arrangement affects C-signaling within the NFBs, we examined the radial pattern of gene expression in TCs and spores at each time point. At 27 h, the RFIs were uniform radially for both the *dev* and *fmgE* strains. By 30 h, the mean RFIs for TCs and spores of the *dev* strain had increased slightly near the center, while the mean RFI for spores of the *fmgE* strain was maximal at 20 to 40  $\mu\text{m}$  from the center ( $p = 0.032$ ). At 36 h, the mean RFIs for TCs and spores of the *dev* strain were uniform from the center to 30  $\mu\text{m}$ , then decreased 10-fold toward the edge ( $p = 0.028$  and  $0.004$ ), and the pattern persisted until 48 h ( $p = 0.076$  and  $0.016$ ). This pattern is similar to the radial pattern of the spore proportion at 48 h (Fig. 2.3A). Remarkably, the mean RFIs for TCs and spores of the *fmgE* strain at 36 h were maximal at 20 to 40  $\mu\text{m}$  from the center ( $p = 0.052$  and  $0.012$ ) (Fig. 2.5A), similar to the radial pattern of the spore proportion at 36 h (Fig. 2.3A). By 48 h, the mean RFIs for TCs and spores of the *fmgE* strain had increased near the center (Fig. 2.5A), resulting in a pattern similar to that of the *dev* strain at 36 and 48 h (Fig. 2.5A) and the spore proportion at 48 h (Fig. 2.3A). We conclude that genes differentially sensitive to C-signaling exhibit distinct radial patterns of expression, as predicted by our model.  $P_{dev}$  activity increases at  $<20$   $\mu\text{m}$  from the radial center in TCs and spores as NFBs mature, while  $P_{fmgE}$  activity increases maximally at 20 to 40  $\mu\text{m}$  from the center at 36 h

before increasing closer to the center.

To determine if the spatial arrangement of rods early in the development of NFBs is predictive of C-signal-dependent gene expression in TCs and spores, we performed a cross-correlation analysis of the radial patterns of rod arrangement at 27, 30, and 36 h (Fig. 2.3B) with the radial patterns of gene expression at 30, 36, and 48 h (Fig. 2.5A). The early patterns of cell density and rod neighbor alignment correlated strongly with the later patterns of RFI for TCs and spores of the *dev* strain (Fig. 2.5B, left and center), similar to the correlations with spore proportion (Fig. 2.3C and 2.3D). In contrast, rod tangential orientation early correlated poorly with RFI for TCs and spores of the *dev* strain later (Fig. 2.5B, right). Rod alignment at high cell density without tangential orientation appears to be sufficient to explain the radial pattern of  $P_{dev}$  activity. On the other hand, rod tangential orientation at 30 h correlated moderately well with RFI for TCs and spores of the *fmgE* strain at 36 h (Fig. 2.5C, right), similar to the correlations between rod tangential orientation and spore proportion (Fig. 2.3E). Interestingly, cell density and rod neighbor alignment early correlated strongly with RFI for TCs of the *fmgE* strain later (Fig. 2.5C, left and center). Rod alignment at high cell density without tangential orientation appears to largely explain the radial pattern of  $P_{fmgE}$  activity in TCs as NFBs mature (although tangential orientation may contribute to the pattern of  $P_{fmgE}$  activity in TCs at 36 h), but tangential orientation appears to be necessary to explain the pattern of  $P_{fmgE}$  activity in spores at 36 h. We conclude that the arrangement of cells early correlates differently with  $P_{dev}$  and  $P_{fmgE}$  activity in TCs and spores later, with greater sensitivity of  $P_{dev}$  than  $P_{fmgE}$  to a low level of C-signaling likely accounting for the distinct radial patterns of expression as NFBs mature (Fig. 2.5A). A requirement for abundant C-signaling appears to make  $P_{fmgE}$  activity a particularly good marker of spore maturation.

## Discussion

Our results provide new information about the spatiotemporal patterns of cellular differentiation and C-signal-dependent gene expression near the bottom of NFBs. In contrast to

a previous report that examined *M. xanthus* development on agar [146], we found that rods are aligned at high density near the radial center of early NFBs formed under submerged culture conditions (Fig. 2.3B). We also observed no strong evidence of TC or spore transport toward the center, as had been proposed previously [145]. Our results support a new model to explain the radial distribution of spores in maturing NFBs – efficient C-signaling between tangentially oriented rods leads to an initial increase of spores 20 to 40  $\mu\text{m}$  from the center at 36 h PS, and less efficient C-signaling between aligned rods at high density  $<20$   $\mu\text{m}$  from the center results in some spores by 36 h and more by 48 h (Fig. 2.3 and 2.5). The model is summarized in Figure 2.6. By communicating several features of cellular arrangement, short-range C-signaling produces precise patterning of gene expression and cellular differentiation in time and space. Our work has implications for understanding the formation of other bacterial biofilms and multicellular animals.

We did not observe an inner domain of three-fold lower cell density, as inferred previously based on fluorescence intensity measurements of confocal sections of NFBs formed by cells labeled with lipophilic dye [146]. A subsequent study also revealed an inner domain of lower fluorescence in confocal sections of NFBs formed by cells expressing *gfp* [107]. Individual cells were not enumerated in the previous studies. Other differences from our study include observation of NFBs from above (rather than below), which may affect light scattering and thus fluorescence intensity, and development on agar (rather than plastic), which may alter motility and thus cellular arrangement. However, fruiting bodies formed on agar and viewed in cross-section after high pressure freezing were uniformly spore-filled [67]. Cells did appear to be arranged differently within 20  $\mu\text{m}$  of the radial center of NFBs viewed from the bottom, than in the surrounding annulus (Fig. 2.1A, 2.1B, 2.2A, and 2.4A). We found that rods are aligned at high density near the center of early NFBs (Fig. 2.3B and 2.6). In this region,  $P_{dev}$  activity increased in TCs and spores as NFBs matured (Fig. 2.5A). A high-affinity site for cooperative binding of FruA (proposed to be activated by C-signaling) and MrpC stimulates transcription from  $P_{dev}$  [25, 177]. Hence, we infer that rods aligned at high density within

20  $\mu\text{m}$  of the center engage in an intermediate level of C-signaling sufficient for  $P_{dev}$  activity (Fig. 2.6). In comparison, we found that rods in the annulus 20 to 40  $\mu\text{m}$  from the center of early NFBs are not only aligned with neighbors but also tangentially oriented (Fig. 2.3B and 2.6). Cell density decreases sharply farther from the center across this region of early NFBs (Fig. 2.3A). Importantly, in this region activity of  $P_{fmgE}$  (as well as of  $P_{dev}$ ) increased by 36 h in TCs and spores (Fig. 2.5A). Two sites for cooperative binding of FruA and MrpC are located upstream of  $P_{fmgE}$ , a higher-affinity site that acts negatively by competing with a lower-affinity site that acts positively [25, 159, 176, 177]. Therefore, we infer that rods 20 to 40  $\mu\text{m}$  from the center of early NFBs engage in a high level of C-signaling sufficient for  $P_{fmgE}$  activity (Fig. 2.6). Rods  $>40$   $\mu\text{m}$  from the center of early NFBs appear to engage in a low level of C-signaling (Fig. 2.6) based on low  $P_{dev}$  and  $P_{fmgE}$  activity in TCs and spores as NFBs mature (Fig. 2.5A), presumably due to low cell density and nearly random rod alignment and orientation in that region of early NFBs (Fig. 2.3A and 2.3B).

Early work demonstrated the importance of cell motility for efficient C-signaling [81, 90]. Mechanical alignment of nonmotile cells facilitated C-signaling, but aligned nonmotile cells formed only 16% as many spores as motile cells [84]. A study of the role of C-signaling in rippling behavior [147] led to the notion that end-to-end contacts between rods in the outer domain of NFBs favor C-signaling [73, 145]. Modeling of rippling behavior instead predicted that side-to-side contacts between rods favor C-signaling [200]. Our results correlate several features of cellular arrangement in early NFBs with distinct radial patterns of C-signal-dependent gene expression as NFBs mature (Fig. 2.5). We propose that restricted movement and/or metabolism of rods aligned at high density within 20  $\mu\text{m}$  of the radial center of early NFBs leads to an intermediate level of C-signaling (Fig. 2.6). Lower cell density together with rod alignment and tangential orientation in the annulus 20 to 40  $\mu\text{m}$  from the center of early NFBs is proposed to allow rod movement, proximity (possibly contact), and/or metabolism that favors a high level of C-signaling (Fig. 2.6). Spore differentiation requires efficient C-signaling [83, 93, 105]. Expression of numerous developmental genes depends on

C-signaling and some of these genes affect sporulation [89, 91, 143]. Expression of the *dev* operon appears to allow a small protein, DevI [141], to delay sporulation for 6 h [140]. FmgE, which has a metalloprotease domain, enhances sporulation 8-fold [159]. We propose that differential expression of *dev*, *fmgE*, and likely many other C-signal-dependent genes, results from different levels of C-signaling due to the cellular arrangement in early NFBs, and produces the observed spatiotemporal pattern of spores as NFBs mature (Fig. 2.3).

Implications of patterning by C-signaling for understanding the formation of other bacterial biofilms and multicellular animals. By communicating several features of cellular arrangement, short-range C-signaling produces precise spatiotemporal patterning of gene expression and differentiation within NFBs. Earlier during development, C-signaling alters cellular movements so that stable mounds are built [40, 72, 201]. Exactly how C-signaling interfaces with motility control is unclear, but it involves the *frz* chemosensory system that regulates reversal of the direction of cell movement [158, 166]. The versatility of C-signaling for pattern formation by *M. xanthus* has implications for development of other bacterial biofilms, which are patterned by myriad short-range interactions [12, 32, 120, 136, 161]. Such interactions may communicate multiple aspects of cellular arrangement, judging from our results, which would be important to consider in designing therapeutic strategies that target the biogeography of polymicrobial infections [161].

Eukaryotes use a variety of mechanisms, broadly categorized as lineage-dependent or signal-dependent, to build multicellular structures with differentiated cells [41, 68]. Among the signal-dependent mechanisms, some depend on cell-to-cell contact (juxtacrine) and others work over a short distance (paracrine), so analogies to C-signaling abound, particularly in animal development. For example, similarities between juxtacrine Notch signaling and C-signaling have long been evident [75], but the postulated C-signal receptor has defied identification. If C-signaling instead acts over a short distance, analogies to paracrine signaling via growth factor [110, 172], Hedgehog [17], and Wnt [31] family proteins would be more apt. Notably, all the paracrine signaling pathways, as well as contact-dependent Notch signaling

pathways [15], play multiple roles in pattern formation during development of complex multicellular animals. Much remains to be learned about the molecular mechanisms of patterning by C-signaling and its evolutionary implications.

## Materials and Methods

**Bacterial strains, plasmids, and primers.** The construction of plasmids and bacterial strains is described in the SI Appendix. The strains, plasmids, and primers used in this study are listed in Table T2.1.

**Growth and development.** *M. xanthus* was grown at 32°C in CTTYE liquid medium (1% Casitone, 10 mM Tris-HCl [pH 8.0], 1 mM  $\text{KH}_2\text{PO}_4$ - $\text{K}_2\text{HPO}_4$ , 8 mM  $\text{MgSO}_4$ , [final pH 7.6], 0.2% yeast extract) with shaking at 350 rpm. Development was performed under submerged culture conditions [95] in 8-well  $\mu$ -slides (Ibidi) with starvation buffer MC7 (10 mM morpholinepropanesulfonic acid [MOPS, pH 7.0], 1 mM  $\text{CaCl}_2$ ) as described previously [139]. Cells from log-phase CTTYE cultures were collected by centrifugation and resuspended in MC7 to 1,000 Klett units. The cell suspension (26  $\mu\text{L}$ ) was added to 174  $\mu\text{L}$  of MC7 in each well. To collect samples for quantification of sonication-resistant spores, cells were scraped from the bottom of a well and the entire contents of a well was aspirated into a 1.5 mL microcentrifuge tube. MC7 (300  $\mu\text{L}$ ) was added, the sample was sonicated, and spores were counted microscopically as described previously [139].

**Microscopy.** Images of NFBs were acquired with a Nikon A1 Laser Scanning Confocal Microscope, which was configured on a Nikon Ti inverted platform with an XY automated stage and a 100X objective. Fluorescence from FM 4-64 and tdTomato were examined using a 560-nm laser for excitation and a 595/50 band pass emission filter. Fluorescence from mNeonGreen was examined using a 488 nm laser for excitation and a 525/50 band pass emission filter. Time-lapse confocal images were taken every 15 min. Images near the bottom of NFBs were the first optical section above the bottom of the well, in which cells could be clearly visualized, so 0.25 to 0.5  $\mu\text{m}$  above the bottom of the well.

**Cell segmentation.** Segmentation was done using a custom MATLAB (Mathworks Inc.)



script. The pixel intensity of each image in each image stack was normalized by subtracting the background intensity (mode of the intensity distribution) and dividing by the intensity at the 99.99th percentile. To smooth high-frequency noise, a three-dimensional Gaussian filter was applied ( $\sigma = 2$  pixels) to each image stack. To estimate the local curvature of fluorescent objects in three dimensions, we calculated the second-order partial derivatives of the voxel intensities with respect to each dimension (Hessian matrix). The direction and magnitude of the local curvature was determined by decomposing the Hessian matrix into eigenvectors and eigenvalues. The core of bright tubular and round structures (such as fluorescent rods and spores) are characterized by large negative curvatures in at least 2 directions. Therefore, new images were created by summing and inverting the two largest eigen values of each voxel to amplify the fluorescence signal likely to come from labeled cells. The resulting image was eroded using a 3x3x3 structure to further separate close objects prior to segmentation, binarized using a threshold at the 99th percentile to identify cells, and opened with a 2x2x2 structure to filter stray voxels. Small bright objects of less than 100 voxels ( $0.308 \mu\text{m}^3$ ) were filtered. To separate cells that were not properly segmented, objects that were larger than the median volume of the segmented objects were processed further. First, the skeleton of the object was extracted to calculate the number of endpoints (more than two indicates merged objects). Then branches with the highest angle from the trunk were separated into independent objects until the remaining objects had only two endpoints. After the image segmentation was completed, the morphological properties of each object were calculated (volume, centroid coordinates, equivalent diameter, surface area). The medial axis length and mean cell diameter along the cell axis were calculated from the cell skeleton transformation. The mean fluorescence intensity of each object was calculated from the original images after subtracting the background intensity for each stack. To filter cells that were truncated at the image boundaries, cells with centroids outside the z-heights range of 1 to  $4.5 \mu\text{m}$  (for  $5 \mu\text{m}$ -high z-stacks) or 1 to  $9.5 \mu\text{m}$  (for  $10 \mu\text{m}$ -high z-stacks) were discarded.

Cellular morphology classification. Cells were classified based on their morphologies by fitting a Gaussian mixture model (GMM) with four components using the `mclust` package [153]. The multivariate mixture was fitted once on the ensemble of segmented objects across NFBs and time points using as variates: equivalent diameter, square root of cell surface area, length of the cell medial axis, and mean cell width. All variates were log-transformed before fitting to produce distributions that were approximately Gaussian for each variable. The cells were assigned to one of four groups (rods, TCs, spores, debris) based on maximum likelihood. Cell classification was assessed by visual inspection of three-dimensional renderings of cells randomly sampled from each cluster across NFBs. The renderings were reconstructed from the image stacks using the Volume Viewer plugin of ImageJ [144]. Analysis of the Bayesian information criterion supports a mixture of 4 components for clustering the ensemble of fluorescent objects.

Comparisons of morphological parameters. The distributions of equivalent diameter, skeleton size (referred as length in Fig. 2.10 and 2.11), and mean width were modeled using an intercept-only log-normal model and the posterior probability distributions of the means were estimated with Bayesian sampling using `brms` [22], with a separate intercept for each cell class and the interaction between fruiting bodies and time points treated as a random effect:

$$Morph\_param \sim 0 + cell\_class + (cell\_class|fruiting\_body\_id : time\_point) \quad (2.1)$$

Sphericity was modeled using an intercept-only beta model with a separate intercept for each morphological classification. The significance of the differences between means was calculated empirically from the posterior probability distributions of each parameter. Bayesian sampling was done with 4 chains with 4,000 iterations each with default parameters. The calculated credible interval is the Bayesian analog of the confidence interval. It is

defined as the highest density interval within which a parameter value falls with a particular probability. There is 90% probability that the population parameter is within the 90% credible interval given the model and the sample data. For a broad overview of this kind of technique, see [92]. Plots were generated in the R statistical environment using the ggplot [183] and tidybayes [78] packages.

Estimating the proportions of cell classes and fluorescence intensity across time points. The number of cells in each class within the fruiting body was fitted with a multinomial logistic regression. Time was treated as an ordinal predictor and variability between fruiting bodies was treated as a random effect. The model was fitted using Bayesian sampling using the brms package [22] to determine the median and 90% credible intervals of Nb\_cellclass, the number of cells for each cell class at each time point:

$$\begin{aligned} Nb\_cellclass|trials(total\_cell\_nb) \sim \\ mo(time\_point) + (mo(time\_point)|fruiting\_body\_id) \end{aligned} \quad (2.2)$$

To model fluorescence intensity across time points and cell classes, the same approach was used using a log-normal model and adding promoter and cell classes as interacting factors in the regression

$$\begin{aligned} mvbind(ch1\_fluo, ch2\_fluo) \sim \\ promoter + cell\_class + mo(time\_point) : promoter : cell\_class + \\ (promoter + cell\_class + mo(time\_point) : promoter : cell\_class|fruiting\_body\_id) \end{aligned} \quad (2.3)$$

Bayesian sampling was done with 8 chains with 5,000 iterations each with default parameters. Plots were generated in the R statistical environment using the ggplot [183] and tidybayes [78] packages.

Estimating the proportions of cell classes and fluorescence intensity along the radii of NFBs. The proportion of each cell class as a function of distance within the fruiting body

was fitted with a multinomial logistic regression with a generalized additive model (GAM) of tensor products to smooth the spatial dependency using the `mgcv` package [192]. Variability between fruiting bodies was treated as a random effect and each time point was fitted independently. The model was fitted using Bayesian sampling using the `brms` package [22] to determine the mean and 90% credible intervals of  $P\_cellclass$ , the probability of a given cell class at a given location:

$$cell\_class|trials(1) \sim 1 + P\_cellclass \, t2(distance) + t2(distance|fruiting\_body\_id) \quad (2.4)$$

To model the cosine similarity of the orientations between cells or between cells and the tangent to the NFB circumference, and to model the mean relative fluorescence intensity of each cell as a function of space, we used the same approach but replaced the multinomial model with a beta or log-normal model. Bayesian sampling was done with default parameters except for ‘`adapt_delta`’=0.99 and ‘`inits`’=0.

Cross-correlation analyses of radial patterns within fruiting bodies. The cross correlations between cell alignment, orientation, or density with the proportion of spores or fluorescence intensities, were calculated by sampling the posterior probability distributions from the GAM models 1,000 times every 1  $\mu\text{m}$  from 0 to 60  $\mu\text{m}$  and calculating the Pearson correlation coefficient between sample pairs for the different combinations of parameters and time points. The mean correlations and the empirical p-values were calculated from the posterior probability distributions of the correlation coefficients.

## Acknowledgements

We dedicate this work to the memory of Dale Kaiser, whose insights and enthusiasm inspired countless scientists, including L.K. as one of his PhD students. Daniel Parrell initiated our confocal microscopy efforts and provided important suggestions and assistance during the early part of the work. We are grateful to Melinda Frame and the Center for

Advanced Microscopy at Michigan State University for assistance with microscopy. We thank Oleg Igoshin and Zhaoyang Zhang for suggestions regarding data analysis, Montse Elias-Arnanz for providing pMR3691, and Oleg Igoshin for helpful comments on the manuscript. This research was supported by National Science Foundation grants MCB-1411272 and IOS-1951025, and by salary support for L.K. from Michigan State University AgBioResearch.

## Supporting Information

### Results

Quantification of C-signal-dependent gene expression of individual cells in nascent fruiting bodies. Laser and microscope settings were predetermined based on preliminary experiments with the *dev* and *fmgE* strains mixed with strain DK1622. We avoided excessive light exposure so as not to interfere with development and not to saturate fluorescence from any cells at any times poststarvation (PS). We collected z-stacks using the same settings at different times PS so that fluorescence intensities can be compared directly for a given color and strain mixture. The magnitude of differences over time can also be compared for the *dev* and *fmgE* strains mixed with strain DK1622. For example, the difference in median (black bar) red fluorescence intensity of spores at 48 h compared with rods at 27 h is 10-fold greater for the *fmgE* strain than for the *dev* strain (Fig. 2.14). We note that green fluorescence intensity of rods, TCs, and spores decreased similarly at later times PS (Fig. 2.14). The decrease appears to be a characteristic of  $P_{van}$ , since red fluorescence intensity from  $P_{van}$ -*tdTomato* also decreased (Fig. 2.15; note that cells with low fluorescence intensity were not excluded in this experiment, which may explain the lower median green fluorescence intensity as compared with Fig. S8B).  $P_{van}$  is presumably recognized by the major sigma factor [70] and the level of this sigma factor has been shown to decrease during development [64]. Although transcription from  $P_{van}$  appears to decrease during development, green fluorescence intensity decreased consistently across cell classes and strains, so we reasoned that green fluorescence could serve as an internal control for normalization of red fluorescence from individual cells, allowing comparisons between cell classes and between the *dev* and *fmgE*

strains. Therefore, for each cell, the red fluorescence intensity was divided by the green fluorescence intensity to yield the relative fluorescence intensity (RFI) (Fig. 2.14).

## Discussion

Transitioning cells are intermediate in shape between rods and spores, but similar to spores in terms of C-signal-dependent gene expression. Although TCs in situ have not been described previously, shortened cells have been observed after resuspending cells from NFBs [11]. Multiple lines of evidence indicate that cell shortening involves conversion of fatty acids from membrane lipids into cytoplasmic lipid bodies [11,67] (3, 4). Interestingly, *csgA* mutant cells fail to shorten and exhibit only 20% of the wild-type lipid body area/cell [11], so CsgA cardiolipin phospholipase activity has been proposed to be important for lipid body formation [14]. Lipid bodies gradually disappear as spores form, suggesting that lipids provide metabolic fuel for cellular differentiation [67]. Since lipid bodies can be visualized by fluorescence microscopy after staining with Nile red [67], the relationships between TCs, cell shortening, and lipid body formation and utilization can be investigated within NFBs using the methods described herein. In addition to loss of 87% of the cell membrane as lipid bodies form [11], cellular shape change requires peptidoglycan remodeling. The peptidoglycan sacculus of *M. xanthus* is mostly if not completely degraded when rod-shaped vegetative cells are induced with glycerol to form spores [18]. This unicellular process differs from the multicellular process of starvation-induced fruiting body formation studied herein, yet in both cases a polysaccharide coat appears to rigidify spores [86,119] (7, 8) and the cytoskeletal protein MreB appears to recruit peptidoglycan remodeling enzymes during both formation and germination of spores [119,201] (8, 9). Nonuniform recruitment or activity of peptidoglycan degrading enzymes could account for the irregular shapes and different fates of TCs (Fig. 2.1, 2.2, 2.9). It is also possible that the peptidoglycan becomes flexible, so the shapes and fates of TCs may be influenced by other cells and the extracellular matrix in NFBs. Our efforts to track the fate of TCs demonstrated that some become spores, but the fate of most was unclear. Importantly, C-signal-dependent gene expression in TCs

is similar to that in spores (Fig. 2.4 and 2.5), suggesting that many TCs have entered the spore differentiation pathway.

## Methods

Construction of plasmids and bacterial strains. To construct pYH8, which was used to induce production of tdTomato by addition of vanillate, primers Van-tdTom F and Van-tdTom R were used to generate a PCR product of the *tdTomato* sequence, using plasmid tdTomato-N1 (Addgene, item ID 54642). The product was digested with NdeI and EcoRI and mixed with pMR3691 (also digested with the same restriction enzymes) in a Gibson assembly reaction to enzymatically join the overlapping DNA fragments [58]. The reaction mixture was transformed into *Escherichia coli* strain DH5 $\alpha$  with outgrowth in Luria-Bertani (LB) medium [149] prior to plating on LB agar (1.5%) supplemented with 15  $\mu$ g/mL tetracycline for selection at 37°C. The DNA sequence of the joined fragments was verified, and the plasmid was transformed into *M. xanthus* strain DK1622 using electroporation [77], with outgrowth in Casitone-Tris (CTT) (1% Casitone, 10 mM Tris-HCl (pH 8.0), 1 mM KH<sub>2</sub>PO<sub>4</sub>-K<sub>2</sub>HPO<sub>4</sub>, 8 mM MgSO<sub>4</sub>, (final pH 7.6)) medium prior to plating on CTT agar (1.5%) supplemented with 15  $\mu$ g/mL tetracycline to select transformants. Integration of pYH8 was verified by colony PCR using primers pMR3691 MCS G-F and pMR3691 MCS G-R. The resulting strain was named YH8. A similar method was used to generate pYH7, which bears P<sub>van-mNeonGreen</sub> (primers Van-mNeonGreen F and Van-mNeonGreen R were used to generate a PCR product with pNSI-ptre-mNeonGreen (Dr. Daniel Ducat, personal communication) as the template). pYH7 was transformed into strains DK1622 and LS2442 to create strains YH7 and YH48, respectively. To construct pYH14, a DNA fragment containing the *dev* promoter region was amplified from *M. xanthus* strain DK1622 genomic DNA with P<sub>dev</sub> F and P<sub>dev</sub> R primers, which contain EcoRI and BamHI sites, respectively. The *tdTomato* sequence was amplified from plasmid tdTomato- N1 (see above) using primers TomatoF and TomatoR, which contain BamHI and HindIII sites, respectively. The two fragments were joined to BamHI-EcoRI-digested pSWU19 in a ligation reaction. The reaction mixture was

transformed into *E. coli* strain DH5 $\alpha$  with outgrowth in LB medium prior to plating on LB agar supplemented with 50  $\mu\text{g}/\text{mL}$  kanamycin for selection at 37°C. The DNA sequence of the joined fragments was verified, and the plasmid was transformed into *M. xanthus* strains YH7 and YH48 using electroporation [77] , with outgrowth in CTT medium prior to plating on CTT agar supplemented with 40  $\mu\text{g}/\text{mL}$  kanamycin to select transformants. Integration of pYH14 was verified by colony PCR and the resulting strains were named YH14 and YH46. The *dev* promoter region fragment of pYH14 was replaced by an *fmgE* promoter region fragment to construct pYH15. The *fmgE* promoter region was amplified from genomic DNA with primers P<sub>*fmgE*</sub> F and P<sub>*fmgE*</sub> R. pYH15 was transformed into YH7 and YH48, resulting in strains YH15 and YH45.



| Strain            | Description   | Source     |
|-------------------|---|------------|
| <i>E. coli</i>    |   |            |
| DH5               | $\lambda^-$ $\Phi$ 80dlacZ $\Delta$ M15 $\Delta(lacZYA-argF)$ U169 <i>recA1 endA1 hsdR17</i> (r <sub>K</sub> -m <sub>K</sub> - <i>supE44 thi-1 gyrA relA1</i> | [63]       |
| <i>M. xanthus</i> |   |            |
| DK1622            | Laboratory wild-type strain   | [74]       |
| LS2442            | DK1622 with and in-frame deletion of <i>csgA</i>  | [14]       |
| YH7               | DK1622 with MXAN_0018-MXAN_0019::pYH7   | This study |
| YH8               | DK1622 with MXAN_0018-MXAN_0019::pYH8   | This study |
| YH14              | YH7 with attB::pYH14  | This study |
| YH15              | YH7 with attB::pYH15  | This study |
| YH45              | YH48 with attB::pYH15   | This study |
| YH46              | YH48 with attB::pYH14   | This study |
| YH48              | LS2442 with MXAN_0018-MXAN_0019::pYH7   | This study |

Table 2.1: **Strains used in this study.**

| Plasmid | Description   | Source     |
|---------|---|------------|
| pMR3691 | MXAN_0018-MXAN_0019-PR3-4:: <i>vanR</i> - <i>P<sub>van</sub>::lacZ</i> , Tet <sup>r</sup>                                 | [70]       |
| pSWU19  | Cloning vector with Mx8 <i>attP</i> locus for integration at <i>attB</i> in the <i>M. xanthus</i> genome, Km <sup>r</sup> | [193]      |
| pYH7    | MXAN_0018-MXAN_0019-PR3-4:: <i>vanR</i> - <i>P<sub>van</sub>::mNeonGreen</i> , Tetr                                       | This study |
| pYH8    | MXAN_0018-MXAN_0019-PR3-4:: <i>vanR</i> - <i>P<sub>van</sub>::tdTomato</i> , Tetr   | This study |
| pYH14   | pSWU19 with <i>P<sub>dev</sub>-tdTomato</i> , Km <sup>r</sup>   | This study |
| pYH15   | pSWU19 with <i>P<sub>fmgE</sub>-tdTomato</i> , Km <sup>r</sup>  | This study |

Table 2.2: **Plasmids used in this study.**

| <b>Primer</b>        | <b>Description</b>                                  | <b>Source</b> |
|----------------------|---|---------------|
| Van-tdTom-F          | GATGCGAGGAAACGCATATGGTGAGCAA<br>GGGCGAG             | This study    |
| Van-tdTom-R          | GTACGCGTAACGTTCTGAATTCTTACTTGT<br>ACAGCTCGTCCATG    | This study    |
| Van-<br>mNeonGreen-F | GATGCGAGGAAACGCATATGGTCAGCAA<br>AGGTGAAGAAGAC       | This study    |
| Van-<br>mNeonGreen-R | GTACGCGTAACGTTCTGAATTCCTTGTACA<br>GTTCGTCCATACCCATC | This study    |
| Pdev-F               | GCTAGAATTCCGCCGTTGGCTCGGATGC                        | This study    |
| Pdev-R               | GCTAGGATCCGACCTGAGCTGATACCAAC                       | This study    |
| PfmgE-F              | GCTAGAATTCCATTAACGGGCGATCTT<br>CCTC                 | This study    |
| PfmgE-R              | GCTAGGATCCCTAAAGGCGATTGCGTTC<br>CTGC                | This study    |
| Tomato-F             | GCTAGGATCCGAAAGGAGGCGCATATGG<br>TGAGCAAGGGCGAG      | This study    |
| Tomato-R             | GCTAAAGCTTTTACTTGTACAGCTCGTC<br>CATG                | This study    |
| pMR3691 MCS-<br>G-F  | CACGATGCGAGGAAACGCA                                 | This study    |
| pMR3691 MCS-<br>G-R  | CACCGGTACGCGTAACGTTC                                | This study    |

Table 2.3: **Primers used in this study.**

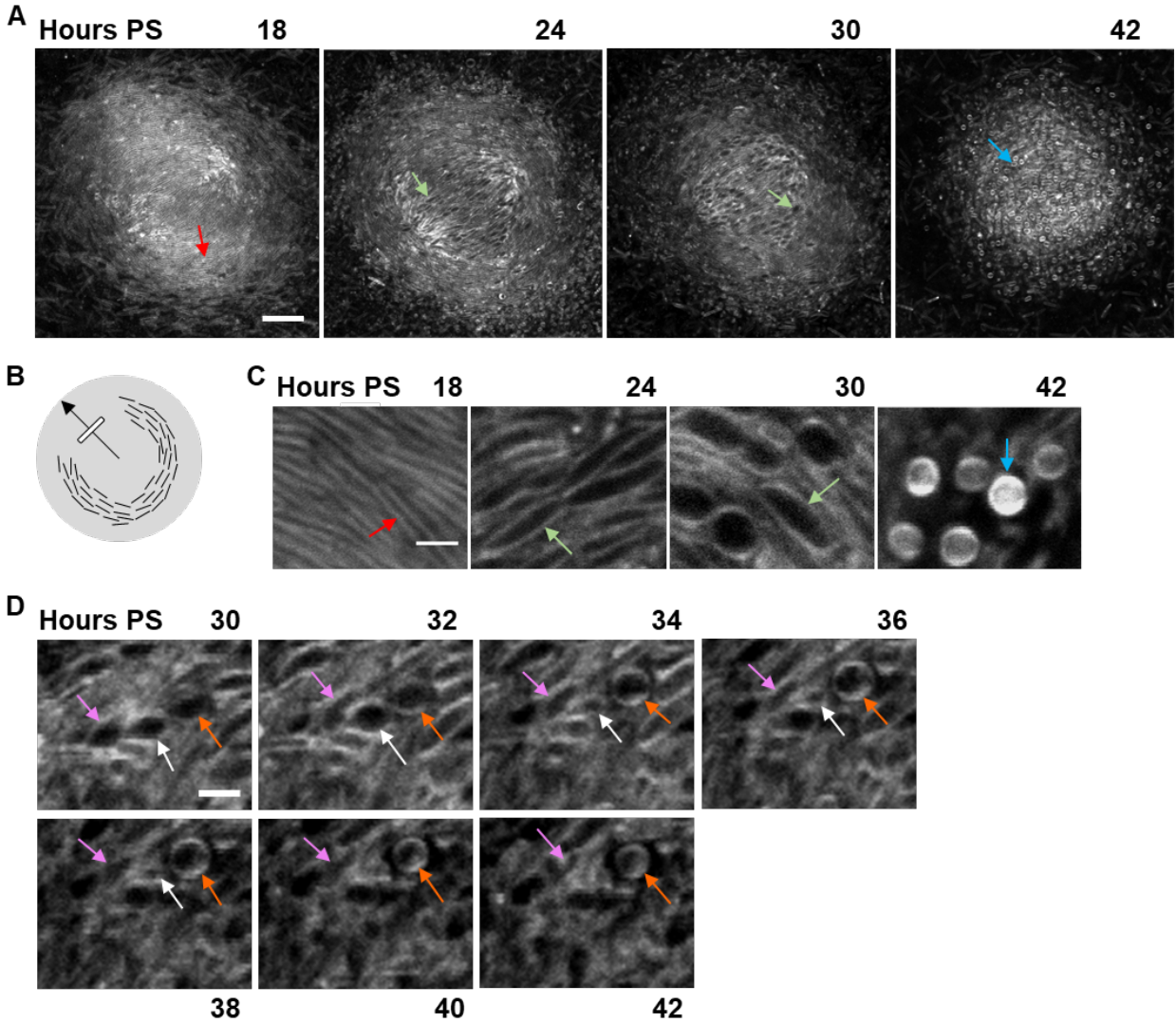
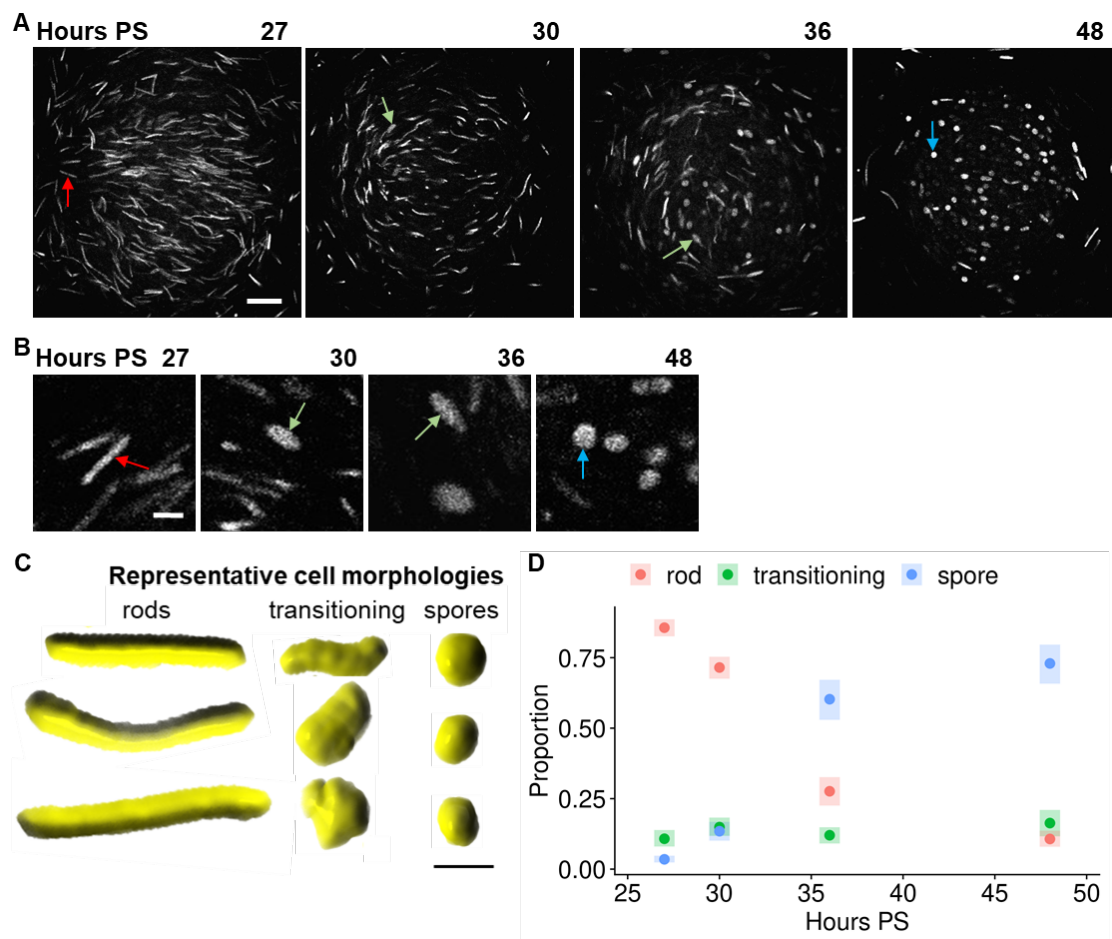
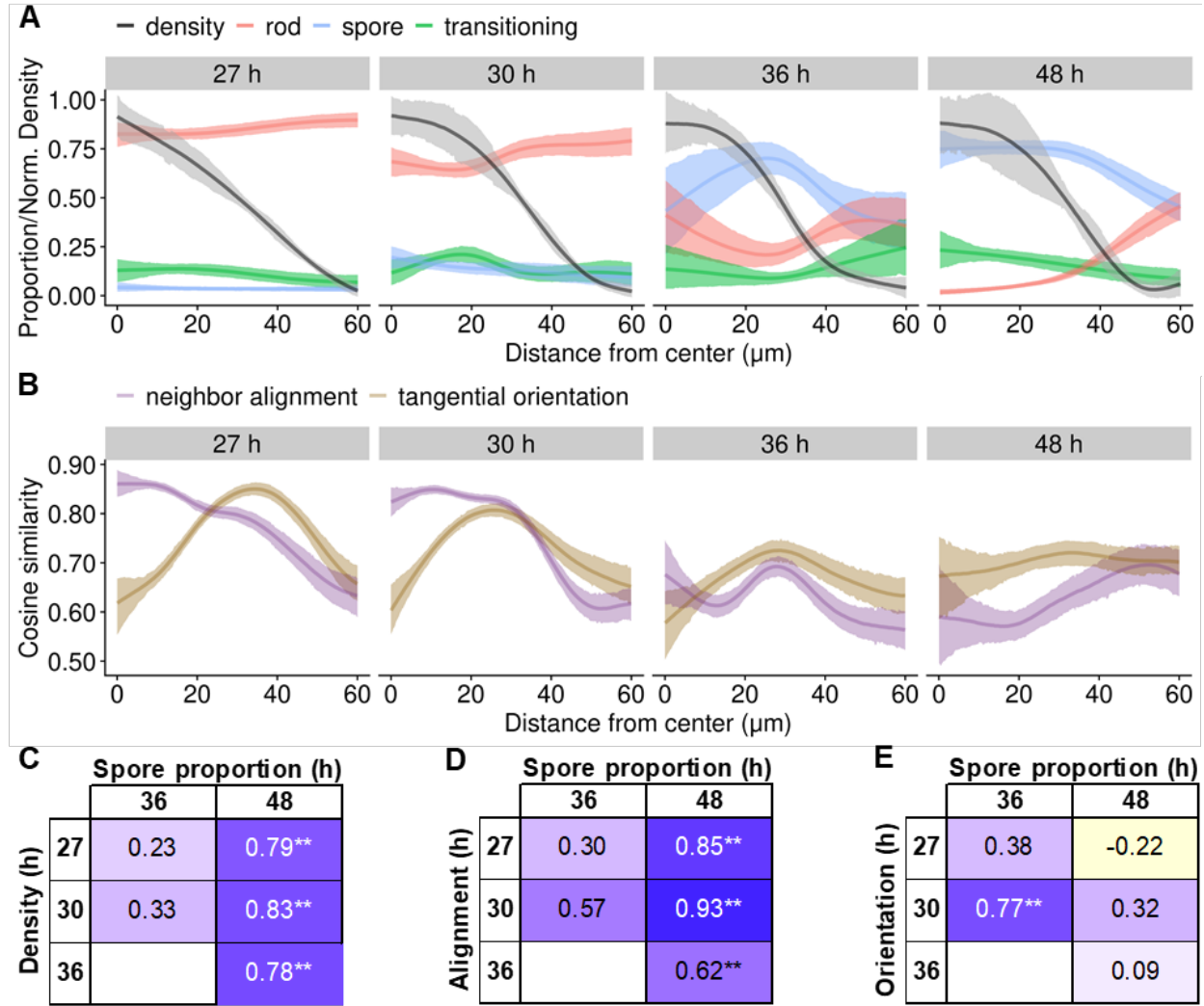


Figure 2.1: **Visualization of transitioning cells near the bottom of a nascent fruiting body.** *M. xanthus* wild-type strain DK1622 was starved to initiate development under submerged culture conditions. FM 4-64 (2  $\mu\text{g}/\text{mL}$ ) was added at the start of starvation. Confocal images were acquired at the indicated times poststarvation (PS) to show FM 4-64 staining of cell membranes (red fluorescence shown as grayscale). Images are representative of three biological replicates. (A) Image of an optical section near the bottom of the same nascent fruiting body (NFB) during development. Red arrow, rod-shaped cell. Green arrows, transitioning cells (TCs). Blue arrow, spore. Bar, 20  $\mu\text{m}$ . (B) Cartoon depiction of tangential orientation. Upper left quadrant depicts one tangentially oriented rod. Arrow indicates radius of NFB (gray). Other quadrants depict many smaller tangentially oriented rods. (C) Enlarged images of representative cells. Images from panel A were enlarged to show cells with a particular shape (a rod at 18 h, TCs at 24 and 30 h, and a spore at 42 h). Colored arrows point to the same cell classes as in panel A. Bar, 2  $\mu\text{m}$ . (D) Tracking the fate of TCs. Images from Movie S2. Orange arrows, TC that becomes a spore. Magenta arrows, TC that changes shape. White arrows, TC whose fate is uncertain. Bar, 2  $\mu\text{m}$ .

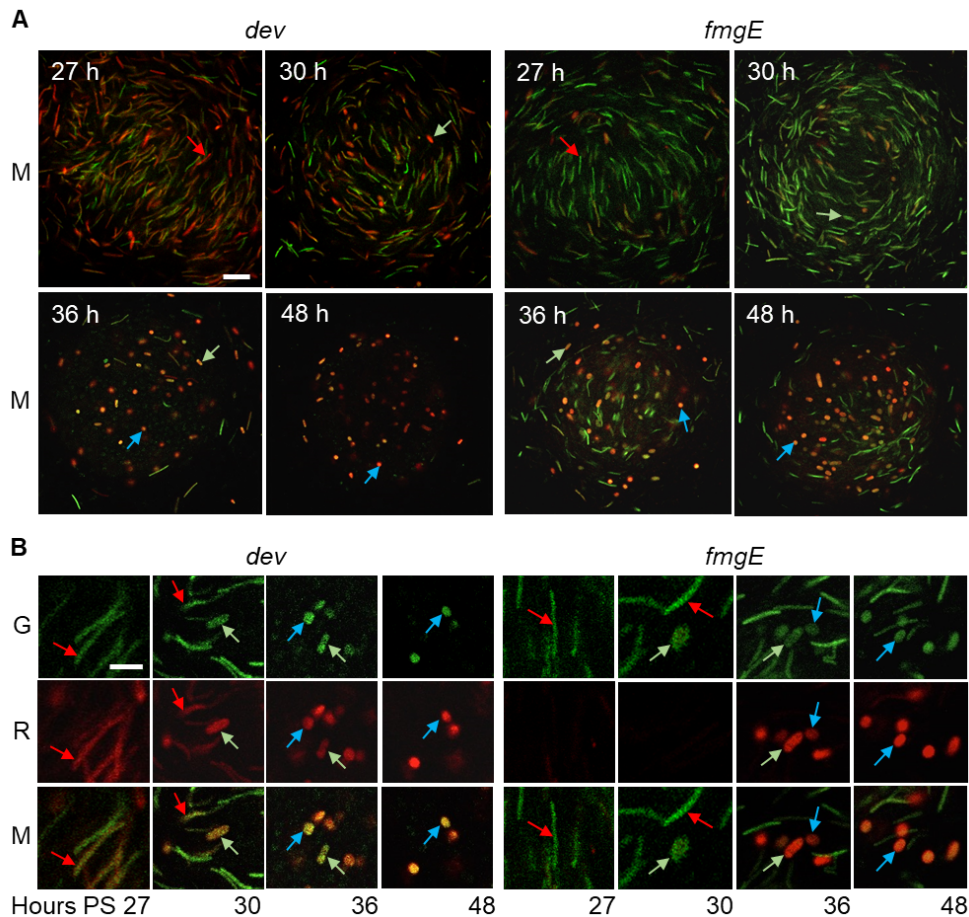


**Figure 2.2: Classification of cellular morphologies and the temporal patterns of rods, transitioning cells, and spores.** Labeled cells that produce a fluorescent protein under control of a vanillate-inducible promoter (strain YH7, YH8, YH14, or Y15) were mixed with unlabeled wild-type cells (strain DK1622) at a ratio of 1 to 5 and starved under submerged culture conditions. Vanillate (0.5 mM) was added at the start of starvation. Z-stacks of optical sections of the same nascent fruiting body (NFB) for each of three biological replicates were acquired at the indicated times poststarvation (PS) to show fluorescence from tdTomato or mNeonGreen. Segmented cells from all z-stacks were classified as rods, transitioning cells (TCs), or spores (5256, 1459, and 3213 cells, respectively). (A) Confocal images of the same NFB. Images show an optical section near the bottom of a representative NFB formed by the mixture of strains YH8 and DK1622. Red fluorescence from YH8 cells is shown as grayscale. Red arrow, rod. Green arrows, TCs. Blue arrow, spore. Bar, 20  $\mu$ m. (B) Enlarged images of representative cells. Images from panel A were enlarged to show cells with a particular shape (a rod at 27 h, TCs at 30 and 36 h, a spore at 48 h). Colored arrows point to the same cell classes as in panel A. Bar, 2  $\mu$ m. (C) Representative cellular morphologies. Three-dimensional renderings reconstructed from the image stacks are shown for three cells of each class. Apparent features  $<0.5$   $\mu$ m are artifacts created by shading during the rendering process. Bar, 1  $\mu$ m. (D) Proportion of each cell class at the indicated times PS. Dot, median. Shaded region, 90% credible interval.



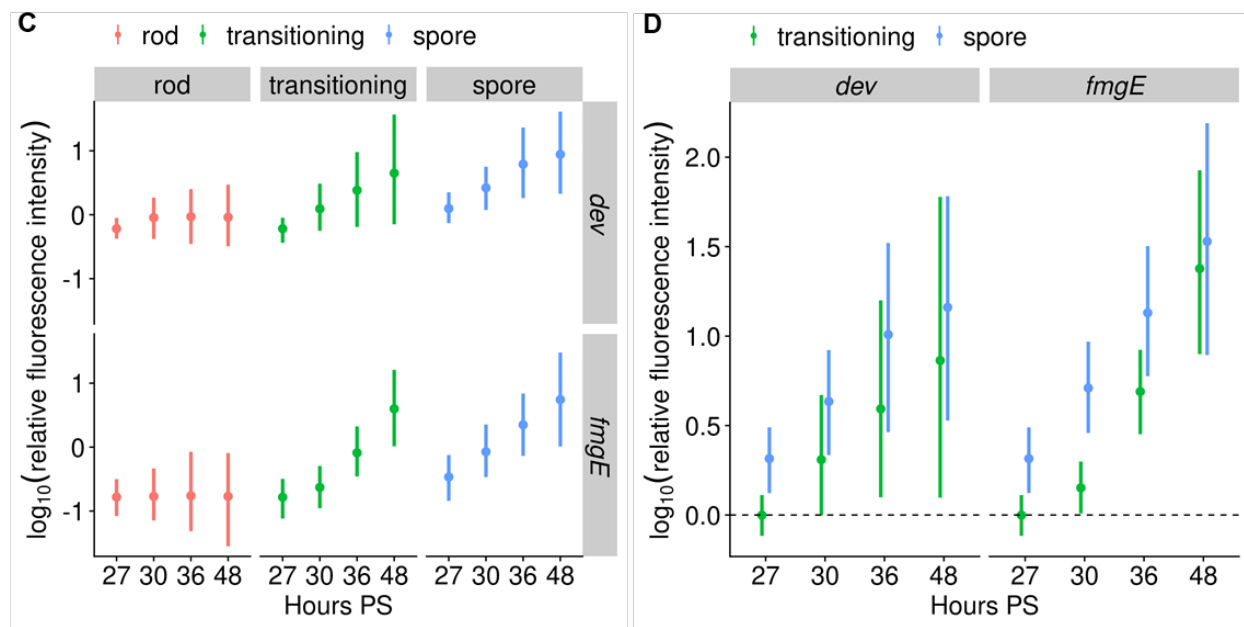
**Figure 2.3: Radial patterns of cell density, alignment, and orientation early correlate with spore proportion later as nascent fruiting bodies mature.** In the experiments described in the legend of Figure 2.2 and in the text, segmented cells from all z-stacks were classified as rods, transitioning cells, or spores. (A) Radial distribution of cell classes and density. The proportion of each cell class at different times poststarvation is shown radially from the center (0  $\mu\text{m}$ ) to the edge (60  $\mu\text{m}$ ) of nascent fruiting bodies (NFBs). Total cell density is normalized to the maximum between the center and the edge of NFBs. Line, median. Shaded region, 90% credible interval. (B) Radial patterns of rod neighbor alignment and tangential orientation. Alignment is the cosine of the angle of a rod to its neighbors averaged using a Gaussian kernel ( $\sigma = 2.5 \mu\text{m}$ ) (1, perfect alignment; 0, orthogonal). Orientation is the cosine of the angle of a rod with the circumference of the NFB (1, tangent to the circumference; 0, orthogonal). (C) Cross-correlations between the radial patterns of total cell density, rod neighbor alignment, and rod tangential orientation early, and that of the spore proportion later. Values are Pearson correlation coefficients and color shade indicates strength of positive (purple) or negative (yellow) correlation (\*  $p < 0.05$ , \*\*  $p < 0.01$ ).

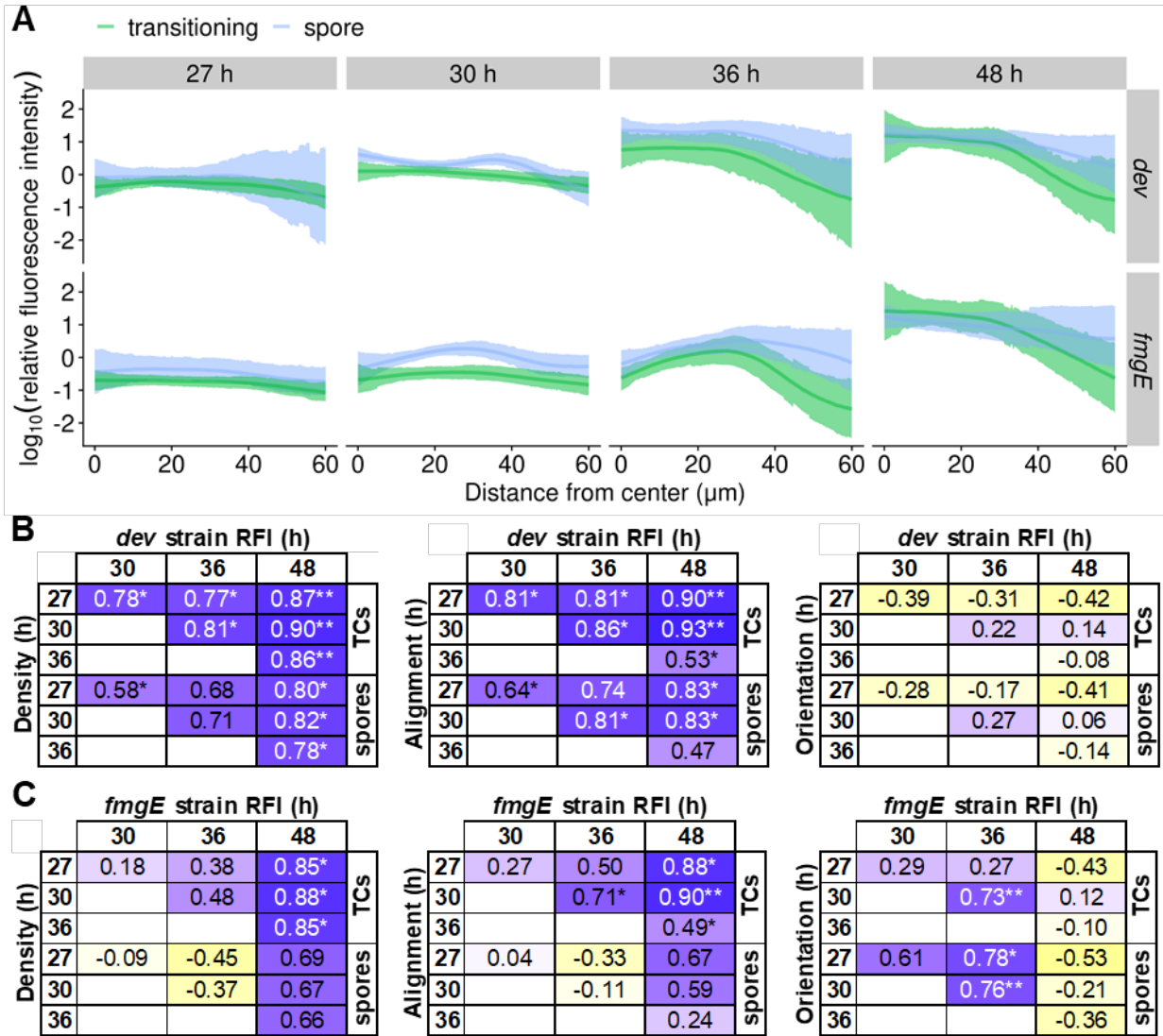




**Figure 2.4: C-signal-dependent gene expression increases in transitioning cells and spores as nascent fruiting bodies mature.** As described in the Figure 2.2 legend, strains YH14 bearing  $P_{dev}$ -tdTomato (designated *dev*) or YH15 bearing  $P_{fmgE}$ -tdTomato (designated *fmgE*) were mixed with unlabeled wild-type strain DK1622 at a ratio of 1 to 5 and co-developed under submerged culture conditions. Both the *dev* and *fmgE* strains also contain Pvan-mNeonGreen. Vanillate (0.5 mM) was added at the start of starvation. Z-stacks of confocal images of the same nascent fruiting body (NFB) were acquired at the indicated times post starvation (PS), measuring both red and green fluorescence. Images are representative of three biological replicates. (A) Confocal images of the same NFB. Images show merged (M) red and green fluorescence. Red arrows, rods. Green arrows, transitioning cells (TCs). Blue arrows, spores. Bar, 20  $\mu$ m. (B) Enlarged images of representative cells. Images from panel A were enlarged to show green (G), red (R), or M fluorescence of cells with a particular shape (rods at 27 and 30 h, TCs at 30 and 36 h, spores at 36 and 48 h). Colored arrows point to the same cell classes as in panel A. Bar, 2  $\mu$ m. (C) Quantification of gene expression. Segmented cells from z-stacks were classified as rods, TCs, or spores based on green fluorescence. For each cell class, the  $\log_{10}$ (relative fluorescence intensity) (i.e., red/green) at each time PS is plotted for the *dev* and *fmgE* strains as indicated. Dot, median. Vertical line, 90% credible interval. (D) Gene expression in TCs and spores. The relative fluorescence intensity of TCs and spores was normalized to that of rods at 27 h (dashed line at  $y = 0$ ). Dot, median. Vertical line, 90% credible interval.

Figure 2.4 (cont'd)





**Figure 2.5: Radial patterns of cell density, alignment, and orientation early correlate with C-signal-dependent gene expression later as nascent fruiting bodies mature.** In the experiment described in the legend of Figure 2.4 and in the text, both red and green fluorescence was quantified for all segmented cells from the z-stacks. The results from three biological replicates were combined. (A) Radial distribution of C-signal-dependent gene expression. For transitioning cells and spores, the  $\log_{10}(\text{relative fluorescence intensity})$  (i.e., red/green) for the *dev* and *fmgE* strains at each time poststarvation is shown radially from the center (0  $\mu\text{m}$ ) to the edge (60  $\mu\text{m}$ ) of nascent fruiting bodies. Line, median. Shaded region, 90% credible interval. (B) Cross-correlations between the radial patterns of total cell density, rod alignment, and rod tangential orientation early, and that of the *dev* strain relative fluorescence intensity (RFI) later. Values are Pearson correlation coefficients and color shade indicates strength of positive (purple) or negative (yellow) correlation. (C) Cross-correlations as in panel B except with the *fmgE* strain RFI (\*  $p < 0.05$ , \*\*  $p < 0.01$ ).



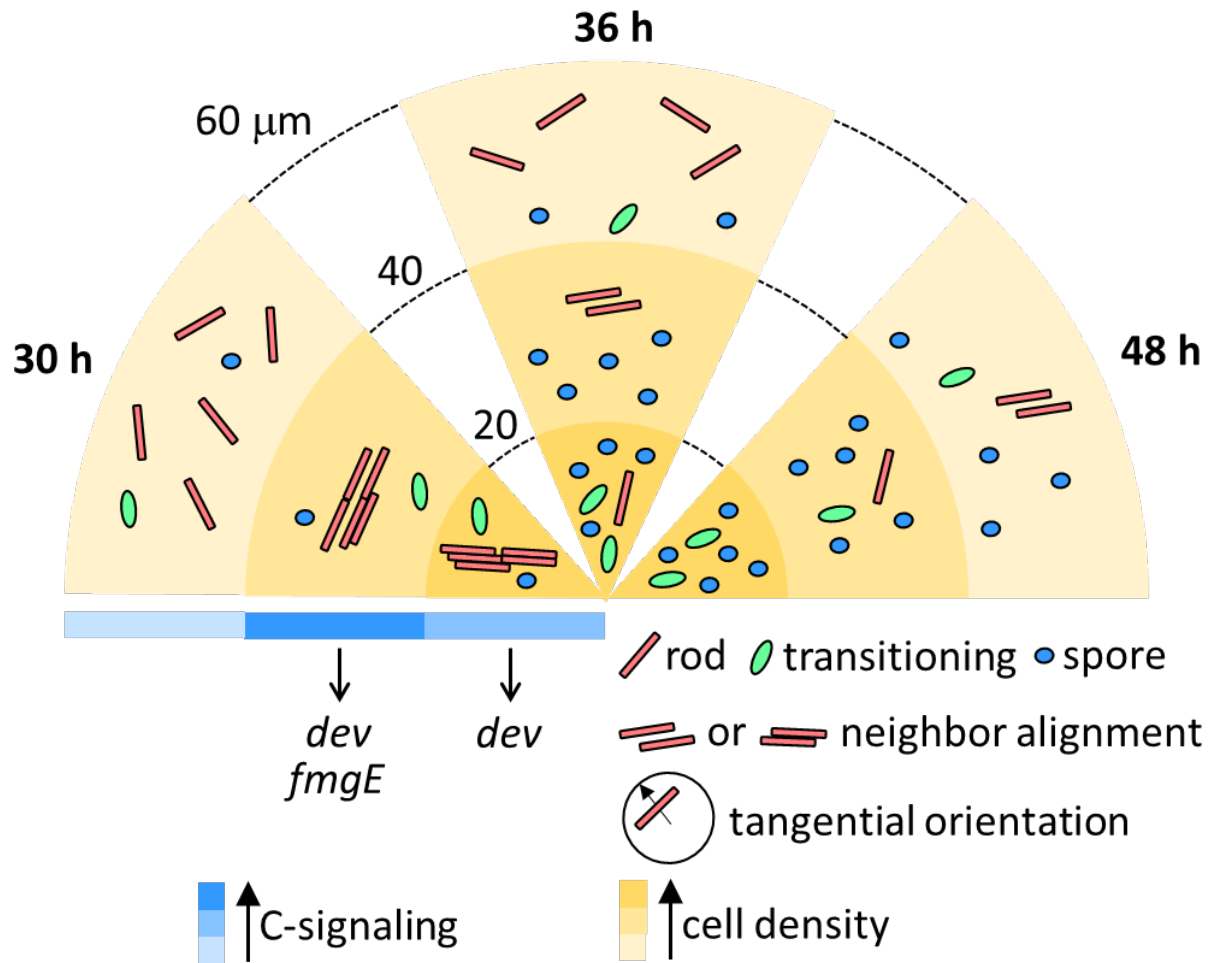


Figure 2.6: Model showing the inferred radial pattern of C-signaling based on observed patterns of cellular arrangement, morphology, and gene expression as nascent fruiting bodies mature. The radial distributions of cell classes, cell density, and the neighbor alignment and tangential orientation of rods are depicted at three time points based on Figure 2.3A and 2.3B. The inferred radial pattern of C-signaling at 30 h poststarvation is based on correlations with later C-signal-dependent *dev* and *fmgE* expression in transitioning cells and spores (Fig. 2.5).

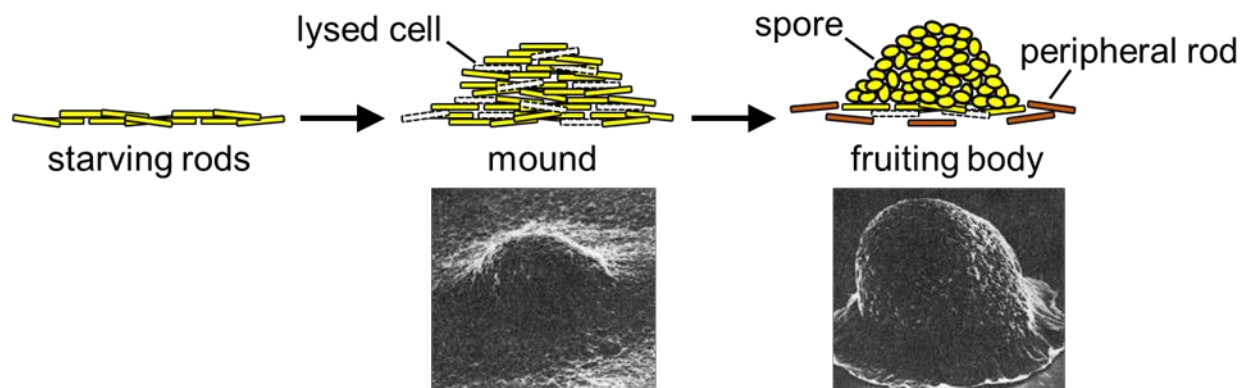


Figure 2.7: **Cartoon and scanning electron micrographs of *M. xanthus* development.** The cartoon at the top depicts starving rod-shaped cells, which coordinate their movements to build a mound, with some cells lysing (dashed outline) during the process. The mound matures into a fruiting body as some rods differentiate into round spores, while other rods remain outside the fruiting body as peripheral rods. The scanning electron micrographs at the bottom show the mound and fruiting body stages. The cartoon is adapted from [89] and the micrographs are from [94] .

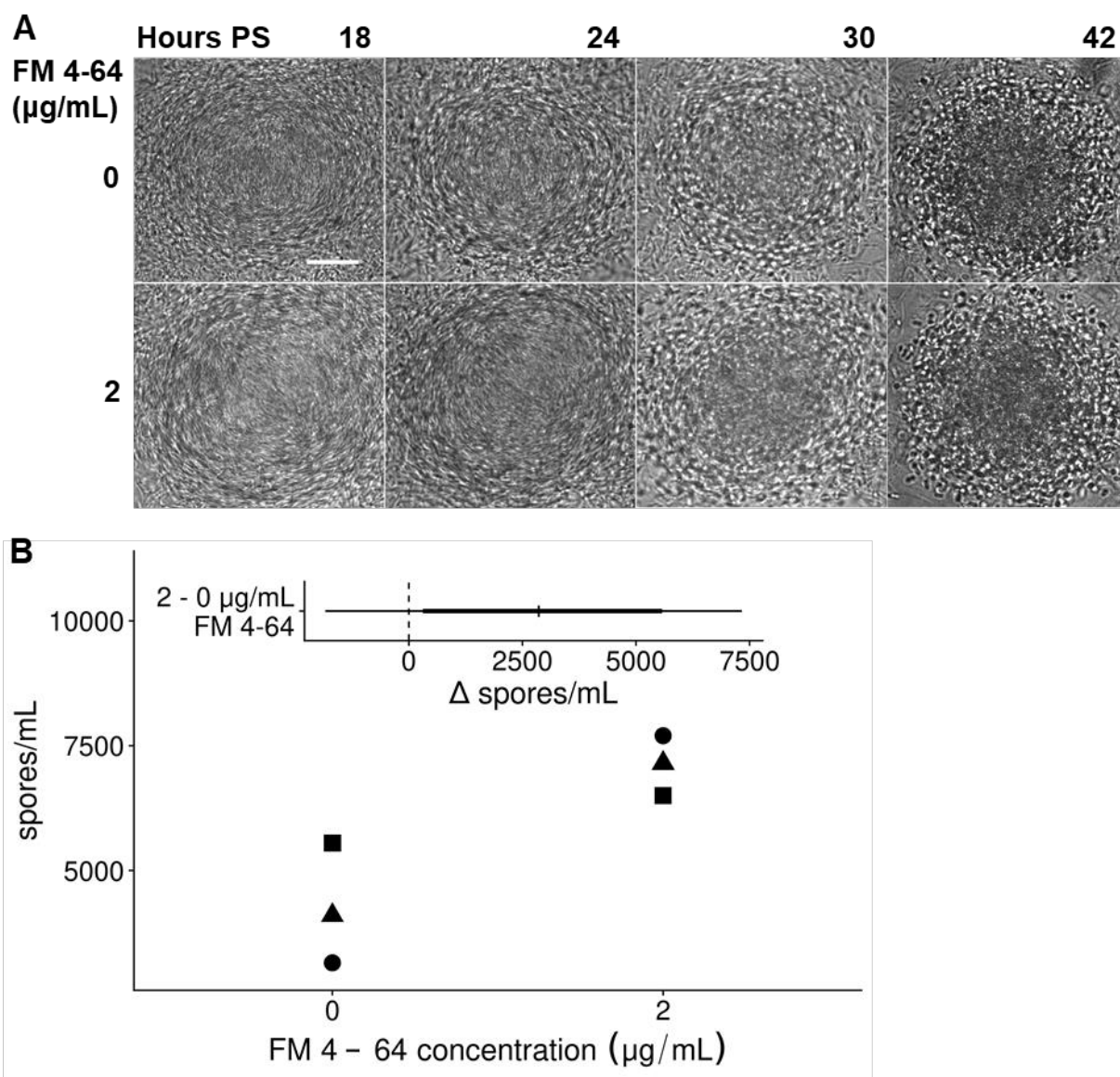


Figure 2.8: **FM 4-64 does not interfere with *M. xanthus* development.** Wild-type strain DK1622 was subjected to starvation under submerged culture conditions. FM 4-64 (2 μg/mL) was added at the start of starvation or withheld. (A) Bright-field images of the bottom of nascent fruiting bodies (NFBs). The same NFB was examined at the indicated times poststarvation (PS). Darkening of mounds typically reflects the formation of spores. Images are representative of three biological replicates. Bar, 20 μm. (B) Quantification of sonication-resistant spores. Samples were collected at 42 h PS, sonicated, and the number of spores/mL was determined. In the main graph, each symbol represents one biological replicate. The insert shows the posterior probability distribution of the difference in the number of sonication-resistant spores/mL in response to 2 vs. 0 μg/mL of FM 4-64. Vertical line, mean. Thick and thin horizontal lines, 90% and 98% credible intervals, respectively.

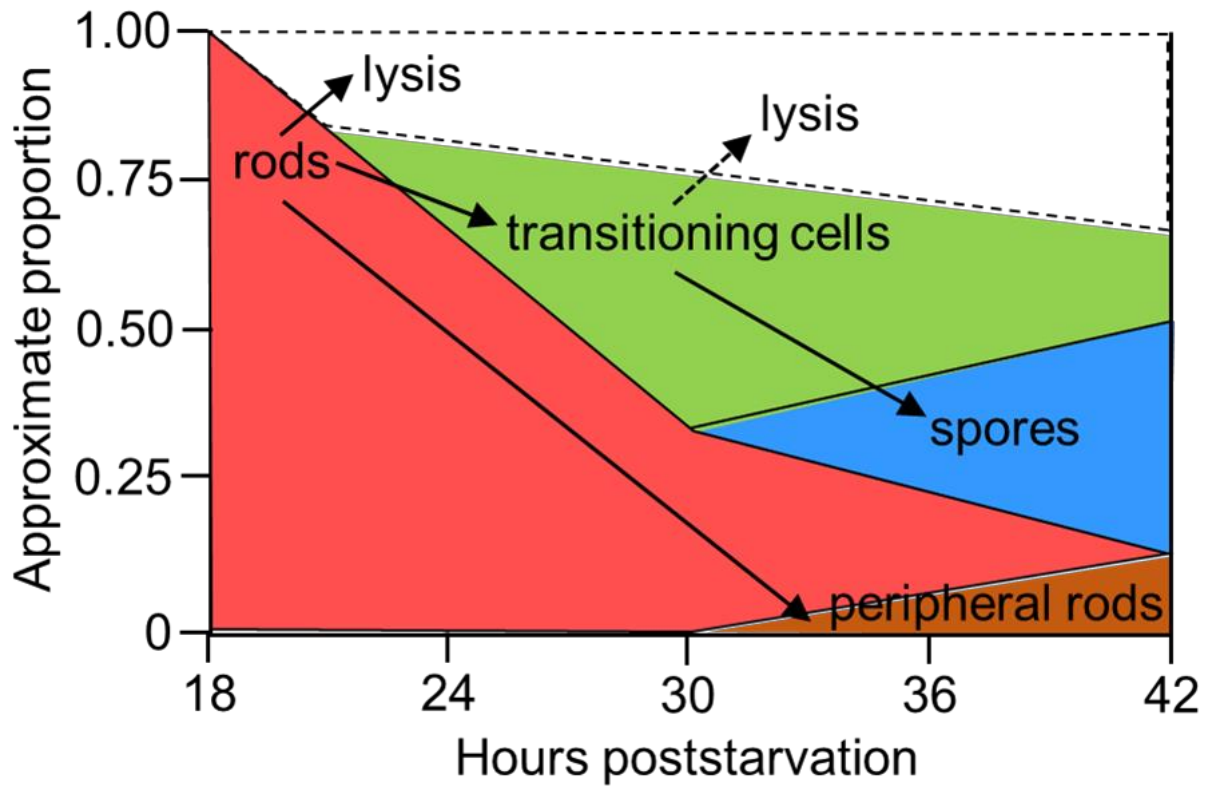


Figure 2.9: **Temporal pattern of cellular shape near the bottom of nascent fruiting bodies.** Based on our observations of cells undergoing submerged culture development and visualized by membrane staining with FM 4-64, the proportion of cells exhibiting the indicated shapes at different times poststarvation was approximated. Rods undergo lysis or become transitioning cells or peripheral rods. Transitioning cells may undergo lysis or persist, but some become spores.

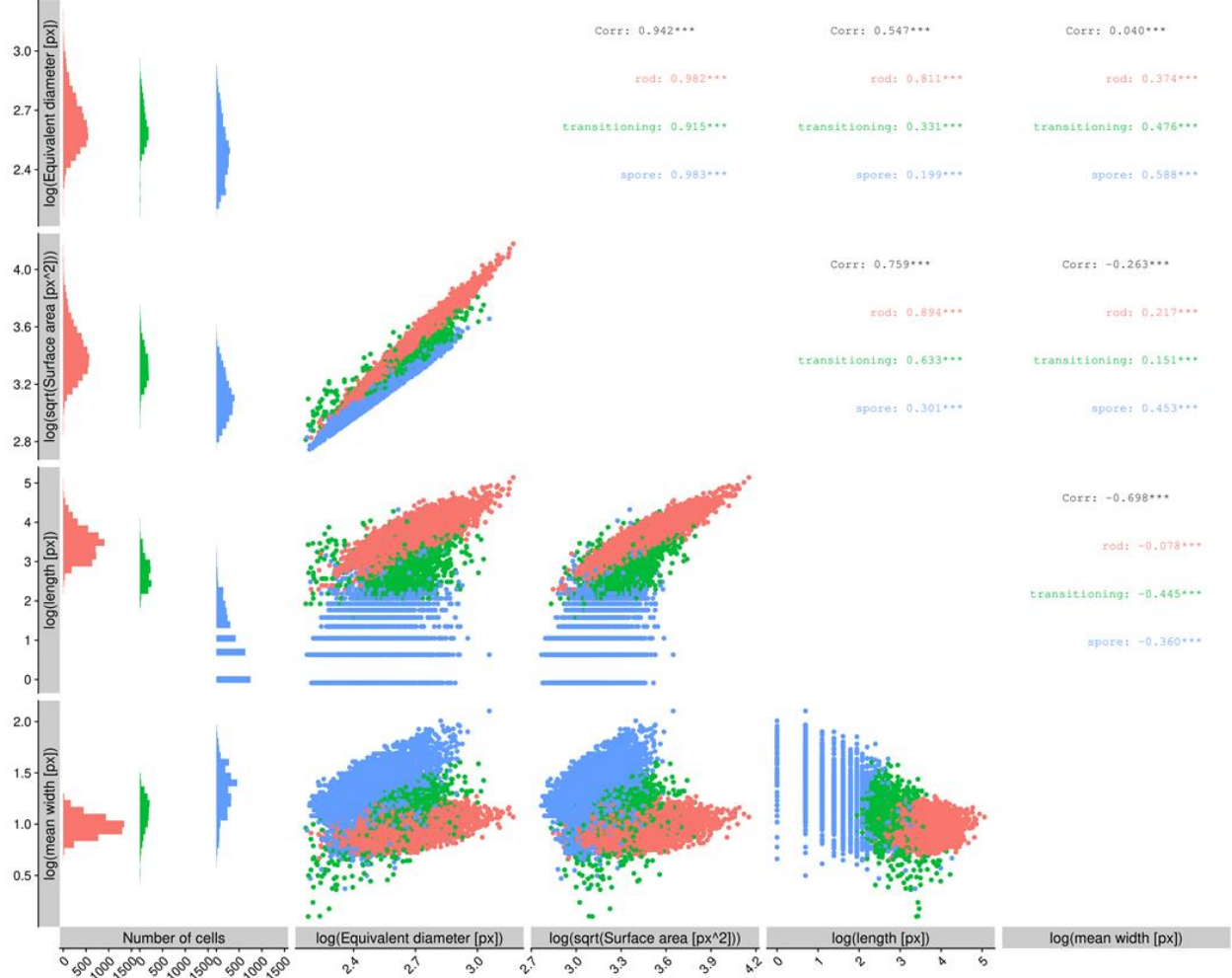


Figure 2.10: **Cell classification using Gaussian mixture modeling.** Each point on the scatter plots represents a cell from the aggregated data across time and fruiting bodies, collected in experiments described in the legend of Figure 2 and in the text. Clusters were determined to be rods (red), transitioning cells (green), or spores (blue) according to their morphology (equivalent diameter, surface area, length, and mean width). The Pearson correlation coefficients are indicated for each pair of properties for all cells (black font) and for each cell class as indicated (\*\*\*) indicates  $p$ -value  $< 0.001$ ). Histograms on the left are the marginal distributions for each cell class for the corresponding morphological property.

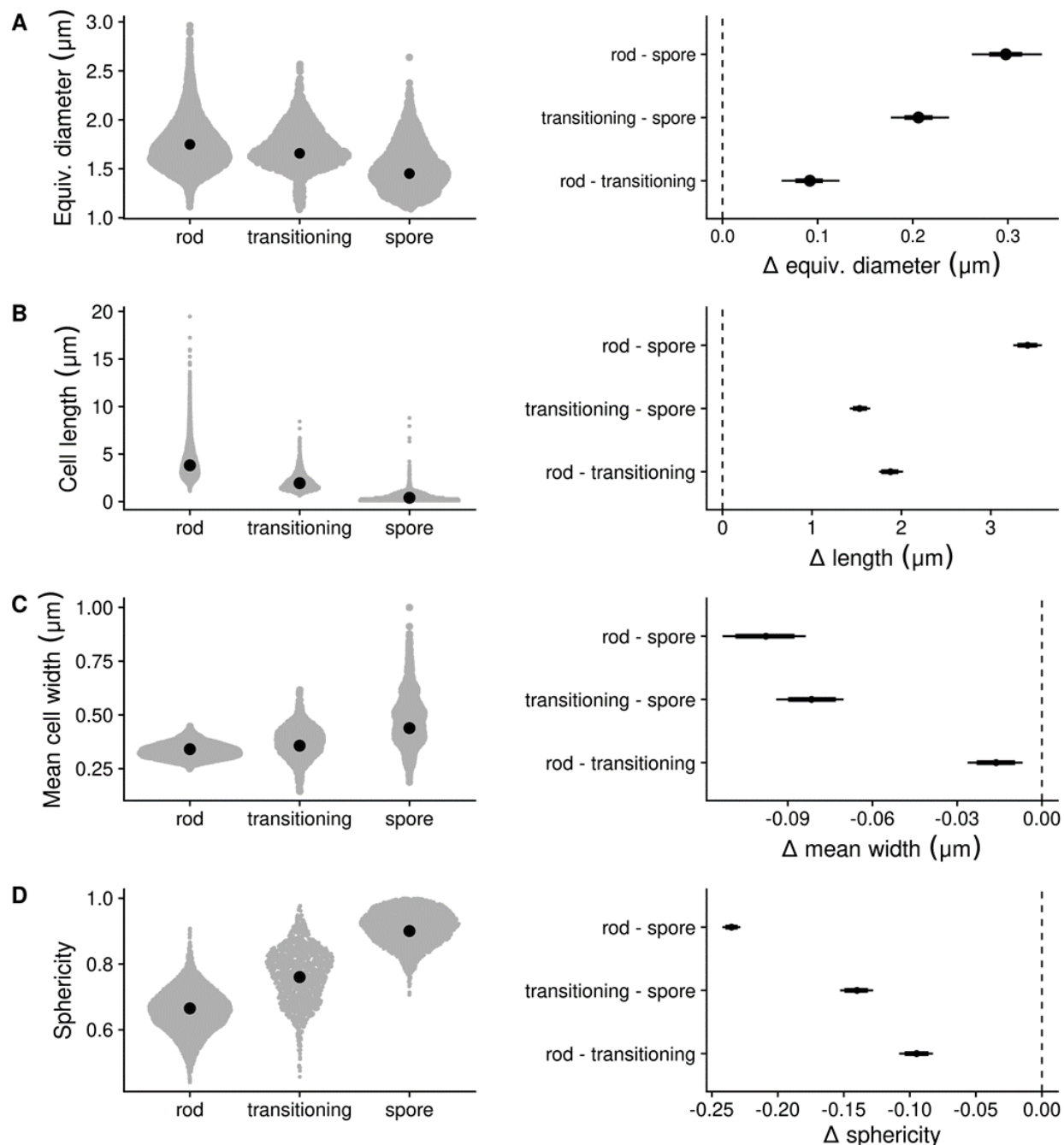


Figure 2.11: **Morphological property distributions of cells classified as rods, transitioning cells, or spores.** In the experiments described in the legend of Figure 2 and in the text, segmented cells from all z-stacks were classified as rods, transitioning cells, or spores. The graph on the left in each panel shows the distribution of the indicated property for each cell class. Gray dot, individual cell. Black dot, median. The graph on the right in each panel shows the posterior probability intervals of the property difference for the indicated pair of cell classes. Dot, median. Thick and thin horizontal lines, 90% and 98% credible intervals, respectively. (A) Equivalent diameter. (B) Cell length. (C) Mean cell width. (D) Sphericity.

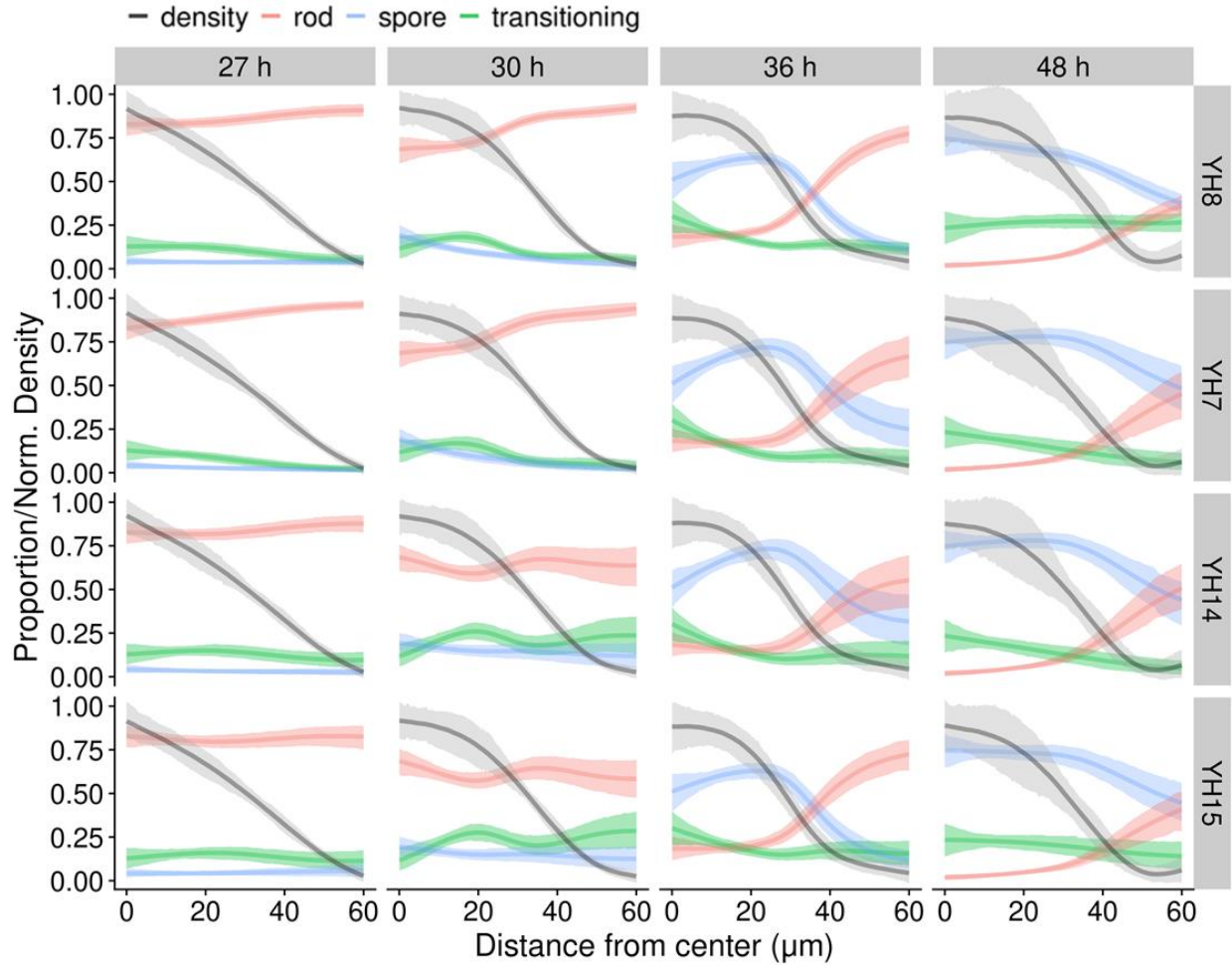


Figure 2.12: **Spatiotemporal distribution of cell classes and density in individual nascent fruiting bodies.** In the experiments described in the legend of Figure 2 and in the text, segmented cells from z-stacks were classified as rods, transitioning cells, or spores. The proportion of each cell class at the indicated time poststarvation is shown radially from the center (0  $\mu\text{m}$ ) to the edge (60  $\mu\text{m}$ ) of one representative nascent fruiting body (NFB) formed by mixing the strain indicated on the right with unlabeled strain DK1622. Total cell density is normalized to the maximum between the center and the edge of the NFB. Line, median. Shaded region, 90% credible interval.



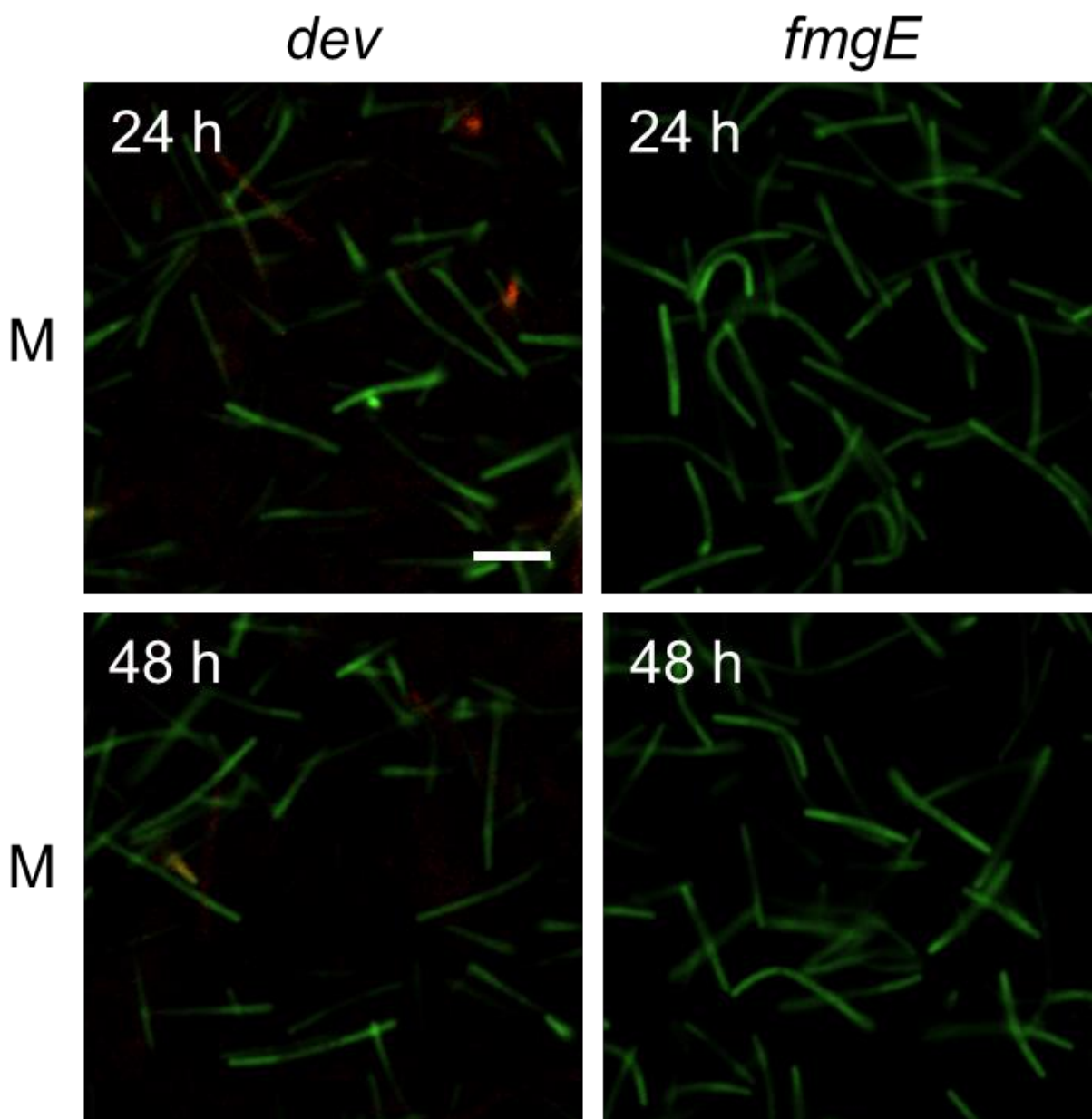


Figure 2.13: **A *csgA* mutant fails to form nascent fruiting bodies and fails to express C-signal-dependent genes.** *M. xanthus csgA* mutant strains YH46 bearing  $P_{dev}$ -*tdTomato* (designated *dev*) or YH45 bearing  $P_{emphfmgE}$ -*tdTomato* (designated *fmgE*) were starved under submerged culture conditions. Both the *dev* and *fmgE* strains also contain  $P_{van}$ -*mNeonGreen*. Vanillate (0.5 mM) was added at the start of starvation. Wells were examined at 6-h intervals from 18 to 48 h poststarvation (PS) using confocal microscopy to visualize both red and green fluorescence. Nascent fruiting bodies failed to form, cells remained rod shaped, and the rods exhibited green fluorescence but showed very little red fluorescence. Confocal images near the bottom of the biofilm were acquired at the indicated times PS. Images show merged (M) red and green fluorescence. Bar, 2  $\mu$ m.



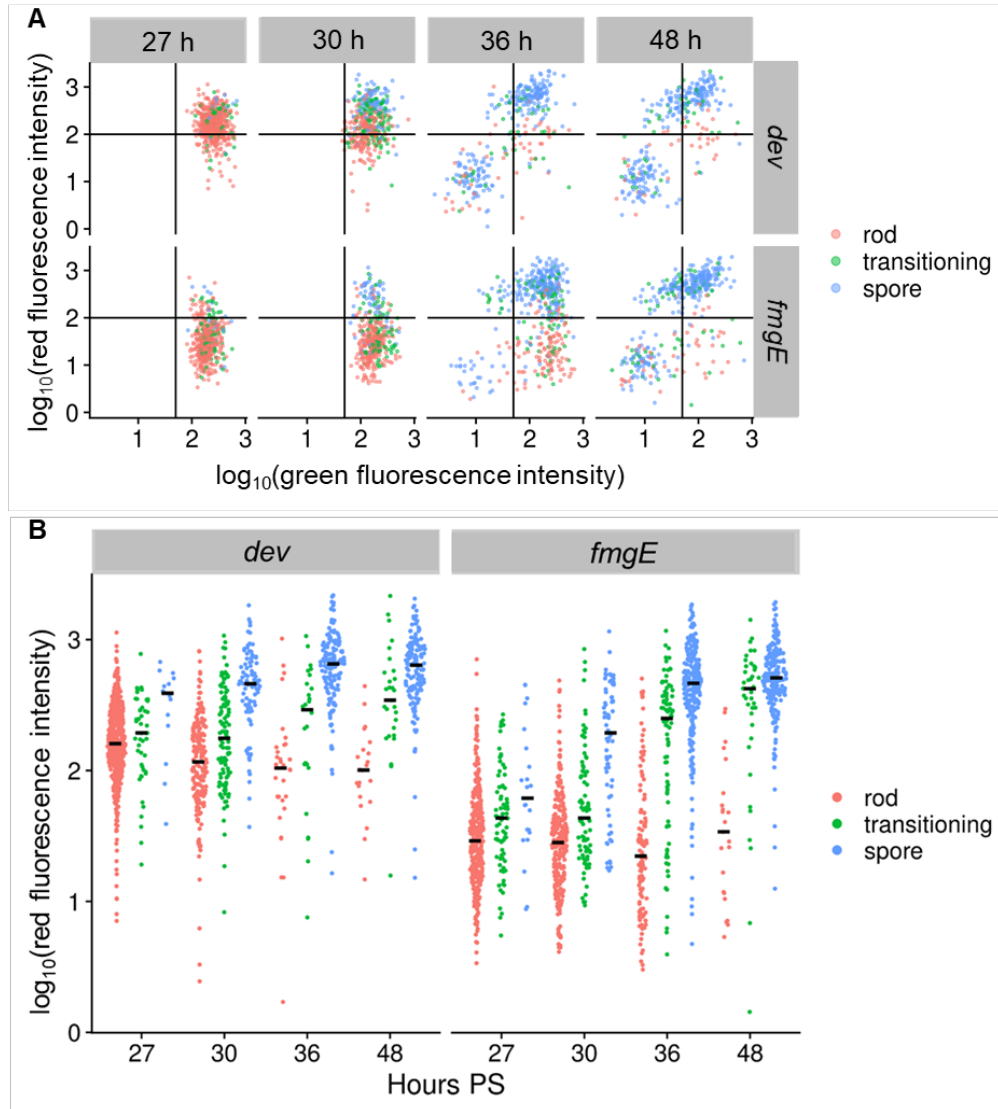
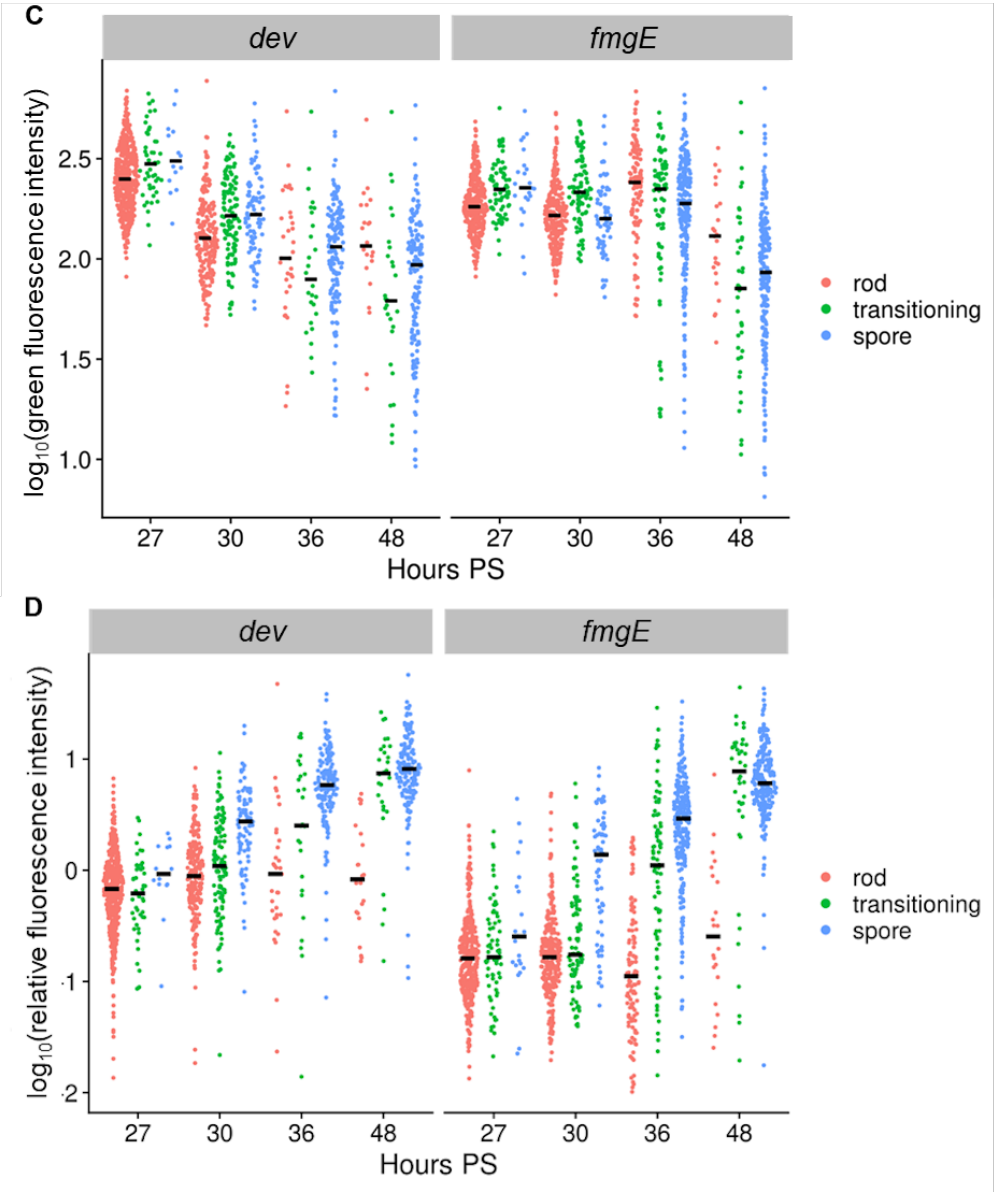


Figure 2.14: **C-signal-dependent gene expression in individual cells of nascent fruiting bodies.** In the experiment described in the legend of Figure 4 and in the text, segmented cells from z-stacks were classified as rods, transitioning cells, or spores based on green fluorescence. The results from three biological replicates were combined. (A) Distributions of red and green fluorescence intensities. For each cell (dot), the  $\log_{10}(\text{red fluorescence intensity})$  and the  $\log_{10}(\text{green fluorescence intensity})$  at each time poststarvation (PS) is plotted for the *dev* and *fmgE* strains as indicated. Cells in the lower left quadrant of each plot (i.e., exhibiting low red and green fluorescence) were eliminated from analyses in panels B-D. (B) Distributions of red fluorescence intensity. For each cell class, the  $\log_{10}(\text{red fluorescence intensity})$  at each time PS is plotted for the *dev* and *fmgE* strains as indicated. Dot, individual cell. Bar, median. (C) Distributions of green fluorescence intensity. Same as panel B except the  $\log_{10}(\text{green fluorescence intensity})$  is plotted. (D) Distributions of relative fluorescence intensity. Same as panel B except the  $\log_{10}(\text{relative fluorescence intensity})$  (i.e., red/green) is plotted.

Figure 2.14 (cont'd)



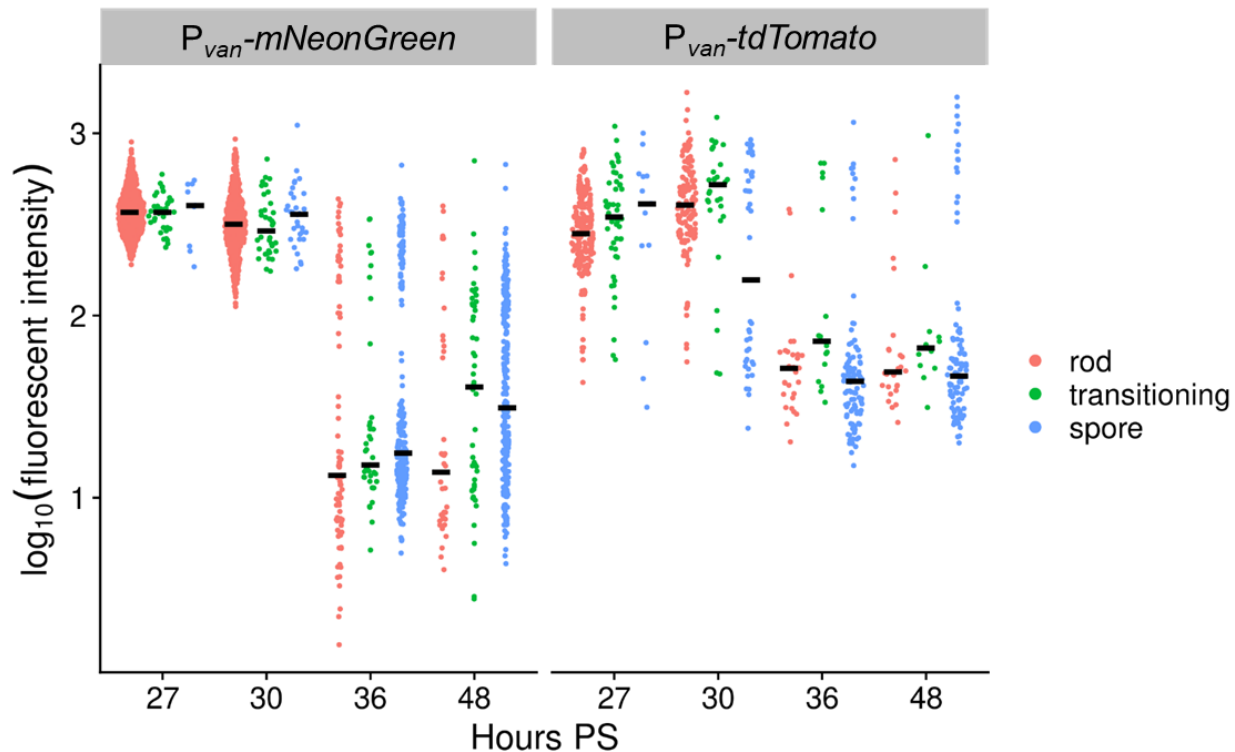


Figure 2.15: **Vanillate-inducible gene expression in individual cells of nascent fruiting bodies.** As described in the Figure 2 legend, strain YH7 bearing  $P_{van}\text{-}mNeonGreen$  or YH8 bearing  $P_{van}\text{-}tdTomato$  were mixed with unlabeled wild-type strain DK1622 cells at a ratio of 1 to 5 and co-developed under submerged culture conditions. Vanillate (0.5 mM) was added at the start of starvation. Z-stacks of confocal images of the same nascent fruiting body were acquired at the indicated times poststarvation (PS) without changing the microscope settings. Segmented cells from z-stacks were classified as rods, transitioning cells, or spores. The results from three biological replicates were combined. For each cell class, the distribution of  $\log_{10}(\text{fluorescence intensity})$  (i.e., green or red) at each time PS is plotted. Dot, individual cell. Bar, median.

## **CHAPTER 3 - COST-BENEFIT ANALYSIS OF PERITRICHIOUS FLAGEL- LATION IN LOW VISCOSITY ENVIRONMENTS**

Authors: Joshua L. Franklin, Sarya Derado, Marc Erhardt, and Yann S. Dufour

## Abstract

Flagellar motility is common in bacteria, but we know little about the cost-benefit trade-offs of synthesizing more than one flagellum. Rough estimates from biochemical data and rapid loss-of-function in evolutionary experiments [138] suggest a relative cost of one to two percent per flagellum in *Escherichia coli*. Given such a high cost, what is the cost-benefit relationship that determines the optimal number flagella? First, we determined the metabolic cost of flagella from optical density measurements of *Salmonella enterica* batch cultures using Bayesian inference with a mixed-effect model. We found that the rate of growth decreases linearly by 1.2% [0.9 – 1.5% 99% CrI] per flagellum. Then, we quantified the swimming performance of cells with different number of flagella by tracking swimming cells with fluorescent flagellar hooks. We found that at low viscosity, the number of flagella does not measurably impact swimming performance. Therefore, we derived a theoretical model to predict if the expression of multiple flagella could be driven by the absence of mechanisms to segregate flagella between daughter cells during division. We predicted the optimal number of flagella given the cost-benefit ratio of producing flagella and provided additional support using agent-based simulations of cells swimming in exponential gradients of nutrients. Our predictions suggest that the synthesis of multiple flagella can be driven by the random inheritance of flagella for up to four or five flagella, as seen in *E. coli* or *S. enterica*, but additional selective pressures are required to explain the expression of higher flagella numbers, as is found in *Bacillus subtilis*.

## Introduction

Owing to their small size, bacteria inhabit a world in which access to resources is severely limited by diffusion [135, 198]. As a result, active motility confers a significant benefit and is widespread in bacteria [43, 71, 79]. While undirected motility alone is advantageous, chemotaxis further increases the benefits by allowing cells to bias their movement toward favorable conditions even in the presence of moderate turbulence or physical obstacles [168, 178, 181].

Bacteria use several mechanisms for motility, but flagellar motility is the most common and well-studied [71]. Flagella location, numbers, and morphology vary considerably between bacterial lineages. Some lineages may have a single flagellum at the cell pole (e.g., *Vibrio cholerae*) or distribute dozens of flagella over the cell body (e.g., *Bacillus subtilis*). While it is expected that different environments apply different selective pressures, the constraints, benefits, and trade-offs underlying the evolution of diverse flagellar morphologies across bacterial lineages is still unclear [54, 123, 168, 197].

Generally, bacteria expressing a single flagellum tend to be found in aqueous environments while bacteria expressing many flagella live in more physically heterogeneous environments such as the gut of animal hosts or soil [151]. Some lineages, such as *Vibrio alginolyticus* or *Azospirillum brasilense*, possess multiple flagellar systems or can regulate the production of lateral flagella to allow swarming motility or swimming in high viscosity environments [5, 56, 114]. For *B. subtilis*, increasing the number of flagella affects swimming behavior and improves spreading on semi-solid surfaces [80, 121, 132, 132]. For *Shewanella oneidensis*, different flagellation patterns were favored when experimental conditions were changed [195]. However, predicting the benefits of flagellation patterns across environments is complicated by the fact that flagella can be used for more than motility, such as attaching to surfaces, forming biofilms, or pathogenesis [8, 28, 34, 129]. Nevertheless, making many flagella appears to be beneficial in various contexts.

Making and powering flagella has a significant metabolic cost for bacteria [9, 51, 109, 115]. Non-motile mutants often outcompete their motile siblings in batch cultures with a significant growth rate advantage [96, 134, 157]. Synthesizing hundreds of thousands of flagellin monomers to extend the flagellar filament (which can account for more than 1% of total protein synthesis) has imposed a strong selective pressure to optimize the composition of flagellins toward energy efficient amino acids [9, 157]. In addition, the flagellar motor expends trans-membrane electrochemical potentials (either proton- or sodium-motive-force), which otherwise drive metabolic processes [160]. The substantial cost of flagellar motility

explains why bacterial lineages that have been examined have evolved elaborate and tight regulation of flagellar synthesis [3, 29, 62, 111, 118, 202]. Therefore, metabolic investment in flagellar motility creates a strong fitness trade-off that should determine an optimal number and arrangement of flagella for bacterial lineages in different environments.

A functional and quantitative characterization of fitness trade-offs is necessary to predict evolutionary trajectories and explain the diversification of complex phenotypic traits across lineages. Despite being well characterized, some aspects of the flagella system in *Enterobacteriales*, such as *Escherichia coli* and *Salmonella enterica*, are not understood from an evolutionary point of view. These bacteria assemble three to five peritrichous flagella but are neither fast swimmers at low viscosity nor proficient at swarming on wet surfaces when compared to other bacterial lineages [45, 130, 131]. In addition, their swimming behavior does not appear to be sensitive to variation in the number of flagella [113, 171]. Therefore, it is not immediately evident why *E. coli* or *S. enterica* make three to five flagella on average.

Here, we characterize the cost-benefit trade-off for peritrichous flagellation with a low number of flagella using a genetically engineered strain of *S. enterica*. We gradually induced flagellar expression and fluorescently labeled flagellar hooks to count the number of flagella in single cells. We then determined the direct cost of flagellar expression by precisely measuring the growth rates of populations making different numbers of flagella. We also used single cell tracking and labels to determine that variation in flagella number has little effect on swimming performance. However, we demonstrated that making multiple flagella can provide a significant fitness advantage when the partition of flagella is not controlled during cell division using an agent-based model of flagellar inheritance.

## Results

### **The mean length of flagellar filaments is robust to changes in FlhDC activity.**

To characterize how change in the number of flagella affects growth rate and swimming performance in *S. enterica* serovar *typhimurium* LT2, we generated a mutant strain in which we can ectopically control flagellar expression and count flagella in single cells (EM8242).

We placed the master regulator of flagellar gene expression (FlhDC) under the control of an inducible promoter (pTet) and introduced a cysteine residue in FlgE (S171C) to enable specific fluorescent labelling of flagellar hooks. Another strain producing FlgE-S171C but with a wild type flhDC locus was generated for comparison (TH9677).

Flagellar hooks produce small and consistent fluorescent foci that can be accurately counted using automated image analysis as opposed to flagellar filaments that have complex overlapping shapes (Figure 3.1A). We visually verified in 500 cells that fluorescently labeled flagellar hooks were always detected at the base of flagellar filaments using our automated image analysis protocol.

Next, we characterized the production of flagellar hooks as a response to ectopic induction of pTet-flhDC with anhydrotetracycline (AnTc) (Figure 3.1B). In EM8242, no fluorescent foci were detected in the absence of inducer and production saturated to 2 foci per cell on average with AnTc concentrations above 1.5 pg/mL. In TH9677, the wild type flhDC promoter produced 4 foci per cell on average. The distribution of the number of foci in each population was most consistent with a negative binomial distribution with a constant overdispersion parameter.

Finally, we examined if the number of hooks expressed had an effect on the lengths of the flagellar filaments produced in each cell (Figure 3.1C). The length of each flagellum was measured by manually tracing flagella with an image processing software for 100 cells randomly selected from each of the 5 induction conditions. The distribution of filament lengths was fitted with a gamma distribution with a constant shape parameter. The effect of the number of hooks on the filament length was estimated to be 0.05  $\mu\text{m}/\text{hook}$   $([-0.07, 0.16]$  95% credible interval (CrI)). There is no strong evidence that changing FlhDC expression has a large effect of the lengths of the produced flagellar filaments in our experimental conditions. Therefore, we can assume that the overall cost for producing flagella is proportional to the number of detected flagellar hooks.



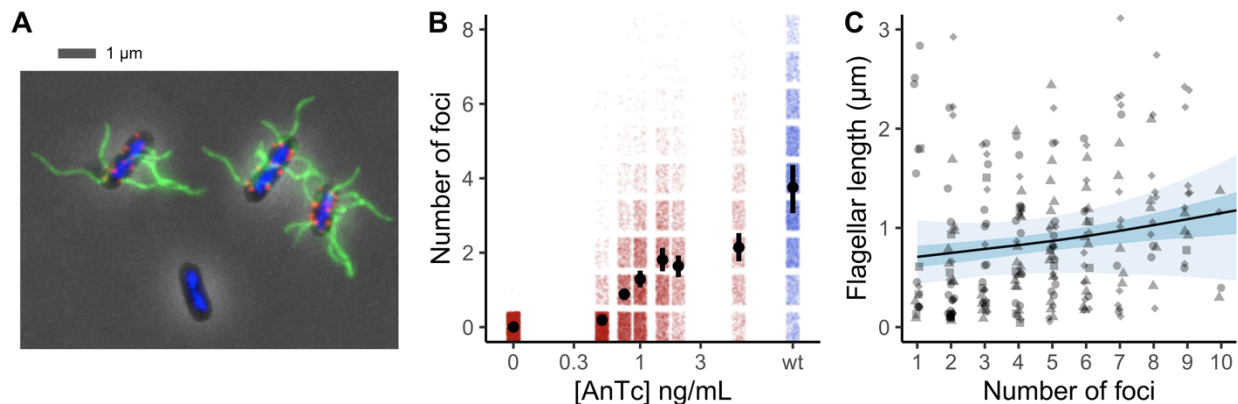


Figure 3.1: **Ectopic control of flagellar expression and flagellar hook detection in *S. enterica*.** A) Composite image of fluorescently labeled EM8242 after induction of flagellar expression ([AnTc] = 2,000 pg/mL). DNA was stained with DAPI (blue), flagellar hook with a maleimide conjugated fluorescent dye (red), and flagellar filament with anti-flagellin antibodies conjugated to a fluorescent dye (green). B) Number of flagellar hook foci as a function of inducer concentration in single cells (750 cells per condition). Each color dot represents a cell (red: EM8242, blue: TH9677). The estimated mean for each population is indicated in black (dot: mean, line: 95% CI). C) Mean lengths of flagella as a function of the number of foci in each cell (symbols represent different AnTc concentrations). The relationship was fitted using a linear model with a gamma response (black line: mean, light and dark blue: 50% and 95% CI).

### Cell growth decreases proportionally to the number of flagella produced.

To quantify the effect of flagellar synthesis on the average population growth rate, we measured the change in optical density of batch cultures producing different numbers of flagella. Expression of FlhDC was controlled in EM8242 by adding varying concentrations of AnTc. Cultures were grown in Vogel-Bonner E minimal salts with 2% w/v yeast extract added to prevent the bimodal regulation of FlhDC by RflP [180].

The difference in growth rates between a strain producing no flagella (EM8242 without AnTc) or 4 flagella (TH9677) was very small, as expected (Figure 3.2A). Cultures were inoculated with cells already growing exponentially in the same conditions to avoid any lag in growth and the growth rates were calculated while culture were still at low cell density ( $\text{OD}_{600} < 0.4$ ). To rule out any direct effect of AnTc, we also grew TH9677 in the presence of 20 ng/mL AnTc and found no negative effect on growth rate (Figure 3.6).

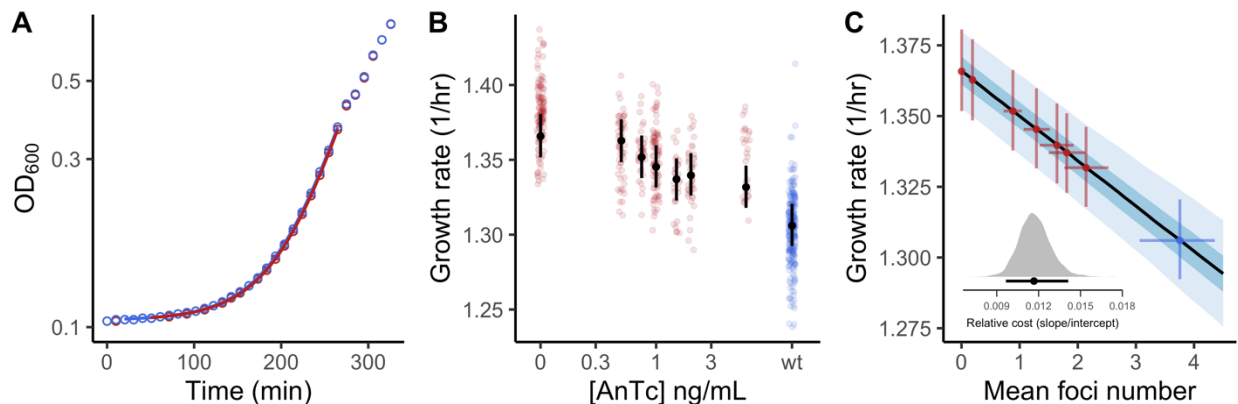
Because experimental variability was relatively large compared to the expected effect of flagellar synthesis, precise estimates of growth rates were obtained through extensive replication (between 30 and 200 per conditions). As a result, we were able to resolve the monotonic decrease in growth rate as flagellar synthesis increases ( $p\text{val} < 10^{-4}$ ) (Figure 3.2B). As expected, TH9677 grew slower than fully induced EM8242.

Finally, the cost of flagellar synthesis was estimated by directly comparing the populations growth rates against mean foci numbers by matching populations according to the AnTc concentrations used (Figures 3.1B and 3.2B). The relationship between the two phenotypes was consistent with a linear model within the range of conditions tested (Figure 3.2C). We estimated the slope and intercept of the linear relationship by integrating the models we used to describe each phenotype independently. The mean growth rate of cells that do not produce flagella was estimated to be  $1.36 \text{ hr}^{-1}$  ( $[1.35, 1.38]$  95% CrI) while the reduction in growth rate for each flagellum produced was estimated to be  $0.0159 \text{ hr}^{-1}/\text{foci}$  ( $[0.0131, 0.0192]$  95% CrI). Therefore, the relative cost for each flagellum produced is 1.17% ( $[0.96\%, 1.41\%]$  95% CrI) in our growth conditions.

Our direct measurement of the relative metabolic cost of producing flagella in *S. enterica* is consistent with prior estimates obtained from competition experiments and biochemical calculations [9, 134].

### **Swimming performance at low viscosity is not sensitive to the number of flagella.**

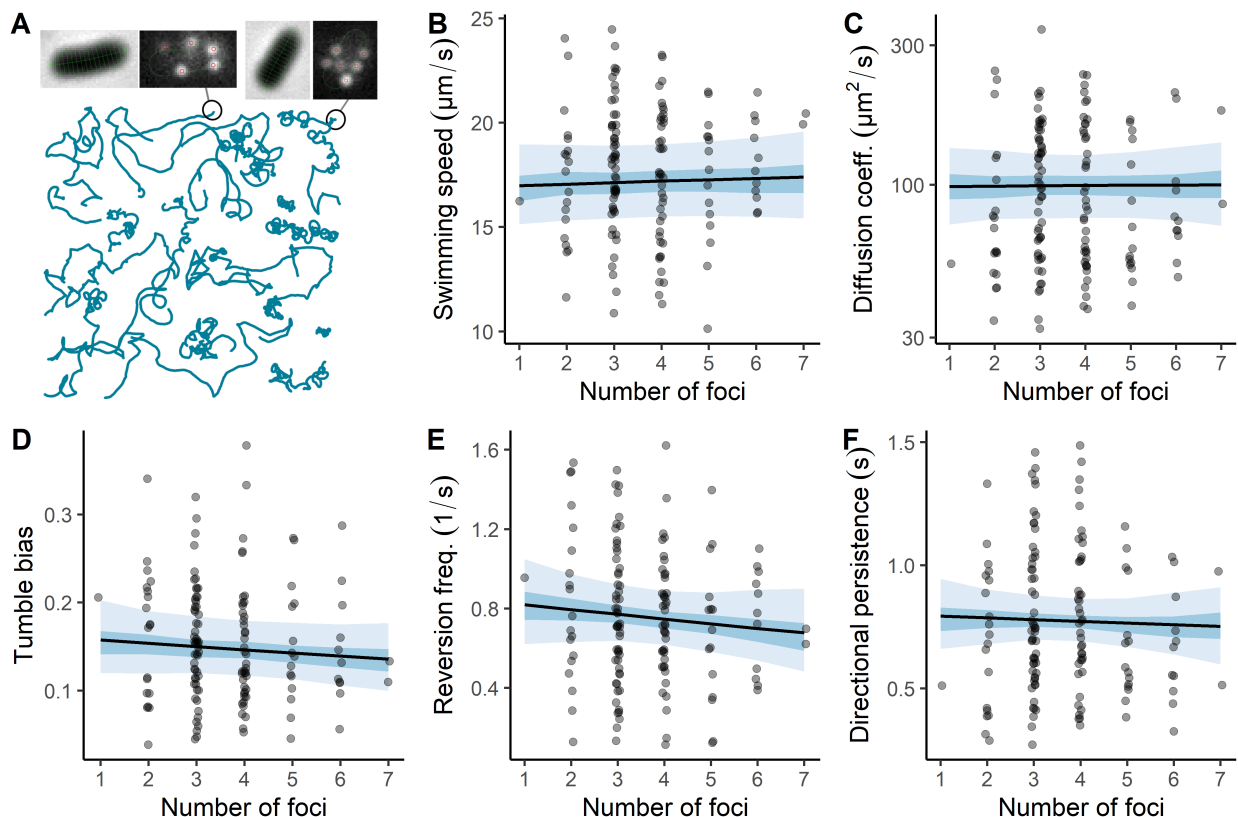
While there is a clear cost for producing several flagella, it is not yet clear how their number would affect the swimming behavior of *S. enterica* in liquid environments. We tested if varying number of flagella had an effect on the swimming performance of single cells using an experimental assay previously described [45]. We tracked the trajectories of single TH9677 cells swimming freely at low viscosity for several minutes and then captured high-magnification phase contrast and epifluorescence micrographs (Figure 3.3A). Morphological parameters and the number of fluorescently labeled flagellar hooks were determined for each cell. Cells that were not motile or had a size (length or width) not within 3 standard



**Figure 3.2: The relative metabolic cost of flagellar synthesis.** A) Calculating growth rate by monitoring change in culture optical density at 600 nm ( $OD_{600}$ ) over time. Representative measurements from non-induced EM8242 (red circle) and TH9677 (blue circle) are overlaid. Growth rates ( $r$ ), background OD, and OD of the inoculum were estimated by least-square fitting of an exponential growth model (solid lines) while cells were at low density. B) Calculated growth rates as a function of inducer concentration for EM8242 (red) and TH9677 (blue). Each dot represents a single batch culture (between 30 and 200 per condition). The estimated growth rate for each condition is indicated in black (dot: mean, line: 95% CrI). C) Regression of growth rate as a function of number of foci. Each dot represents a single inducer concentration (EM8242 in red) or wild type FlhDC activity (TH9677 in blue). The crossed lines represent the 95% credible interval from Figure 3.1B and 3.2B. The black line represents the mean of the regression model (dark and light blue areas: 50% and 95% CrI). Inset: Posterior probability distribution of the relative cost (slope/intercept) of one flagellum of the growth rate (dot: mean, line: 95% CrI).

deviations of the population mean were excluded from the following analyses.

We tested if the number of flagellar hooks had a measurable effect on several behavioral parameters calculated from the cell trajectories (Figure 3.3B-E). Linear regression analyses indicate that swimming performance at low viscosity is not sensitive to change in the number of flagella (swimming speed/foci = 0.004 mm/s [-0.016,0.024] 95% CrI; diffusion coefficient/foci = 0.001 mm<sup>2</sup>/s [-0.066,0.068] 95% CrI; tumble bias/foci = -0.025 [-0.082,0.030] 95% CrI; reversion frequency/foci = -0.032 s<sup>-1</sup> [-0.103,0.045] 95% CrI; directional persistence/foci = -0.009 [-0.064,0.043] 95% CrI). In addition, regression analyses against cell length and width did not reveal any significant effect, as expected, since cells of unusual size were excluded.



**Figure 3.3: Swimming performance of TH9677 with varying number of flagella.** A) Illustrative examples of single cell tracking followed by detection of fluorescently labeled flagellar hooks after immobilizing cells with a UV-activated hydrogel. B-F) Linear regressions of behavioral parameters calculated from cell trajectories against number of foci from labeled flagellar hooks. Varying number of foci does not explain the variability in swimming performance. Each dot represents a single cell. The black line represents the mean of the regression model (dark and light blue areas: 50% and 95% CrI).

### Making multiple flagella is beneficial even when swimming performance is not affected.

The high cost of flagellar synthesis and the lack of effect from varying flagella number on swimming performance raise the question of why *S. enterica* makes more than one flagellum. One model may be that flagellar motility in *S. enterica* has not evolved to be optimal at low viscosity. Alternatively, because the location of new flagella on the cell body is not controlled (although flagella are less likely to be near the poles) (Figure 3.6) [151], the inheritance of flagella as cells divide may create a selective trade-off. The random allocation of flagella to

daughter cells can generate non-flagellated cells that may be at a severe disadvantage since it can take several minutes to synthesis a new flagellum [51].

To examine the trade-off between the cost of making flagella and reducing the production of non-flagellated cells, we created an agent-based model to simulate population growth and evolutionary dynamics. The growth rate of each agent in a population is determined by its position along a linear chemical gradient minus the proportion of resources allocated for flagellar synthesis. The agent position along the gradient is determined by the equilibrium between a chemotactic force (inversely proportional to the chemical concentration) and gravity (Figure 3.4A). Non-flagellated agents eventually sediment to the bottom of the gradient where they have access to fewer resources. Agents make flagella by allocating a fraction of the acquired resources to flagellar synthesis until a predetermined cost is overcome. New flagella are placed randomly along the agent body length. Three models for cell growth, which affect the inheritance of flagella at cell division, were examined (Figure 3.4B). *S. enterica* elongates by inserting new peptidoglycan diffusively along the sidewall (random model), but other flagellated bacterial lineages insert new peptidoglycan near a cell pole (e.g. *Agrobacterium*) or the mid cell (e.g. *Caulobacter*) [24, 27].

To determine the optimal number of flagella for different conditions, we evolved populations of agents by allowing the rate of flagellar synthesis to mutate as agent replicate until the rate reached steady state. We varied the cost of making a flagellum and the reward for being flagellated by varying the chemotactic force pulling agents up the gradient (the resulting growth reward was calculated from the population average). As a result, we found that the optimal number of flagella is a sublinear function of the growth reward that is scaled down when flagellar cost is increased (Figure 3.8). Remarkably, the optimal number of flagella collapse to a single function when the variable is changed to the reward over cost ratio (Figure 3.4C). However, the mode of growth (thus, flagellum allocation) still affects the scale of the relationship.

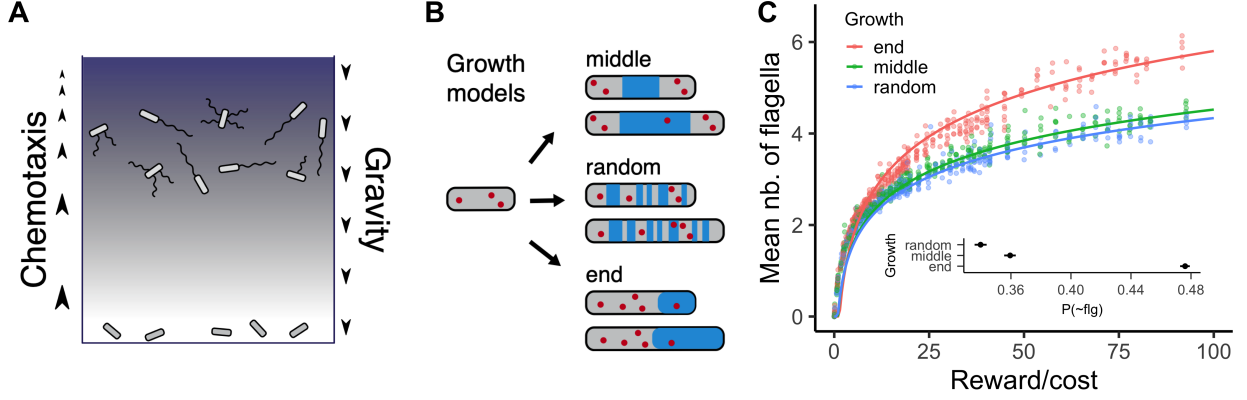


Figure 3.4: ***In silico* evolution of the optimal number of flagella in different conditions.** A) Schematic representation of the agent-based model. Flagellated agents swim up with a speed inversely proportional to the chemical gradient (blue) and are pulled down by gravity. Position along the gradient and determines the resources available for growth and flagellar synthesis. The rate of flagellar synthesis is allowed to evolve by natural selection as the population grows until steady state is reached. B) New flagella (red dots) are inserted randomly on the agent body. Cell elongation follows one of three alternative models for cell wall synthesis (blue) to determine the allocation of flagella to daughter cells. C) Results from repeated *in silico* evolutionary experiments with different combinations of flagellar cost and growth reward from chemotaxis. Each dot represents the mean number of flagella per cell in an independent population after its evolutionary trajectory reached steady state. The solid lines represent the best fit from our mathematical model (with  $P(-\text{flg})$  as a free parameter). Inset: Mean and 95% CrI for the parameter  $P(-\text{flg})$  fitted to simulation results.

### Making multiple flagella is beneficial even when swimming performance is not affected.

To gain further insights into the functional trade-off determining the optimal number of flagella, we derived a mathematical model describing the average population growth rate ( $r$ ) when flagella are randomly allocated between daughter cells:

$$r = p^n(r_0 - cn) + (1 - p^n)(r_0 - cn + \delta) \quad (3.1)$$

where,  $r_0$  is the basal population growth rate with no flagellar synthesis or reward from chemotaxis,  $c$  is the growth cost per flagellum produced,  $n$  is the average number of flagella produced,  $\delta$  is the growth reward for having at least one flagellum, and  $p$  is the probability of not inheriting each flagellum after cell division. By taking the derivative of equation 3.1 as a

function of  $n$ , we can calculate the number of flagella that should be produced to maximize the average population growth rate ( $\frac{dr}{dn} = 0$ ):

$$n = \frac{\log \frac{-1}{\log p} \frac{c}{\delta}}{\log p} \quad (3.2)$$

Equation 3.2 revealed that the optimal number of flagella is determined by  $\delta/c$  (the reward to cost ratio) and  $p$  (the probability of not inheriting a flagellum) (Figure 3.5A). The functional form of equation 3.2 fitted the results well even though the population dynamics in the agent-based evolutionary experiments are expected to be more complex. The nonlinear growth dynamics generated by the different growth modes and chemotactic behavior average into an effective probability of flagellar inheritance ( $p$ ). Therefore, we estimated  $p$  from the results of the agent-based simulations to characterize the influence of different cell growth modes on the inheritance of flagella in peritrichous bacteria (Figure 3.4C inset). As expected, growing from the cell pole is more likely to generate non-flagellated cells ( $P_{\text{end}}(\neg\text{flg}) = 0.47$ ), while a more balanced allocation of flagella resulted from diffusive elongation ( $P_{\text{random}}(\neg\text{flg}) = 0.34$ ) or mid cell elongation ( $P_{\text{middle}}(\neg\text{flg}) = 0.36$ ). Our model also predicted that an extremely large reward to cost ratio would be necessary to justify making more than half a dozen flagella if the only goal was to minimize the probability in producing non-flagellated cells during growth (Figure 3.5A).

Equation 3.2 can be rearranged to calculate the expected growth reward ( $\delta$ ) from producing a given number of flagella ( $n$ ) at a given cost ( $c$ ):

$$\delta = -cp^{-n}/\log p \quad (3.3)$$

We calculated the expected reward using the parameters we estimated for *S. enterica* and assuming a simple evolutionary model. To maintain the production of 4 flagella on average, the reward from chemotaxis must increase the average population growth rate by at least 80% (Figure 3.5B) but this relationship is sensitive to the mode of flagellar inheritance. As

expected, fewer flagella are necessary when both daughter cells are more likely to inherit at least one. Finally, the optimal number of flagella,  $n$ , is more sensitive to the probability,  $p$ , when the reward to cost ratio,  $d/c$ , is high (Figure 3.5C). When the reward to cost ratio is very small, the ability to maximize the inheritance of flagella ( $p = 0$ ) becomes critical to make flagellar motility worthwhile. Overall, peritrichously flagellated bacterial lineages should disfavor growing cell wall from the poles, and oligotrophic environments may strongly select for mechanisms to ensure flagellar inheritance.

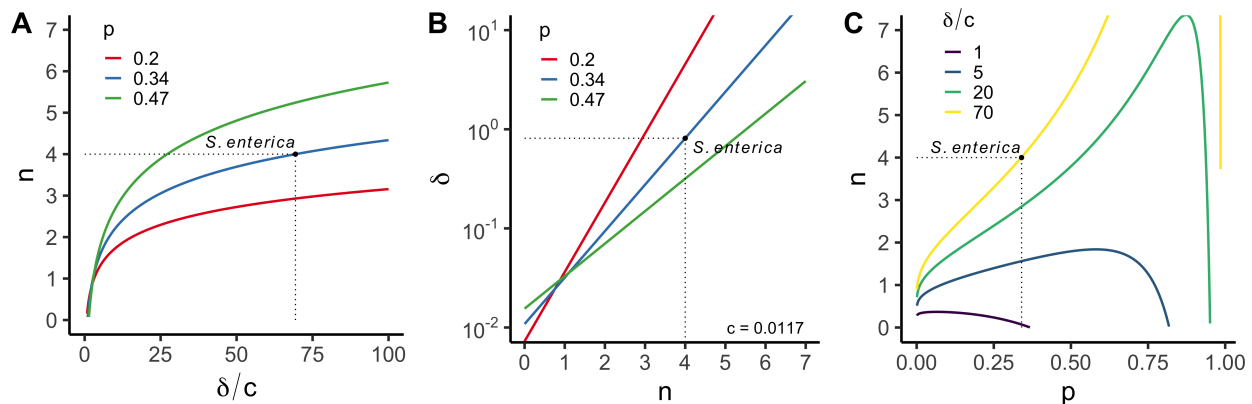


Figure 3.5: **Mathematical predictions of the optimal number of flagella.** A) Optimal number of flagella ( $n$ ) as a function of the reward to cost ratio for chemotaxis ( $\delta/c$ ) for different probability of not inheriting a flagellum after cell division ( $p$ ) from equation 3.2. B) Expected growth reward ( $d$ ) as a function of number of flagella ( $n$ ) for different probability of not inheriting a flagellum ( $p$ ) and a fixed cost of 1.17% of the growth rate per flagellum ( $c$ ) from equation 3.3. C) Expected number of flagella ( $n$ ) as a function of the probability of not inheriting a flagellum ( $p$ ) for different reward to cost ratios ( $\delta/c$ ) from equation 3.2. The parameter values estimated in this study for *S. enterica* are indicated with dotted lines.

## Discussion

We find a growth rate cost of approximately 1% per flagellum, which is approximately in line with previous evolutionary loss-of-function experiments, and biochemical estimates [138]. A growth rate decline of 1% per flagellum is substantial in evolutionary terms [138], with our reference *S. enterica* strain growing approximately 4.5% slower than our inducible strain when expressing no flagella (Figure 3.2B).

The substantial cost of producing flagella indicates that cells derive some benefit from



making multiple flagella. But we know that cells with a single polar flagellum can swim well [122], and our FAST assay data indicates that peritrichously flagellated cells with a single flagellum can also swim effectively. Additionally, given that our inducible strain could stably express fewer flagella than the reference strain, it seems clear that peritrichously flagellated cells could evolve a lower flagella number. Taken together, these facts imply that the expression of multiple peritrichous flagella is maintained by active selection.

However, we did not find any strong effect of flagella number on a variety of swimming parameters, though we could not rule out weak effects. The FAST assay is currently limited to examining swimming behavior in a low viscosity regime, and it is possible that flagella number plays a more important role at higher viscosities. Indeed, this is supported by evolution experiments repeatedly passaging cells in swimplates [123]. Nevertheless, it seems that in our experimental setup, flagella number does not play a strong role in swimming behavior, excepting the obvious difference between zero and one or more flagella.

The fact that we did not find a strong relationship between flagella number and swimming behavior supports other explanations of peritrichous flagellation, such as offspring inheritance, attachment, immunomodulation, etc. Though, we note that these alternative explanations would still have been worth investigating even if we had found a strong link to swimming performance, as natural selection acts on the fitness of the cell as a whole. We decided to focus on offspring inheritance, as it is one of the simplest viable explanations, and is amenable to computer simulation.

Assuming cells do not receive a direct benefit from possessing more than one flagellum, our simulations show that the optimal number of flagella grows logarithmically as a function of the reward/cost ratio. We also found the logarithmic fit robust to variations in how peptidoglycan was inserted into the cell wall. The logarithmic nature of our model implies that the optimal number of flagella first grows quickly at low reward/cost ratios, then slows down as the reward/cost ratio becomes large (Figure 3.4C). This leads us to the conclusion that cells with many flagella (e.g. *B. subtilis*) violate our model assumptions. Otherwise,

our model implies that *B. subtilis* would be experiencing a flagellar reward/cost ratio on the order of  $2^{20}$  (one million), which seems implausibly high. A more likely explanation is that highly flagellated cells derive direct benefits from possessing more than one flagellum, perhaps for swarming or attachment purposes [121].

Our work demonstrates the benefits of the flagellar localization processes found in some bacteria. Flagellar localization ensures that both offspring cells receive one flagellum, and so requires only one new flagellum be produced per division. Our logarithmic model of optimal flagella number has no upper bound, indicating that even cells with four to five flagella per cell will produce some unflagellated offspring. Essentially, peritrichously flagellated cells producing four to five flagella pay 2-3% more than cells with localized flagella, while still not being certain that both offspring will receive at least one flagellum. This trade-off seems of dubious value, unless cells either directly benefit from peritrichous flagellation for another reason or have not evolved flagellar localization machinery due to evolutionary constraints. Perhaps regulatory machinery is also metabolically expensive, flagellation has recently been (re)introduced into a lineage and insufficient time has passed to acquire localization machinery, or evolutionary obstacles (mutation supply, antagonistic pleiotropy, etc.) hinder the evolution of regulatory machinery.

Future work involves better characterizing the role of flagella number in swimming behavior, and evolutionary modeling of flagella number in different environmental contexts. Additionally, studies into other non-swimming uses of peritrichous flagella will provide valuable insight about potentially diverse mechanisms of selection. In the end, we suspect that peritrichous flagellation is maintained via several concurrent mechanisms, each of which contributes to the overall selective pressure to maintain expression of multiple flagella.

## Materials and Methods

We performed all biological experiments using *Salmonella enterica* serovar *typhimurium* LT2. We used strain TH9677 (Dhin-5717::FRT flgE6506(S171C)) as our “wild-type” reference strain, as it allows for maleimide-fluorophore staining of the flagellar hook and dis-

ables flagellar phase variation. We then created EM8242 (Dhin-5717::FRT flgE6506 (S171C) PflhDC5451::Tn10dTc[del-25] DTetA4(Daa6-396)), which allowed us to control flagellar expression using anhydrotetracycline (AnTc) induction of a pTet-*flhDC* promoter.

We incubated all cultures in Vogel-Bonner Minimal Medium E salts buffer containing 2% w/v yeast extract (VB2YE). We grew cells overnight in VB2YE broth at 32°C with shaking, then performed a 1:1000 dilution of overnight culture into fresh inducer-containing media, then incubated under the same conditions for another five hours to condition the cultures. We measured the optical density (OD) of the conditioned cultures at a wavelength of 590nm using a spectrophotometer, then diluted to a target OD of 0.002 in fresh inducer-containing media. We then transferred 200uL of inoculated culture to wells on a 96-well plate (CELLSTAR). Once filled, we secured the lid with parafilm then poked four ventilation holes in the corners of the lid using a hot 18-gauge needle. We then transferred the plate to a Tecan (Austria) Sunrise 96-well plate reader, which incubated the plate at 32°C with shaking. The plate reader collected OD readings every 5-15 minutes for 8-24 hours during the experiment.

OD readings are transformed to account for non-linearity in the spectrophotometer readings before trimming. To limit our fit to regions of exponential growth, we trimmed sequences by keeping only points in which both  $\frac{dOD}{dt}$  and  $\frac{d^2OD}{dt^2}$  were positive. We then used the trimmed data to fit an exponential growth model via maximum likelihood to estimate the growth rate.

When counting foci, we incubate the cells using the same protocol as the growth curves, with an additional five-hour incubation at 32°C with shaking and induction. After the final incubation we centrifuge 500uL of culture at 2500g for five minutes, then resuspend in Vogel-Bonner salts buffer (VB2YE without yeast extract) as a washing step. We repeat the wash, then add AlexaFluor594-maleimide conjugate (ThermoFisher) at a concentration of 50-100uM and incubate with shaking for 10 minutes at room temperature. To remove excess stain, we wash the cells three times as before. To immobilize the cells in a single plane of focus we place 2uL of stained cells onto 1% agarose pads before imaging. We collect 61

images per slide, manually discarding images that have obvious defects such as dust, bubbles, excessively high or low cell density, etc. We use a customized version of SuperSegger [164] to detect and segment cells, along with a custom written foci-detection function to count foci within each cell.

To calculate the cost of flagellar expression, we wrote a model in the STAN probabilistic programming language to fit a mixed-effect model to both the foci count and growth rate data with a linear model to measure the cost of flagellar expression. We used a negative binomial model for the foci count data, and a lognormal model for the growth rate data. We then fit a linear model to determine the slope and intercept of growth rate as a function of foci number.

The FAST assay was performed as previously described [45].

We used an agent-based model (ABM) written in NetLogo [184] to model flagellar inheritance. The model consists of a world inhabited by simulated bacterial cells. Gravity pulls cells towards the bottom of the simulation at a constant rate. Cells with at least one flagellum swim upwards as if they were chemotaxing along a linear attractant gradient, resulting in a drift velocity inversely proportional to nutrient concentration [44]. The combination of gravity and chemotaxis result in cells forming a band within the world, and the user can change the location of the band by tuning the proportionality constants of gravity and chemotaxis. All cells receive a basal amount of energy for existing, and an additional energy reward as a linear function of their height in the world. Together, this means that cells which produce and maintain at least one flagellum are rewarded with additional energy compared to unflagellated cells. There is no additional energy reward for possessing more than one flagellum.

Energy collected by the cells is partitioned between growth and flagellar synthesis via a cell-owned parameter,  $\pi$ . Each cell possesses its own value for  $\pi$ , which it inherited from its parent with a small change in value (i.e. a mutation). The value of  $\pi$  influences the survival and reproduction of the cell, so natural selection will result in the average population

$\pi$  shifting towards the optimal value for a given combination of simulation parameters.

When the amount of energy allocated to flagellar synthesis passes a user-selectable cost threshold, the cell synthesizes a new flagellum which is placed at a random position along the medial axis of the cell. Meanwhile, energy allocated to growth causes the cell to increase in length according to a user-selected growth model (see below), and when the length passes a user-set threshold the cell divides at the midpoint into two offspring cells. To model the passage of flagella to offspring cells, the user can select from three different growth models that control how peptidoglycan is inserted into the cell wall: mid-growth, modeling a situation in which peptidoglycan is inserted only at the cell midpoint; end-growth, which simulates peptidoglycan insertion only at one pole, and random growth, in which the inserted peptidoglycan is inserted randomly along the cell body on each simulation update. The coordinates of each flagellum are adjusted as the cell grows according to the growth method. During division daughter cells are formed from each half of the parent cell, preserving the relative positions the inherited flagella.

## Supporting Information

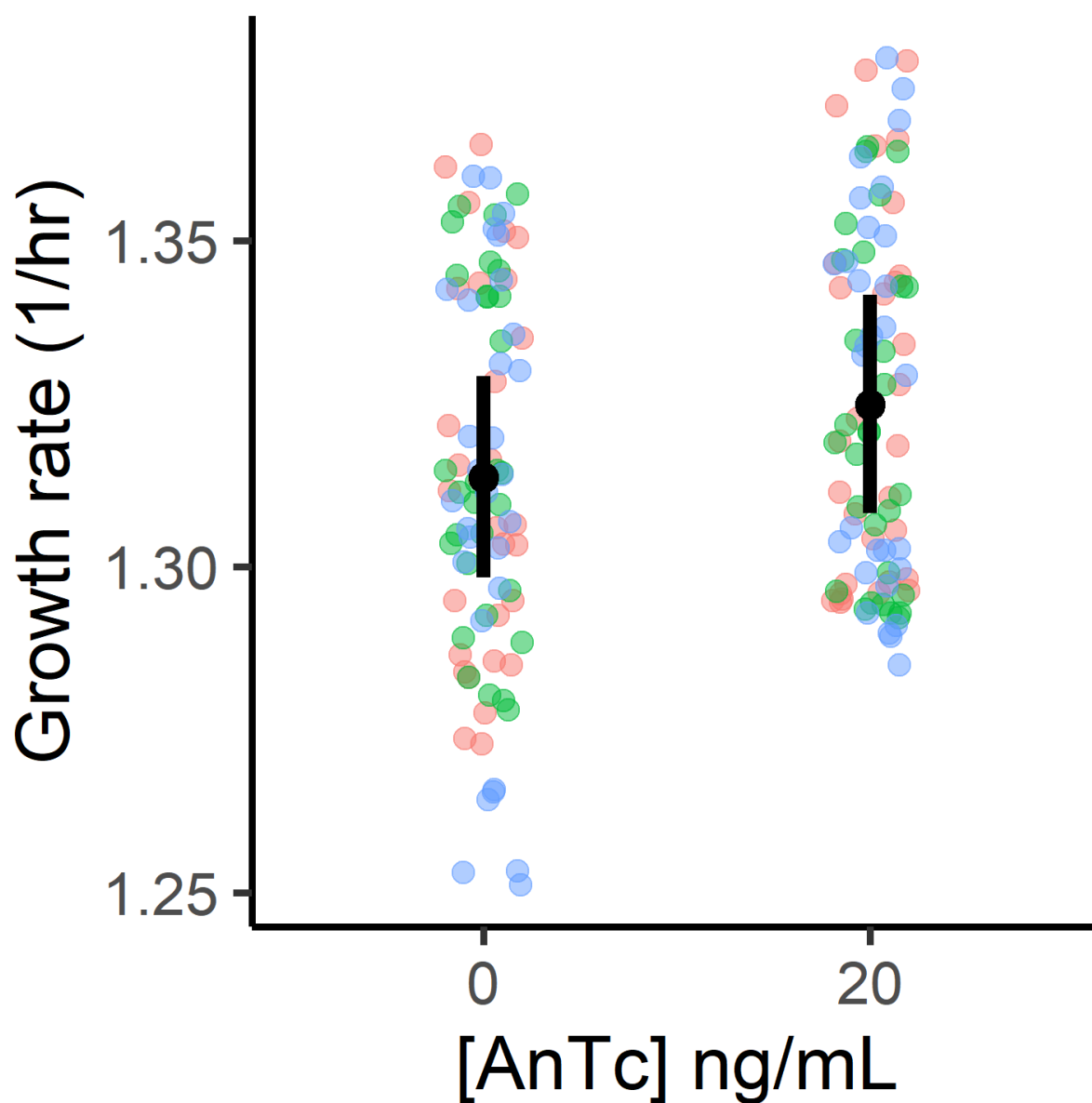


Figure 3.6: **Effect of AnTc on the growth rate of TH9677.** Each dot represents a single batch culture, and each color represents an independent experiment. The estimated growth rate for each condition is indicated in black (dot: mean, line: 95% CrI). The addition of AnTc had a very small positive effect on growth ( $5.6\text{e-}4 \text{ hr}^{-1}(\text{ng/ml})^{-1}$  [ $2.3\text{e-}4, 8.9\text{e-}4$ ] 95% CrI).

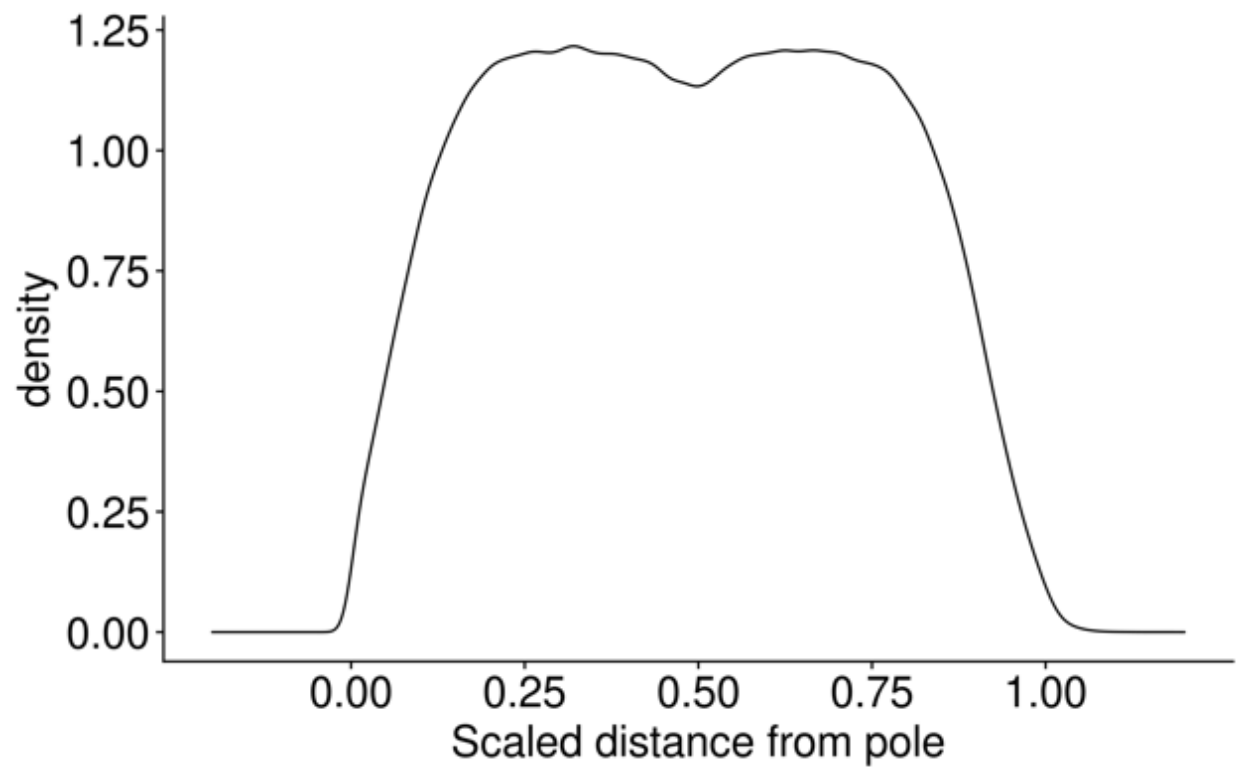


Figure 3.7: **Relative distribution of foci along the cell midline.** Most foci are located between the cell poles, with a nearly even distribution from 0.2 to 0.8, and a slight dip at the cell midpoint. The pole and midpoint may be crowded with other cellular machinery (eg. FtsZ complex).

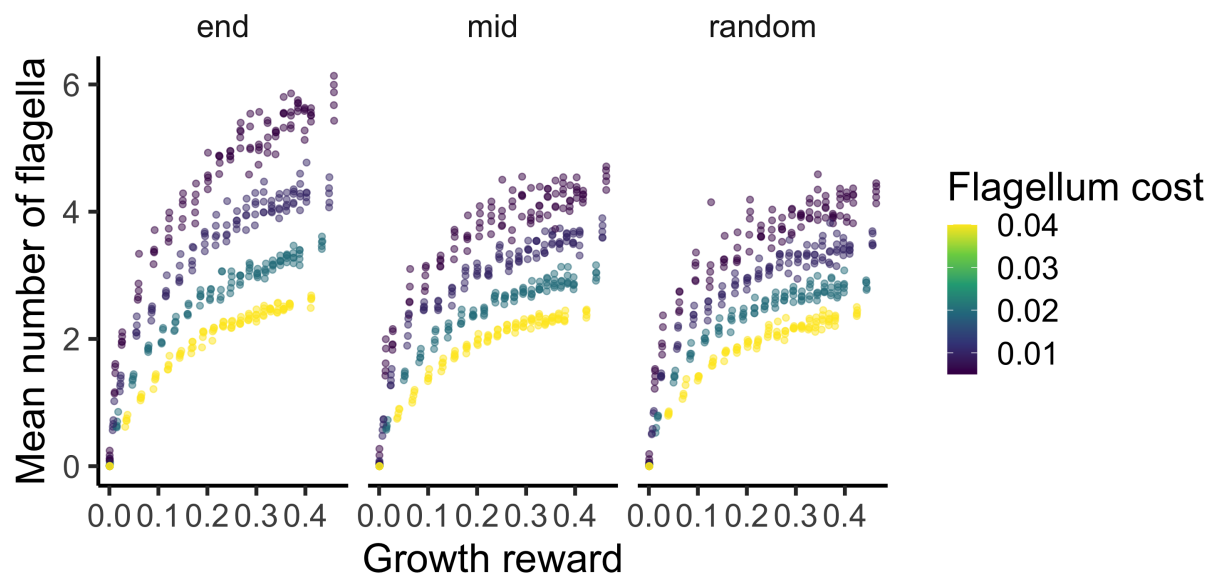


Figure 3.8: **Optimal number of flagella for different combinations of flagellum costs and rewards.** Results from repeated *in silico* evolutionary experiments with different combinations of flagellar cost (color scale) and growth reward from chemotaxis. Each dot represents the mean number of flagella per cell in an independent population after its evolutionary trajectory reached steady state. Simulations were repeated using three growth models for new cell wall synthesis and flagella allocation between daughter cells (end: growth near cell pole, mid: growth from middle of cell length, random: diffusive growth along cell length).



## **CHAPTER 4 - CONCLUSIONS AND FUTURE DIRECTIONS**

Author: Joshua L. Franklin

## Introduction

Each of the previous chapters describe their contributions to biological knowledge. Here, I will emphasize how the computational and statistical models I used in the previous chapters were essential for the success of each project. In all cases, the chapter would have been impossible, or would have been radically different, had we not utilized computational and statistical modeling. Additionally, I will take this chapter as an opportunity to discuss a couple of future directions more freely than in most scientific articles, which I hope the reader finds engaging and interesting without becoming bogged down in details.

## Chapter 1

### Benefits of using Avida as an evolutionary model

In the first chapter we built upon the Survival of the Flattest (SOF) paper, which demonstrated slowly reproducing but mutationally robust organisms out-competing quickly reproducing but mutationally fragile organisms [189]. Such processes are relevant to RNA viruses, which tend to have high mutation rates. Indeed, the effect itself has been confirmed in RNA viruses, providing support for Avida as a useful evolutionary model of important biological systems [33, 101, 150].

In order to replicate the original experiment, we needed to generate mutationally robust and mutationally fragile organisms, which required precise control over mutation rates [189]. Additionally, we needed to perform competition assays across a range of mutation rates to ensure the SOF effect was occurring. While researchers can manipulate biological mutation rates, tuning a culture to a specific mutation rate would be difficult [165]. Additionally, it takes considerable time and effort to adapt biological organisms to new mutation rates, and there is no guarantee that organisms will not evolve reduced mutation rates [165]. In contrast, Avida allowed us to fix the mutation rate to a known value while also disabling insertion and deletion mutations, which we did not of interest in this study. Easy control over the mutation rate in Avida allowed us to examine 201 independent pairs of organisms,

a degree of replication infeasible in many complex biological experiments.

However, the most important advantage of Avida was the ability to directly quantify the distribution of fitness effects (DFE) of our organisms. Basically, the DFE required us to categorize each mutation as lethal, deleterious, neutral, and beneficial relative to the reference genotype. Avida can easily generate and characterize the fitness of all single- and double- mutants for a genome. However, even Avida cannot test all triple- or higher mutants due to the vast number of possible combinations of mutations. Fortunately, Avida can still precisely estimate the DFE at higher mutational distances by randomly sampling tens of millions of combinations for a given mutational distance. In a biological system, where the smallest viral genomes are about 1 kilobase [26], generating and characterizing every single nucleotide change would be possible, if laborious. With effort, a researcher could generate a library of all 4.5 million possible double mutants, but the competitions assays required to categorize each mutant would be infeasible.

Instead, a researcher might try to characterize fitness by measuring the change in the relative frequency of genotypes before and after a round of growth, similar to a TnSeq experiment [174]. However, this kind of assay has more noise than Avida, making it difficult to differentiate between e.g. neutral and slightly deleterious mutations, which may be relevant if mutational robustness and drift robustness are related phenomena [55,98]. Avida can classify the effect of each mutation as lethal, deleterious, neutral, or beneficial by directly measuring the fitness of a mutant. While it is possible for organisms in Avida to have probabilistic phenotypes, it is not common with the settings we used in our experiment. In biological systems fitness measurements must contend with biological variability and other limitations of competition assays (e.g. “neutral” markers). A growth rate approach to measuring fitness, as discussed in chapter 3, could potentially be easier and more accurate. Even so, performing millions of growth curves with substantial replication is far more difficult than using a few hundred hours of CPU time.

## Future directions

We encountered some unusual discrepancies as we tried to replicate the original SOF paper. We had to use a spatially structured environment to generate the SOF effect, rather than the well-mixed environment used in the original SOF study. When investigating, we found that the version of Avida used in the original study possessed a slightly different instruction set [127, 189]. The original instruction set was less flexible in that some of the instructions had order dependencies. For example, “get” and “put” which load a value and send a value, respectively, perform different tasks and so require a particular ordering. Whereas the modern instruction set replaces “get” and “put” with a unified “IO” instruction that both sends the current value then loads a new value automatically [127]. An analogous change has occurred with some of the reproduction instructions as well [127]. The changes from the old to new instruction sets were likely made to increase evolvability by making the instruction set more “robust.”

Instruction sets of varying “robustness” provide a unique opportunity to explore biochemical transitions that may have occurred in early lifeforms. Many scientists hypothesize that life originally used RNA to both store information and catalyze reactions, before transitioning to DNA-based information storage and protein catalysts [10]. It seems possible that such a transition would reduce the mutation rate, due to the higher chemical stability of DNA [10]. However, the transition could also cause radical changes in mutational robustness. If RNA with substantial secondary structure or catalytic activity transitioned into the double-stranded DNA we see today, the DFE would probably change substantially.

We might be able to emulate some aspects of the RNA to DNA transition by evolving organisms in Avida with a “fragile” instruction set, then transitioning the organisms to a semi-compatible “robust” instruction set. We would expect to see radical changes in the DFE and rapid changes in fitness during adaptation. If it is difficult to perform this transition without killing the organisms, it may be possible to more gently transition from one instruction set to another using an intermediate instruction set containing both robust and fragile instructions.

A study of the RNA to DNA transition using Avida would have the limitations frequently associated with computational models. Instruction sets in Avida are imperfect analogies to RNA and DNA. However, researchers have used Avida for several interesting studies that have been validated experimentally, such as the original SOF paper [189]. Additionally, Avida has already been used to investigate aspects of another difficult to study transition: the origin of life [23]. Despite the limitations, I believe that experiments with Avida instruction sets can provide insights into the otherwise opaque RNA to DNA transition.

## Chapter 2

### Benefits of using generalized additive models

In the second chapter we investigated fruiting body formation in *Myxococcus xanthus*. During starvation, *M. xanthus* cells cooperate to form a fruiting body, which requires the differentiation of rod cells into spores [93]. The signaling processes controlling sporulation are driven by contact-dependent C-signaling [93], so we wanted to characterize the spatial distribution of gene expression, cell alignment, and cell classification. Ideally, we would describe the spatial structure of these parameters throughout the fruiting body using a smooth, continuous function. Additionally, due to the substantial variability between fruiting bodies, our model would be capable of incorporating information from multiple replicates to accurately reflect this variability. Generalized additive models (GAM) satisfied all these requirements, and it is easy to write Bayesian GAM models using the *brms* R software package [22, 192].

GAMs are a principled way of fitting a flexible curve to a dataset [192]. A GAM consists of a series of basis functions that are combined to generate a single curve that is flexible enough to fit a dataset without becoming “too wiggly” and overfitting noise within the data [192]. These features make GAMs useful for interpolation, where we do not have, or are not concerned with, a specific equation describing the data across some variable such as space or time. For example, we wanted to know the fluorescent intensity across space within a fruiting body, and we suspected that factors such as cell classification would affect

intensity. However, we did not have a specific mathematical equation describing the change in fluorescent intensity as a function of distance from the center of the fruiting body, such as exponential decay. Instead, by fitting a GAM we can generate a smooth curve that describes intensity across space while accounting for relevant variables, but without forcing the curve to have a specific shape.

The flexibility of brms allowed us to use very similar GAMs for continuous variables, such as alignment and fluorescent intensity, and discrete variables, such as cell classification. This is because brms allows the user to specify the response function of a linear model [22]. Essentially, this means that instead a Gaussian distribution around the mean, as in a linear model, we can choose to use an arbitrary probability distribution, such as gamma, Poisson, categorical, etc [22]. So, we can keep similar model structures in terms of explanatory variables, and simply change the response function from Gaussian to categorical when switching from a continuous variable, e.g. alignment, to one described by a categorical distribution, e.g. cell classification. While brms is typically used to fit linear and generalized linear models, it gives users the capability to use GAMs as a “drop-in” replacement, enabling us to easily extend linear models across space, time, or any other variable of interest.

## Chapter 3

### Benefits of using Mixed-effect Models

In chapter three we were interested in precisely quantifying the cost of flagella in order to characterize the cost-benefit trade-offs associated with peritrichous flagellation. We chose to measure the cost using 96-well plate optical density growth curves, which can be noisy and difficult to fit. Biochemical and evolutionary loss-of-function estimates put the cost per flagellum on the order of 1% [138], so we required substantial precision to even detect the growth defect.

To increase the precision of our estimates, we performed many replicate assays. However, we did not simply average our growth rates, because this is not a good model of how the data was generated [112]. This is because many biological assays produce structured data, e.g.

wells are nested within plates which themselves are nested in days. A simplistic view is that wells within plate are isolated from one another and so they are completely independent. However, this is not an accurate view of the situation; wells within a plate often share some sources of noise. For example, we diluted the chemical inducer on a per-plate basis, so all wells on a plate shared the same dilution error. Each plate was sourced from different colonies on an agar plate, so biological variation is shared by all the wells within a plate. We performed incubations on a per-plate basis, so errors from timing or temperature are also shared by all wells on the plate. The result is that wells within a plate need to be treated as related entities.

Additionally, higher-level sources of error such as pipette calibration, ambient temperature, relative humidity, etc. vary from day to day, so multiple plates performed on the same day are not fully independent either. We could take this farther and say that pipetting precision, preparedness, and laboratory skills vary widely from person-to-person, so plates performed by different people need to be grouped together. Similar arguments could be made that laboratory-to-laboratory variation in climate, equipment manufacturers, water sources, etc. mean that we should also take laboratory into account. While we could continue this chain of noise, beyond a certain point it is impractical to e.g. replicate across people and across labs.

Mixed-effect models allowed us to accurately model hierarchical data produced by our growth curves. We grouped our growth rate measurements by plate, which also accounts for the date because we performed one plate per day. The model then estimated a plate effect/deviation for every plate we performed as being drawn from a normal (or t-) distribution with some variance. At a high level, all plates contribute to the posterior distribution (read: estimate) of growth rate, but plates that deviate strongly from the mean are “shrunk” towards the mean in a mathematically principled way [112]. This reduces our tendency to overfit compared to doing no averaging and reduces our tendency to underfit compared to computing a global average [112]. There are also side benefits, such as the “free” posterior dis-

tribution of the plate effects in our study. While we did not utilize that distribution, other researchers may be interested in the distribution of effects coming from dates, personnel, countries, reagents, etc.

### **Future directions**

Our growth curve results indicated substantial cost of producing flagella but did not find a strong relationship between flagella number and swimming behavior. Instead, we postulated that flagellar inheritance might explain the production of multiple peritrichous flagella. Our agent-based model of flagella assumed that a single flagellum was sufficient for maximum swimming performance, and that cells paid a constant cost per flagellum produced. We found that populations of agents evolved different flagellar expression levels depending on the reward for being flagellated, the cost of producing flagella, and the method of peptidoglycan insertion, which affects the distribution of flagella along the cell.

The optimal number of flagella in our model is a logarithmic function of the benefit/cost ratio, implying that an optimal number of flagella greater than six or seven requires high benefit/cost ratios. That is why we suggested that organisms like *Bacillus subtilis* (over 20 flagella) violate our model assumptions, as it seems implausible that the reward/cost ratio is of the order ten million. However, our model of flagellar inheritance is consistent with common enteric bacteria such as *Escherichia coli* and *Salmonella enterica*, which have about four flagella per cell. This begs the question of whether these enteric bacteria experience selective pressures consistent with our model and have consequently optimized their flagellar expression for inheritance purposes. If we could design an environment consistent with our model, we could perform an evolution experiment to test our predictions.

Key to such an experiment is designing an environment with the intended selective pressures. The environment needs to reward swimming and chemotaxing cells, without benefiting nonmotile behaviors such as buoyancy and biofilm production. This clearly rules out shaken media, which causes cells to rapidly lose flagellation, as they pay a large cost and reap no benefit [138]. Instead, shaken media could serve as a control environment to determine



the rate of flagella loss when there is zero benefit of flagella [138]. Serial passage in agar media would select for motility, but that is a high viscosity and porous medium that is outside the scope of our flagellar inheritance model and has been explored in other studies [123]. These studies indicate upregulated flagellar expression in *E. coli*, which implies that agar violates our model assumptions, perhaps because of the higher viscosity [123]. Unshaken liquid media could encourage swimming along a vertical oxygen gradient, but can also favor lineage differentiation due to the spatially structured nature of the environment [138]. Instead, we need to design a more specialized environment that continuously produces the required selective pressure for motility.

To produce such an environment, I have helped design a bacterial “racetrack” microfluidic device. The racetracks consist of a single chamber about 150 $\mu$ m wide, 150 $\mu$ m deep, and about two meters long (Figure 4.1). The racetrack spirals outwards from a central loading chamber until the diameter of the device reaches about 100mm, at which point it merges with an unloading chamber. The result is a racetrack that provides sustained selection for motility, as cells need to continuously swim to access fresh nutrients. The device can optionally utilize chemotaxis, depending on the presence of metabolizable chemoattractants in the media. At a drift velocity of 5 $\mu$ m/s, it would take four to five days for cells to complete the “race”. Once the cells reached the unloading chamber, we could use a syringe to transfer cells to a new racetrack for a new round of selection.

If our racetracks correlate with our agent-based model, we should expect the population to evolve towards the optimal flagella number due to flagellar inheritance. If flagella number reaches a new steady expression level, we can then derive the benefit/cost ratio of producing flagella for that organism. Perhaps most interestingly, if we see no meaningful change in flagella number of wildtype cells over prolonged periods of selection, that could indicate the organism has a life history that is well-aligned with our model.

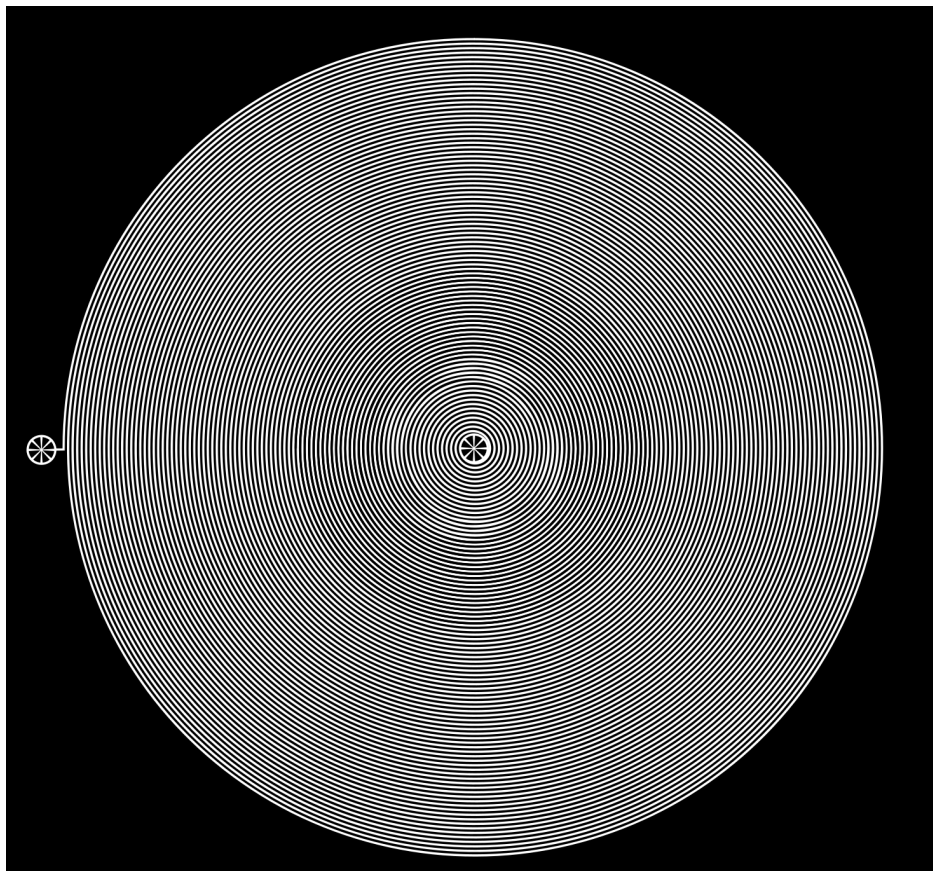


Figure 4.1: **A bacterial “racetrack”.** The width and depth of the channel is about 150 $\mu$ m. The total length of the track is approximately two meters long.

## Closing recommendations

### Embrace agent-based modeling

As my work has emphasized, agent-based modeling (ABM) should be in the toolset of most biologists. ABMs consist of entities (agents) following a set of rules. An ABM of bacterial chemotaxis might allow entities to swim through their environment or tumble, and transition between those states based on the concentration of nutrients. Or in an ABM of *M. xanthus* fruiting body development, entities might possess an accumulator variable that integrates their exposure to a particular signal and change their behavior once the accumulator crosses a threshold. In both cases, the rules are fairly simple and can be inferred from biological knowledge. We can then set our computers to run the model many times

and record the output. We can then observe how the model output changes when tweaking parameters of interest (e.g. swimming speed).

Certainly, agent-based models can be much more complex than these examples. However, they can be useful to flip the script on working “smarter not harder”, by having the computer work harder so that you look smarter. The explosion of computational resources over the past few decades have made running hundreds of permutations of complex agent-based models easy. Additionally, the ABM approach is arguably easier for the average biologist than attempting to derive the (partial) differential equations governing these phenomena. Indeed, without our ABM of flagellar inheritance in chapter three, we may not have found our equation-based model of the optimal number of flagella.

### **Prefer a Bayesian Approach**

Perhaps the most appealing argument for Bayesian statistics is that many scientists already try to interpret frequentist statistics in a Bayesian manner. A classic example of this behavior relates to confidence intervals. A confidence interval is a frequentist interval calculated such that, if an experiment were repeated many times, the calculated interval would overlap the true value of a parameter some percentage of the time (usually 95%) [66]. However, scientists often try to interpret a confidence interval as containing the true value of a parameter with some probability (again, usually 95%). This is nonsense from a frequentist point of view, in which parameters are treated as fixed; the confidence interval either contains the true value or does not [66]. However, Bayesian credible intervals, *conditional on the likelihood, data, and prior distribution*, do contain the true value of a parameter with a specified degree of probability [57]. By shifting to a Bayesian approach, scientists can more confidently follow their intuition when interpreting statistics.

More fundamentally, we are often actually interested in *the value of parameters*, not rejecting (often arbitrary) null hypotheses. For example, imagine a farmer asking a scientist to test the effect of fertilizer treatment on potato yields. The farmer probably would not ask if the scientist had rejected the hypothesis that the fertilizer had no effect. Instead,

the farmer would probably ask *how much* the fertilizer increased yields. This indicates that the farmer is interested in the value of the yield parameter, not rejecting a null hypothesis. Tellingly, it also indicates that the farmer has a prior distribution on the value of the yield parameter due to their prior experience with fertilizer. Bayesian statistics puts the focus on the posterior distribution of the parameters, making interpretation easier for both scientists and the general public.

### **Make computational models accessible for educational purposes**

Computational models allow biologists to address questions that are otherwise difficult to answer. Additionally, they make excellent pedagogical tools because they allow students to play with a subject without the expense, labor, and safety issues of laboratories [16]. Part of my instructional experience during graduate school involved helping to write web-based computational models of diffusion, bacterial motility, and chemotaxis for 8th grade students as part of MSU's Gifted and Talented Education summer program. The students enjoyed the interactivity of the simulations, with some students going to far as to edit the JavaScript code backing the simulation. Additionally, the easy data-logging capabilities of computational models allowed us to perform simple quantitative analyses in class. Overall, it was an excellent opportunity to integrate microbiology, physics, math, and programming into a compelling educational experience.

While the models we used in our classroom were built specifically for education, I encourage researchers to facilitate the use of their research models as educational tools by making them as accessible as possible. To do so, articles using the model should be open access and model code should be open source. However, to make a model highly accessible for educational use requires more effort. Specifically, I suggest the following:

- Documentation is key
  - Fully document every function, parameter, setting, flag, and output
  - Ensure that all provided documentation is accurate and always up to date

- Target your documentation towards an audience of modest technical and scientific knowledge
- Make installation painless
  - Preferably, use a web-based platform or provide pre-built binaries
  - Alternatively, provide build scripts
  - At a minimum, provide comprehensive installation instructions
  - Minimize external dependencies and avoid complex build tools
- Provide interesting examples
  - Write tutorials guiding users through important aspects of the model
  - Design example experiments for users to build upon
  - Consider providing videos or blog posts to interest users
- Be visual and interactive
  - Graphics are particularly engaging
  - Make it easy to save videos and images of the model
  - If appropriate, include touch- and mouse-based controls
- Make real experiments easy
  - Dump data into easily viewed table formats such as CSV, even if only a subset of your output can be saved in this way
  - Avoid using binary, JSON, XML, etc. files as the only output format
- Be genuine
  - Avoid the temptation to write a simplified model for educational use

- Instead, handle complexity through good documentation and software design practices

These recommendations pose a substantial cost, but I suggest that the benefits are larger. Educational tools that organically engage students are intrinsically beneficial and can be used to fulfill grant outreach and engagement requirements. Less obviously, following these recommendations may result in models that are more scientifically useful because they are also more accessible *to other scientists*. Improved accessibility to scientific audiences may also improve reproducibility and encourage model extension/collaboration. Taken together, there is a clear case for making accessibility a higher priority for computational models in the sciences.

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- [1] Christoph Adami, Charles Ofria, and Travis C Collier. Evolution of biological complexity. *Proceedings of the National Academy of Sciences*, 97(9):4463–4468, 2000.
- [2] Aneil F Agrawal and Michael C Whitlock. Mutation load: The fitness of individuals in populations where deleterious alleles are abundant. *Annual Review of Ecology, Evolution, and Systematics*, 43:115–135, 2012.
- [3] Phillip Aldridge and Kelly T Hughes. Regulation of flagellar assembly. *Current Opinion in Microbiology*, 5(2):160–165, 2002.
- [4] M Archetti. Survival of the steepest: Hypersensitivity to mutations as an adaptation to soft selection. *Journal of Evolutionary Biology*, 22(4):740–750, 2009.
- [5] T Atsumi, Y Maekawa, T Yamada, I Kawagishi, Y Imae, and M Homma. Effect of viscosity on swimming by the lateral and polar flagella of vibrio alginolyticus. *Journal of Bacteriology*, 178(16):5024–5026, 1996.
- [6] Rajesh Balagam and Oleg A. Igoshin. Mechanism for collective cell alignment in *Myxococcus xanthus* bacteria. *PLoS Computational Biology*, 11, 8 2015.
- [7] Robert E Beardmore, Ivana Gudelj, David A Lipson, and Laurence D Hurst. Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature*, 472(7343):342–346, 2011.
- [8] Robert Belas. Biofilms, flagella, and mechanosensing of surfaces by bacteria. *Trends in Microbiology*, 22(9):517–527, 2014.
- [9] Howard C. Berg. The rotary motor of bacterial flagella. *Annual Review of Biochemistry*, 72(1):19–54, 2003. PMID: 12500982.
- [10] Harold S. Bernhardt. The rna world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biology Direct*, 7(1):23, Jul 2012.
- [11] Swapna Bhat, Tye O. Boynton, Dan Pham, and Lawrence J. Shimkets. Fatty acids from membrane lipids become incorporated into lipid bodies during *Myxococcus xanthus* differentiation. *PLOS ONE*, 9(6):1–10, 06 2014.
- [12] Saurabh Bhattacharya, Amit K. Baidya, Ritesh Ranjan Pal, Gideon Mamou, Yair E. Gatt, Hanah Margalit, Ilan Rosenshine, and Sigal Ben-Yehuda. A ubiquitous platform for bacterial nanotube biogenesis. *Cell articles*, 27:334–342.e10, 4 2019.
- [13] Christof Biebricher and Manfred Eigen. What is a quasispecies? *Quasispecies: Concept and Implications for Virology*, pages 1–31, 2006.
- [14] Tye O’Hara Boynton and Lawrence Joseph Shimkets. *Myxococcus csga*, *drosophila* sniffer, and human hsd10 are cardiolipin phospholipases. *Genes and Development*, 29:1903–1914, 9 2015.



- [15] Sarah J. Bray. Notch signalling in context. *Nature Reviews Molecular Cell Biology*, 17:722–735, 11 2016.
- [16] Elena Bray Speth, Tammy M. Long, Robert T. Pennock, and Diane Ebert-May. Using aida-ed for teaching and learning about evolution in undergraduate introductory biology courses. *Evolution: Education and Outreach*, 2(3):415–428, Sep 2009.
- [17] James Briscoe and Pascal P. Thérond. The mechanisms of hedgehog signalling and its roles in development and disease. *Nature Reviews Molecular Cell Biology*, 14:418–431, 2013.
- [18] Nhat Khai Bui, Joe Gray, Heinz Schwarz, Peter Schumann, Didier Blanot, and Walde-mar Vollmer. The peptidoglycan sacculus of *Myxococcus xanthus* has unusual structural features and is degraded during glycerol-induced myxospore development. *Journal of Bacteriology*, 191(2):494–505, 2009.
- [19] James J Bull, Lauren Ancel Meyers, and Michael Lachmann. Quasispecies made simple. *PLoS Computational Biology*, 1(6):e61, 2005.
- [20] Matthew J. Bush, Natalia Tschowri, Susan Schlimpert, Klas Flärdh, and Mark J. Buttner. C-di-gmp signalling and the regulation of developmental transitions in streptomycetes. *Nature Reviews Microbiology*, 13:749–760, 12 2015.
- [21] David Butcher. Muller’s ratchet, epistasis and mutation effects. *Genetics*, 141(1):431–437, 1995.
- [22] Paul-Christian Bürkner. Advanced bayesian multilevel modeling with the r package brms. *The R Journal*, 10(1):395–411, 2018.
- [23] Nitash C G, Thomas LaBar, Arend Hintze, and Christoph Adami. Origin of life in a digital microcosm. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2109):20160350, 2017.
- [24] Todd A. Cameron, John R. Zupan, and Patricia C. Zambryski. The essential features and modes of bacterial polar growth. *Trends in Microbiology*, 23(6):347–353, Jun 2015.
- [25] Ashleigh Campbell, Poorna Viswanathan, Terry Barrett, Bongjun Son, Shreya Saha, and Lee Kroos. Combinatorial regulation of the *dev* operon by mrpc2 and frua during *Myxococcus xanthus* development. *Journal of Bacteriology*, 197:240–251, 2015.
- [26] José A. Campillo-Balderas, Antonio Lazcano, and Arturo Becerra. Viral genome size distribution does not correlate with the antiquity of the host lineages. *Frontiers in Ecology and Evolution*, 3:143, 2015.
- [27] Felipe Cava, Erkin Kuru, Yves V Brun, and Miguel A de Pedro. Modes of cell wall growth differentiation in rod-shaped bacteria. *Current Opinion in Microbiology*, 16(6):731–737, 2013. Growth and development: eukaryotes/prokaryotes.

- [28] Bonnie Chaban, H. Velocity Hughes, and Morgan Beeby. The flagellum in bacterial pathogens: For motility and a whole lot more. *Seminars in Cell & Developmental Biology*, 46:91–103, 2015. Biom mineralisation & Motorisation of pathogens.
- [29] Fabienne F. V. Chevance and Kelly T. Hughes. Coordinating assembly of a bacterial macromolecular machine. *Nature Reviews Microbiology*, 6(6):455–465, Jun 2008.
- [30] Dennis Claessen, Daniel E. Rozen, Oscar P. Kuipers, Lotte Søgaaard-Andersen, and Gilles P. Van Wezel. Bacterial solutions to multicellularity: A tale of biofilms, filaments and fruiting bodies. *Nature Reviews Microbiology*, 12:115–124, 2 2014.
- [31] Hans Clevers and Roel Nusse. Wnt/b-catenin signaling and disease. *Cell*, 149:1192–1205, 6 2012.
- [32] Alma Dal Co, Simon van Vliet, Daniel Johannes Kiviet, Susan Schlegel, and Martin Ackermann. Short-range interactions govern the dynamics and functions of microbial communities. *Nature Ecology and Evolution*, 4:366–375, 3 2020.
- [33] Francisco M Codoñer, José-Antonio Darós, Ricard V Solé, and Santiago F Elena. The fittest versus the flattest: Experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathogens*, 2(12):e136, 2006.
- [34] Remy Colin, Bin Ni, Leanid Laganenka, and Victor Sourjik. Multiple functions of flagellar motility and chemotaxis in bacterial physiology. *FEMS Microbiology Reviews*, 07 2021. fuab038.
- [35] Iñaki Comas, Andrés Moya, and Fernando González-Candelas. Validating viral quasispecies with digital organisms: A re-examination of the critical mutation rate. *BMC Evolutionary Biology*, 5(1):5, 2005.
- [36] Christopher R. Cotter, Heinz Bernd Schüttler, Oleg A. Igoshin, and Lawrence J. Shimkets. Data-driven modeling reveals cell behaviors controlling self-organization during *Myxococcus xanthus* development. *Proceedings of the National Academy of Sciences of the United States of America*, 114:E4592–E4601, 6 2017.
- [37] Arthur W Covert, Richard E Lenski, Claus O Wilke, and Charles Ofria. Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences*, 110(34):E3171–E3178, 2013.
- [38] Eugene W Crawford and Lawrence J Shimkets. The *Myxococcus xanthus* *socE* and *csgA* genes are regulated by the stringent response. *Molecular Microbiology*, 37:788–799, 2000.
- [39] James F Crow. Some possibilities for measuring selection intensities in man. *Human Biology*, 30(1):1–13, 1958.
- [40] Patrick D. Curtis, Rion G. Taylor, Roy D. Welch, and Lawrence J. Shimkets. Spatial organization of *Myxococcus xanthus* during fruiting body formation. *Journal of Bacteriology*, 189:9126–9130, 12 2007.

- [41] Eric H. Davidson. Emerging properties of animal gene regulatory networks. *Nature*, 468:911–920, 12 2010.
- [42] J Arjan G M de Visser, Joachim Hermisson, Günter P Wagner, Lauren Ancel Meyers, Homayoun Bagheri-Chaichian, Jeffrey L Blanchard, Lin Chao, James M Cheverud, Santiago F Elena, Walter Fontana, et al. Perspective: Evolution and detection of genetic robustness. *Evolution*, 57(9):1959–1972, 2003.
- [43] Edward F. DeLong, Christina M. Preston, Tracy Mincer, Virginia Rich, Steven J. Hallam, Niels-Ulrik Frigaard, Asuncion Martinez, Matthew B. Sullivan, Robert Edwards, Beltran Rodriguez Brito, Sallie W. Chisholm, and David M. Karl. Community genomics among stratified microbial assemblages in the ocean’s interior. *Science*, 311(5760):496–503, 2006.
- [44] Yann S. Dufour, Xiongfei Fu, Luis Hernandez-Nunez, and Thierry Emonet. Limits of feedback control in bacterial chemotaxis. *PLOS Computational Biology*, 10(6):1–11, 06 2014.
- [45] Yann S. Dufour, Sébastien Gillet, Nicholas W. Frankel, Douglas B. Weibel, and Thierry Emonet. Direct correlation between motile behavior and protein abundance in single cells. *PLOS Computational Biology*, 12(9):1–25, 09 2016.
- [46] Jeffrey A Edlund and Christoph Adami. Evolution of robustness in digital organisms. *Artificial Life*, 10(2):167–179, 2004.
- [47] Manfred Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.
- [48] Manfred Eigen and Peter Schuster. A principle of natural self-organization. part a: Emergence of the hypercycle. *Naturwissenschaften*, 64(11):541–565, 1977.
- [49] Santiago F Elena, Purificación Carrasco, José-Antonio Daròs, and Rafael Sanjuán. Mechanisms of genetic robustness in RNA viruses. *EMBO articles*, 7(2):168–173, 2006.
- [50] Eva Ellehauge, Mads Nørregaard-Madsen, and Lotte Søgaaard-Andersen. The frua signal transduction protein provides a checkpoint for the temporal co-ordination of inter-cellular signals in *Myxococcus xanthus* development. *Mol Microbiol*, 30(4):807–817, Nov 1998.
- [51] Marc Erhardt. Energy requirements for protein secretion via the flagellar type III secretion system. In *Methods in Molecular Biology*, pages 449–457. Springer New York, 2017.
- [52] Adam Eyre-Walker and Peter D Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):610–618, 2007.
- [53] Robert Forster, Christoph Adami, and Claus O Wilke. Selection for mutational robustness in finite populations. *Journal of theoretical biology*, 243(2):181–190, 2006.

- [54] David T Fraebel, Harry Mickalide, Diane Schnitkey, Jason Merritt, Thomas E Kuhlman, and Seppe Kuehn. Environment determines evolutionary trajectory in a constrained phenotypic space. *eLife*, 6, March 2017.
- [55] Joshua Franklin, Thomas LaBar, and Christoph Adami. Mapping the peaks: Fitness landscapes of the fittest and the flattest. *Artificial Life*, 25:250–262, 2019.
- [56] Mikako Fujii, Satoshi Shibata, and Shin-Ichi Aizawa. Polar, peritrichous, and lateral flagella belong to three distinguishable flagellar families. *Journal of Molecular Biology*, 379(2):273–283, 2008.
- [57] Samanta Tapas Ghosh Jayanata K., Delampady Mohan. *An Introduction to Bayesian Analysis*, chapter 2, pages 29–63. Springer, New York, NY, 1 edition, 2006. Page 49 specifically.
- [58] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, April 2009.
- [59] Heather J Goldsby, Anna Dornhaus, Benjamin Kerr, and Charles Ofria. Task-switching costs promote the evolution of division of labor and shifts in individuality. *Proceedings of the National Academy of Sciences*, 109(34):13686–13691, 2012.
- [60] Heather J Goldsby, David B Knoester, Charles Ofria, and Benjamin Kerr. The evolutionary origin of somatic cells under the dirty work hypothesis. *PLoS Biology*, 12(5):e1001858, 2014.
- [61] Thomas M A Gronewold and Dale Kaiser. The act operon controls the level and time of c-signal production for *Myxococcus xanthus* development. *Molecular Microbiology*, 40:744–756, 2001.
- [62] Sarah B. Guttenplan and Daniel B. Kearns. Regulation of flagellar motility during biofilm formation. *FEMS Microbiology Reviews*, 37(6):849–871, 11 2013.
- [63] Douglas Hanahan. Studies on transformation of escherichia coli with plasmids. *Journal of Molecular Biology*, 166(4):557–580, 1983.
- [64] Tong Hao, Dvora Biran, Gregory J. Velicer, and Lee Kroos. Identification of the  $\omega$ regulatory region, a developmental promoter of *Myxococcus xanthus* that is transcribed in vitro by the major vegetative rna polymerase. *Journal of Bacteriology*, 184(12):3348–3359, 2002.
- [65] Antonia Herrero, Joel Stavans, and Enrique Flores. The multicellular nature of filamentous heterocyst-forming cyanobacteria. *FEMS Microbiology Reviews*, 40:831–854, 11 2016.
- [66] Rink Hoekstra, Richard D. Morey, Jeffrey N. Rouder, and Eric Jan Wagenmakers. Robust misinterpretation of confidence intervals. *Psychonomic Bulletin and Review*, 21:1157–1164, 10 2014.

- [67] Egbert Hoiczyk, Michael W. Ring, Colleen A. McHugh, Gertrud Schwär, Edna Bode, Daniel Krug, Matthias O. Altmeyer, Jeff Zhiqiang Lu, and Helge B. Bode. Lipid body formation plays a central role in cell fate determination during developmental differentiation of *Myxococcus xanthus*. *Molecular Microbiology*, 74:497–517, 10 2009.
- [68] Colette A. Ten Hove, Kuan Ju Lu, and Dolf Weijers. Building a plant: Cell fate specification in the early arabidopsis embryo. *Development (Cambridge)*, 142:420–430, 2 2015.
- [69] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.
- [70] Antonio A. Iniesta, Francisco García-Heras, Javier Abellón-Ruiz, Aránzazu Gallego-García, and Montserrat Elías-Arnanz. Two systems for conditional gene expression in *myxococcus xanthus* inducible by isopropyl- $\beta$ -thiogalactopyranoside or vanillate. *Journal of Bacteriology*, 194(21):5875–5885, 2012.
- [71] Ken F. Jarrell and Mark J. McBride. The surprisingly diverse ways that prokaryotes move. *Nature Reviews Microbiology*, 6(6):466–476, Jun 2008.
- [72] Lars Jelsbak and Lotte Søgaard-Andersen. Pattern formation by a cell surface-associated morphogen in *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences*, 99(4):2032–2037, 2002.
- [73] Bryan Julien, A. Dale Kaiser, and Anthony Garza. Spatial control of cell differentiation in *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences*, 97(16):9098–9103, 2000.
- [74] D. Kaiser. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences of the United States of America*, 76(11):5952–5956, Nov 1979. PMC411771[pmcid].
- [75] Dale Kaiser. Cell fate and organogenesis in bacteria. *Trends in Genetics*, 15(7):273–277, Jul 1999.
- [76] Dale Kaiser and Roy Welcht. Dynamics of fruiting body morphogenesis. *Journal of Bacteriology*, 186:919–927, 2 2004.
- [77] Kazem Kashefi and Patricia L. Hartzell. Genetic suppression and phenotypic masking of a *Myxococcus xanthus* frz<sup>-</sup> defect. *Molecular Microbiology*, 15(3):483–494, 1995.
- [78] Matthew Kay. *tidybayes: Tidy Data and Geoms for Bayesian Models*, 2020. R package version 2.3.1.
- [79] Daniel B. Kearns. A field guide to bacterial swarming motility. *Nature Reviews Microbiology*, 8(9):634–644, Sep 2010.
- [80] Daniel B. Kearns. You get what you select for: better swarming through more flagella. *Trends in Microbiology*, 21(10):508–509, 2013.

- [81] S K Kim and D Kaiser. Cell motility is required for the transmission of c-factor, an intercellular signal that coordinates fruiting body morphogenesis of *Myxococcus xanthus*. *Genes & Development*, 4(6):896–905, June 1990.
- [82] Seung K Kim and Dale Kaiser. C-factor: A cell-cell signaling protein required for fruiting body morphogenesis of *M. xanthus*. *Cell*, 61:19–26, 1990.
- [83] Seung K Kim and Dale Kaiser. C-factor has distinct aggregation and sporulation thresholds during myxococcus development. *Journal of Bacteriology*, 173:1722–1728, 1991.
- [84] SK Kim and D Kaiser. Cell alignment required in differentiation of *Myxococcus xanthus*. *Science*, 249(4971):926–928, 1990.
- [85] Alexey S Kondrashov. Muller’s ratchet under epistatic selection. *Genetics*, 136(4):1469–1473, 1994.
- [86] R H Kottel, K Bacon, D Clutter, and D White. Coats from *Myxococcus xanthus*: characterization and synthesis during myxospore differentiation. *Journal of Bacteriology*, 124(1):550–557, 1975.
- [87] David C Krakauer and Joshua B Plotkin. Redundancy, antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences*, 99(3):1405–1409, 2002.
- [88] Lee Kroos. Bacterial development in the fast lane. *Journal of Bacteriology*, 190:4373–4376, 7 2008.
- [89] Lee Kroos. Highly signal-responsive gene regulatory network governing myxococcus development. *Trends in Genetics*, 33:3–15, 1 2017.
- [90] Lee Kroos, Patricia Hartzell, Karen Stephens, and Dale Kaiser. A link between cell movement and gene expression argues that motility is required for cell-cell signaling during fruiting body development. *Genes & Development*, 2:1677–1685, 1988.
- [91] Lee Kroos and Dale Kaiser. Expression of many developmentally regulated genes in myxococcus depends on a sequence of cell interactions. *Genes & Development*, 1:840–854, 1987.
- [92] John K. Kruschke. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142:573–603, 2013.
- [93] Thomas Kruse, Sune Lobedanz, Nils M S Berthelsen, and Lotte Sùgaard-Andersen. C-signal: a cell surface-associated morphogen that induces and co-ordinates multicellular fruiting body morphogenesis and sporulation in *Myxococcus xanthus*. *Molecular Microbiology*, 40:156–168, 2001.
- [94] J M Kuner and D Kaiser. Fruiting body morphogenesis in submerged cultures of *Myxococcus xanthus*. *Journal of Bacteriology*, 151(1):458–461, 1982.

- [95] Jerry M Kunert and Dale Kaiser. Fruiting body morphogenesis in submerged cultures of *Myxococcus xanthus*. *Journal of Bacteriology*, pages 458–461, 1982.
- [96] K Kutsukake and T Iino. Role of the flia-flgm regulatory system on the transcriptional control of the flagellar regulon and flagellar formation in salmonella typhimurium. *Journal of Bacteriology*, 176(12):3598–3605, 1994.
- [97] Thomas LaBar and Christoph Adami. Different evolutionary paths to complexity for small and large populations of digital organisms. *PLoS Computational Biology*, 12(12):e1005066, 2016.
- [98] Thomas LaBar and Christoph Adami. Evolution of drift robustness in small populations. *Nature Communications*, 8(1):1012, 2017.
- [99] Yinghong Lan, Aaron Trout, Daniel M Weinreich, and C Scott Wylie. Natural selection can favor the evolution of ratchet robustness over evolution of mutational robustness. *bioRxiv*, page 121087, 2017.
- [100] Adam S Luring and Raul Andino. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathogens*, 6(7):e1001005, 2010.
- [101] Adam S Luring, Judith Frydman, and Raul Andino. The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol*, 11(5):327–336, 2013.
- [102] Bongsoo Lee, Carina Holkenbrink, Anke Treuner-Lange, and Penelope I. Higgs. *Myxococcus xanthus* developmental cell fate production: Heterogeneous accumulation of developmental regulatory proteins and reexamination of the role of mazf in developmental lysis. *Journal of Bacteriology*, 194:3058–3068, 6 2012.
- [103] Richard E Lenski, Charles Ofria, Travis C Collier, and Christoph Adami. Genome complexity, robustness and genetic interactions in digital organisms. *Nature*, 400(6745):661–664, 1999.
- [104] Richard E Lenski, Charles Ofria, Robert T Pennock, and Christoph Adami. The evolutionary origin of complex features. *Nature*, 423(6936):139–144, 2003.
- [105] Shengfeng Li, Bheong-Uk Lee, and Lawrence J Shimkets. *csgA* expression entrains *Myxococcus xanthus* development. *Genes & Development*, 6:401–410, 1992.
- [106] Sune Lobedanz and Lotte Sogaard-Andersen. Identification of the c-signal, a contact-dependent morphogen coordinating multiple developmental responses in *Myxococcus xanthus*. *Genes and Development*, 17:2151–2161, 9 2003.
- [107] R. Lux, Y. Li, A. Lu, and W. Shi. Detailed three-dimensional analysis of structural features of *Myxococcus xanthus* fruiting bodies using confocal laser scanning microscopy. *Biofilms*, 1(4):293–303, 2004.
- [108] Nicholas A. Lyons and Roberto Kolter. On the evolution of bacterial multicellularity. *Current Opinion in Microbiology*, 24:21–28, 4 2015.

- [109] Esteban Martínez-García, Pablo I. Nikel, Max Chavarría, and Víctor de Lorenzo. The metabolic cost of flagellar motion in *pseudomonas putida* kt2440. *Environmental Microbiology*, 16(1):291–303, 2014.
- [110] Joan Massagué. Tgf $\beta$ signalling in context. *Nature Reviews Molecular Cell Biology*, 13:616–630, 10 2012.
- [111] Linda L McCarter. Regulation of flagella. *Current Opinion in Microbiology*, 9(2):180–186, 2006. Cell Regulation / Edited by Werner Goebel and Stephen Lory.
- [112] Richard McElreath. *Statistical Rethinking; A Bayesian Course with Examples in R and STAN*, chapter 13. CRC Press, Boca Raton, 2020.
- [113] Patrick J Mears, Santosh Koirala, Chris V Rao, Ido Golding, and Yann R Chemla. *Escherichia coli* swimming is robust against variations in flagellar number. *eLife*, 3, February 2014.
- [114] Susana Merino, Jonathan G. Shaw, and Juan M. Tomás. Bacterial lateral flagella: an inducible flagella system. *FEMS Microbiology Letters*, 263(2):127–135, 10 2006.
- [115] James G. Mitchell. The energetics and scaling of search strategies in bacteria. *The American Naturalist*, 160(6):727–740, 2002. PMID: 18707461.
- [116] Sheenu Mittal and Lee Kroos. A combination of unusual transcription factors binds cooperatively to control *Myxococcus xanthus* developmental gene expression. *Proceedings of the National Academy of Sciences*, 106(6):1965–1970, 2009.
- [117] Sheenu Mittal and Lee Kroos. Combinatorial regulation by a novel arrangement of *frua* and *mrpc2* transcription factors during *Myxococcus xanthus* development. *Journal of Bacteriology*, 191:2753–2763, 4 2009.
- [118] Sampriti Mukherjee and Daniel B. Kearns. The structure and regulation of flagella in *bacillus subtilis*. *Annual Review of Genetics*, 48(1):319–340, 2014. PMID: 25251856.
- [119] Frank D. Müller, Christian W. Schink, Egbert Hoiczyk, Emöke Cserti, and Penelope I. Higgs. Spore formation in *Myxococcus xanthus* is tied to cytoskeleton functions and polysaccharide spore coat deposition. *Molecular Microbiology*, 83(3):486–505, 2012.
- [120] Carey D Nadell, Knut Drescher, and Kevin R Foster. Spatial structure, cooperation and competition in biofilms. *Nature Publishing Group*, 2016.
- [121] Javad Najafi, Mohammad Reza Shaebani, Thomas John, Florian Altegoer, Gert Bange, and Christian Wagner. Flagellar number governs bacterial spreading and transport efficiency. *Science Advances*, 4(9), 2018.
- [122] Nguyen T. Q. Nhu, John S. Lee, Helen J. Wang, Yann S. Dufour, and Laurie E. Comstock. Alkaline ph increases swimming speed and facilitates mucus penetration for *vibrio cholerae*. *Journal of Bacteriology*, 203(7):e00607–20, 2021.



- [123] Bin Ni, Bhaswar Ghosh, Ferencz S. Paldy, Remy Colin, Thomas Heimerl, and Victor Sourjik. Evolutionary remodeling of bacterial motility checkpoint control. *Cell Reports*, 18:866–877, 1 2017.
- [124] Martin A Nowak. What is a quasispecies? *Trends in Ecology & Evolution*, 7(4):118–121, 1992.
- [125] Kathleen A O’connor and David R Zusman. Development in *Myxococcus xanthus* involves differentiation into two cell types, peripheral rods and spores. *JOURNAL OF BACTERIOLOGY*, 173:3318–3333, 1991.
- [126] Brendan D O’Fallon, Frederick R Adler, and Stephen R Proulx. Quasi-species evolution in subdivided populations favours maximally deleterious mutations. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1629):3159–3164, 2007.
- [127] C. Ofria and C. O. Wilke. Avida: a software platform for research in computational evolutionary biology. *Artificial Life*, 10(2):191–229, 2004.
- [128] Charles Ofria, David M Bryson, and Claus O Wilke. Avida: A software platform for research in computational evolutionary biology. In Andrew Adamatzky Maciej Komosinski, editor, *Artificial Life Models in Software*, pages 3–35. Springer London, 2009.
- [129] George A. O’Toole and Roberto Kolter. Flagellar and twitching motility are necessary for pseudomonas aeruginosa biofilm development. *Molecular Microbiology*, 30(2):295–304, 1998.
- [130] Jonathan D. Partridge and Rasika M. Harshey. More than motility: Salmonella flagella contribute to overriding friction and facilitating colony hydration during swarming. *Journal of Bacteriology*, 195(5):919–929, 2013.
- [131] Jonathan D. Partridge and Rasika M. Harshey. Swarming: Flexible roaming plans. *Journal of Bacteriology*, 195(5):909–918, 2013.
- [132] Joyce E. Patrick and Daniel B. Kearns. Swarming motility and the control of master regulators of flagellar biosynthesis. *Molecular Microbiology*, 83(1):14–23, 2012.
- [133] Robert T Pennock. Models, simulations, instantiations, and evidence: The case of digital evolution. *Journal of Experimental & Theoretical Artificial Intelligence*, 19(1):29–42, 2007.
- [134] Gordon R. Plague, Krystal S. Boodram, Kevin M. Dougherty, Sandar Bregg, Daniel P. Gilbert, Hira Bakshi, and Daniel Costa. Transposable elements mediate adaptive debilitation of flagella in experimental escherichia coli populations. *Journal of Molecular Evolution*, 84(5-6):279–284, June 2017.
- [135] E. M. Purcell. *Life at Low Reynolds Number*, pages 47–67. 1977.

- [136] Boyang Qin, Chenyi Fei, Andrew A. Bridges, Ameya A. Mashruwala, Howard A. Stone, Ned S. Wingreen, and Bonnie L. Bassler. Cell position fates and collective fountain flow in bacterial biofilms revealed by light-sheet microscopy. *Science*, 369(6499):71–77, 2020.
- [137] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [138] Paul B. Rainey and Michael Travisano. Adaptive radiation in a heterogeneous environment. *Nature*, 394(6688):69–72, Jul 1998.
- [139] Ramya Rajagopalan and Lee Kroos. Nutrient-regulated proteolysis of mrpc halts expression of genes important for commitment to sporulation during *Myxococcus xanthus* development. *Journal of Bacteriology*, 196:2736–2747, 2014.
- [140] Ramya Rajagopalan and Lee Kroos. The *dev* operon regulates the timing of sporulation during *Myxococcus xanthus* development. *Journal of Bacteriology*, 199, 5 2017.
- [141] Ramya Rajagopalan, Sébastien Wielgoss, Gerardo Lippert, Gregory J. Velicer, and Lee Kroos. *devi* is an evolutionarily young negative regulator of *Myxococcus xanthus* development. *Journal of Bacteriology*, 197:1249–1262, 2015.
- [142] Hans Reichenbach. *Myxococcus* spp. myxobacterales - schwarmentwicklung und bildung von protocysten. IWF Göttingen, 1965. <https://doi.org/10.3203/IWF/E-778> Last accessed: 15 Jul 2021.
- [143] Mark Robinson, Bongjun Son, David Kroos, and Lee Kroos. Transcription factor mrpc binds to promoter regions of hundreds of developmentally-regulated genes in *Myxococcus xanthus*. *BMC Genomics*, 15, 2014.
- [144] Curtis T. Rueden, Johannes Schindelin, Mark C. Hiner, Barry E. DeZonia, Alison E. Walter, Ellen T. Arena, and Kevin W. Eliceiri. ImageJ2: ImageJ for the next generation of scientific image data, Nov 2017.
- [145] B. Sager and D. Kaiser. Spatial restriction of cellular differentiation. *Genes Dev*, 7(9):1645–1653, Sep 1993.
- [146] Brian Sager and Dale Kaiser. Two cell-density domains within the *Myxococcus xanthus* fruiting body (multicellular development/pattern formation/morphogenesis/myxobacterales). *Proc. Natl. Acad. Sci. USA*, 90:3690–3694, 1993.
- [147] Brian Sager and Dale Kaiser. Intercellular c-signaling and the traveling waves of myxococcus. *Genes & Development*, 8, 1994.
- [148] Shreya Saha, Pintu Patra, Oleg Igoshin, and Lee Kroos. Systematic analysis of the *Myxococcus xanthus* developmental gene regulatory network supports posttranslational regulation of frua by c-signaling. *Molecular Microbiology*, 111:1732–1752, 6 2019.

- [149] J. Sambrook, E.F Fritsch, and T. Maniatis. *Molecular cloning: a laboratory manual ; 2nd ed.* Cold Spring Harbor Laboratory Press, 1989.
- [150] Rafael Sanjuán, José M Cuevas, Victoria Furió, Edward C Holmes, and Andrés Moya. Selection for robustness in mutagenized RNA viruses. *PLoS Genetics*, 3(6):e93, 2007.
- [151] Jan S. Schuhmacher, Kai M. Thormann, and Gert Bange. How bacteria maintain location and number of flagella? *FEMS Microbiology Reviews*, 39(6):812–822, 07 2015.
- [152] Peter Schuster and Jörg Swetina. Stationary mutant distributions and evolutionary optimization. *Bulletin of Mathematical Biology*, 50(6):635–660, 1988.
- [153] Luca Scrucca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *The R Journal*, 8(1):289–317, 2016.
- [154] Jun seok Lee, Bongjun Son, Poorna Viswanathan, Paul M. Luethy, and Lee Kroos. Combinatorial regulation of *fmgD* by *mrpc2* and *frua* during *Myxococcus xanthus* development. *Journal of Bacteriology*, 193(7):1681–1689, 2011.
- [155] Lawrence J Shimkets and Sheilah J Asher. Use of recombination techniques to examine the structure of the *csg* locus of *Myxococcus xanthus*. *Molecular and General Genetics*, 211:63–71, 1988.
- [156] Oleksii Sliusarenko, David R. Zusman, and George Oster. Aggregation during fruiting body formation in *Myxococcus xanthus* is driven by reducing cell movement. *Journal of Bacteriology*, 189:611–619, 1 2007.
- [157] Daniel R. Smith, Matthew R. Chapman, Thomas J. Silhavy, and James Tiedje. Economical evolution: Microbes reduce the synthetic cost of extracellular proteins. *mBio*, 1(3):e00131–10, 2010.
- [158] L Sogaard-Andersen and D Kaiser. C factor, a cell-surface-associated intercellular signaling protein, stimulates the cytoplasmic *frz* signal transduction system in *Myxococcus xanthus*. *Proceedings of the National Academy of Sciences*, 93(7):2675–2679, 1996.
- [159] Bongjun Son, Yu Liu, and Lee Kroos. combinatorial regulation by *mrpc2* and *frua* involves three sites in the *finge* promoter region during *Myxococcus xanthus* development. *Journal of Bacteriology*, 193:2756–2766, 6 2011.
- [160] Yoshiyuki Sowa and Richard M. Berry. Bacterial flagellar motor. *Quarterly Reviews of Biophysics*, 41(2):103–132, 2008.
- [161] Apollo Stacy, Luke McNally, Sophie E. Darch, Sam P. Brown, and Marvin Whiteley. The biogeography of polymicrobial infection. *Nature Reviews Microbiology*, 14:93–105, 2 2016.
- [162] Stan Development Team. RStan: the R interface to Stan, 2020. R package version 2.21.2.

- [163] Giovanni Strona and Kevin D Lafferty. Environmental change makes robust ecological networks fragile. *Nature Communications*, 7:12462, 2016.
- [164] Stella Stylianidou, Connor Brennan, Silas B. Nissen, Nathan J. Kuwada, and Paul A. Wiggins. Supersegger: robust image segmentation, analysis and lineage tracking of bacterial cells. *Molecular Microbiology*, 102(4):690–700, 2016.
- [165] Toon Swings, Bram Van den Bergh, Sander Wuyts, Eline Oeyen, Karin Voordeckers, Kevin J Verstrepen, Maarten Fauvart, Natalie Verstraeten, and Jan Michiels. Adaptive tuning of mutation rates allows fast response to lethal stress in *Escherichia coli*. *eLife*, 6:e22939, may 2017.
- [166] L Søgaaard-Andersen, F J Slack, H Kimsey, and D Kaiser. Intercellular c-signaling in *Myxococcus xanthus* involves a branched signal transduction pathway. *Genes & Development*, 10(6):740–754, 1996.
- [167] Lotte Søgaaard-Andersen, Martin Overgaard, Sune Lobedanz, Eva Ellehauge, Lars Jelsbak, and Anders Aa Rasmussen. Coupling gene expression and multicellular morphogenesis during fruiting body formation in *Myxococcus xanthus*. *Molecular Microbiology*, 48:1–8, 2003.
- [168] John R. Taylor and Roman Stocker. Trade-offs of chemotactic foraging in turbulent water. *Science*, 338(6107):675–679, 2012.
- [169] Jennifer K. Teschler, David Zamorano-Sánchez, Andrew S. Utada, Christopher J.A. Warner, Gerard C.L. Wong, Roger G. Linington, and Fitnat H. Yildiz. Living in the matrix: Assembly and control of vibrio cholerae biofilms. *Nature Reviews Microbiology*, 13:255–268, 5 2015.
- [170] Shashi Thutupalli, Mingzhai Sun, Filiz Bunyak, Kannappan Palaniappan, and Joshua W. Shaevitz. Directional reversals enable *Myxococcus xanthus* cells to produce collective one-dimensional streams during fruiting-body formation. *Journal of the Royal Society Interface*, 12, 8 2015.
- [171] Linda Turner, Liam Ping, Marianna Neubauer, and Howard C. Berg. Visualizing flagella while tracking bacteria. *Biophysical Journal*, 111(3):630–639, Aug 2016.
- [172] Nicholas Turner and Richard Grose. Fibroblast growth factor signalling: From development to cancer. *Nature Reviews Cancer*, 10:116–129, 2 2010.
- [173] Erik Van Nimwegen, James P Crutchfield, and Martijn Huynen. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences*, 96(17):9716–9720, 1999.
- [174] Tim van Opijnen, Kip L. Bodi, and Andrew Camilli. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, 6(10):767–772, Oct 2009.

- [175] Gregory J. Velicer and Michiel Vos. Sociobiology of the myxobacteria. *Annual Review of Microbiology*, 63:599–623, 10 2009.
- [176] Kartik Viswanathan, Poorna Viswanathan, and Lee Kroos. Mutational analysis of the *Myxococcus xanthus*  $\omega$ 4406 promoter region reveals an upstream negative regulatory element that mediates c-signal dependence. *Journal of Bacteriology*, 188:515–524, 1 2006.
- [177] Poorna Viswanathan, Toshiyuki Ueki, Sumiko Inouye, and Lee Kroos. Combinatorial regulation of genes essential for *Myxococcus xanthus* development involves a response regulator and a lysr-type regulator. *Proceedings of the National Academy of Sciences*, 104(19):7969–7974, 2007.
- [178] Nikita Vladimirov and Victor Sourjik. Chemotaxis: how bacteria use memory. 390(11):1097–1104, 2009.
- [179] Hera Vlamakis, Yunrong Chai, Pascale Beauregard, Richard Losick, and Roberto Kolter. Sticking together: Building a biofilm the bacillus subtilis way. *Nature Reviews Microbiology*, 11:157–168, 3 2013.
- [180] Xiaoyi Wang, Santosh Koirala, Phillip D. Aldridge, Christopher V. Rao, and Yves V. Brun. Two tandem mechanisms control bimodal expression of the flagellar genes in salmonella enterica. *Journal of Bacteriology*, 202(13):e00787–19, 2020.
- [181] Yan Wei, Xiaolin Wang, Jingfang Liu, Ilya Nememan, Amoolya H. Singh, Howie Weiss, and Bruce R. Levin. The population dynamics of bacteria in physically structured habitats and the adaptive virtue of random motility. *Proceedings of the National Academy of Sciences*, 108(10):4047–4052, 2011.
- [182] Marvin Whiteley, Stephen P. Diggle, and E. Peter Greenberg. Progress in and promise of bacterial quorum sensing research. *Nature*, 551:313–320, 11 2017.
- [183] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- [184] Uri Wilensky.
- [185] Claus O Wilke. Adaptive evolution on neutral networks. *Bulletin of Mathematical Biology*, 63(4):715–730, 2001.
- [186] Claus O Wilke. Quasispecies theory in the context of population genetics. *BMC Evolutionary Biology*, 5(1):44, 2005.
- [187] Claus O Wilke and Christoph Adami. Interaction between directional epistasis and average mutational effects. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1475):1469–1474, 2001.
- [188] Claus O Wilke and Christoph Adami. Evolution of mutational robustness. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 522(1):3–11, 2003.

- [189] Claus O Wilke, Jia Lan Wang, Charles Ofria, Richard E Lenski, and Christoph Adami. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844):331–333, 2001.
- [190] Jon F Wilkins, Peter T McHale, Joshua Gervin, and Arthur D Lander. Survival of the curviest: Noise-driven selection for synergistic epistasis. *PLoS Genetics*, 12(4):e1006003, 2016.
- [191] John W Wireman and Martin Dworkin. Developmentally induced autolysis during fruiting body formation by *Myxococcus xanthus*. *Journal of Bacteriology*, 129:796–802, 1977.
- [192] Simon N Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Statist. Soc. B*, 73:3–36, 2010.
- [193] Samuel S. Wu and Dale Kaiser. Genetic and functional evidence that type iv pili are required for social gliding motility in *Myxococcus xanthus*. *Molecular Microbiology*, 18(3):547–558, 1995.
- [194] Chunyan Xie, Haiyang Zhang, Lawrence J. Shimkets, and Oleg A. Igoshin. Statistical image analysis reveals features affecting fates of *Myxococcus xanthus* developmental aggregates. *Proceedings of the National Academy of Sciences of the United States of America*, 108:5915–5920, 4 2011.
- [195] Ri-Sheng Yang and Yi-Tao Chen. Flagellation of shewanella oneidensis impacts bacterial fitness in different environments. *Current Microbiology*, 77(8):1790–1799, Aug 2020.
- [196] Zhaomin Yang and Penelope Higgs, editors. *Myxobacteria: Genomics, Cellular and Molecular Biology*. Caister Academic Press, 2014.
- [197] Xiao Yi and Antony M Dean. Phenotypic plasticity as an adaptation to a functional trade-off. *eLife*, 5, October 2016.
- [198] Kevin D. Young. The selective value of bacterial shape. *Microbiology and Molecular Biology Reviews*, 70(3):660–703, 2006.
- [199] Haiyang Zhang, Stuart Angus, Michael Tran, Chunyan Xie, Oleg A. Igoshin, and Roy D. Welch. Quantifying aggregation dynamics during *Myxococcus xanthus* development. *Journal of Bacteriology*, 193:5164–5170, 10 2011.
- [200] Haiyang Zhang, Zalman Vaksman, Douglas B. Litwin, Peng Shi, Heidi B. Kaplan, and Oleg A. Igoshin. The mechanistic basis of *Myxococcus xanthus* rippling behavior and its physiological role during predation. *PLoS Computational Biology*, 8, 9 2012.
- [201] Zhaoyang Zhang, Christopher R Cotter, Zhe Lyu, Lawrence J Shimkets, and Oleg A Igoshin. Data-driven models reveal mutant cell behaviors important for myxobacterial aggregation. *mSystems*, 5, 2020.

- [202] Shiwei Zhu, Seiji Kojima, and Michio Homma. Structure, gene regulation and environmental response of flagella in vibrio. *Frontiers in Microbiology*, 4:410, 2013.