

Tools for Manipulation of Microbial Communities through Bacterial Conjugation

by

Aaron Yip

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2024

© Aaron Yip 2024

Examining Committee Membership

External Examiner: Ting Lu
Professor
Department of Bioengineering
University of Illinois Urbana-Champaign

Supervisors: Brian Ingalls
Professor
Department of Applied Mathematics
University of Waterloo

Marc Aucoin
Professor
Department of Chemical Engineering
University of Waterloo

Internal Members: Valerie Ward
Associate Professor
Department of Chemical Engineering
University of Waterloo

Nasser Mohieddin Abukhdeir
Associate Professor
Department of Chemical Engineering
University of Waterloo

Internal-External Member: Trevor Charles
Professor
Department of Biology
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapters 1, 2, and 7: I am the sole author of this material, under supervision of Brian Ingalls and Marc Aucoin. This material has not been previously published.

Chapter 3: The work included in this chapter is published: A. Yip, O.D. McArthur, K.C. Ho, M.G. Aucoin, B.P. Ingalls. Degradation of polyethylene terephthalate (PET) plastics by wastewater bacteria engineered via conjugation. *Microbial Biotechnology*, 17 (2024), e70015.

I conceptualized the research and provided supervision to Owen McArthur (Undergraduate student, Department of Biology, University of Waterloo). I and O.D.M. designed and constructed plasmids and performed filter-mating experiments. O.D.M. developed the filter-mating protocol for mating with wastewater bacteria and performed early-phase testing for confirming expression of FAST-PETase in transconjugants. I performed PET degradation experiments, with assistance from Kalista Ho (Undergraduate student, Department of Biology, University of Waterloo) in performing protein purifications and preparing PET samples. M.G.A. wrote software for visualizing cell density count data. I analyzed the results, wrote the original manuscript, and revised the manuscript with guidance from B.P.I and M.G.A. and feedback from peer-reviewers. B.P.I and M.G.A. provided guidance, facilities, equipment, and other resources throughout the project. Michael Lynch (Metagenom Bio Life Science) provided the protocol for bITS sequencing.

Chapter 4: The work included in this chapter is published: A. Yip, J. Smith-Roberge, S. Haghayegh-Khorasani, M.G. Aucoin, B.P. Ingalls. Calibrating spatiotemporal models of microbial communities to microscopy data: A review. *PLoS Computational Biology*, 18 (2022), e1010533.

B.P.I. promoted the idea of writing this review. I surveyed literature for experimental methods related to single-cell observations of microbial communities, calibration techniques used in agent-based models of microbial communities, summary statistics used in quantifying spatio-temporal patterns in microbial communities, and opportunities for improvements in model calibration. I wrote the Introduction and sections pertaining to the aforementioned

contributions and the Conclusion, and revised my sections of the manuscript with feedback from peer-reviewers. Julien Smith-Roberge (PhD Candidate, Applied Mathematics, University of Waterloo) surveyed literature pertaining to calibration of partial differential equation models of microbial communities, wrote the section on feature identification using topological data analysis, and contributed to describing summary statistics. Sara Haghayegh-Khorasani (PhD Candidate, Department of Chemical Engineering, University of Waterloo) assisted in surveying for literature related to agent-based modelling of microbial communities. B.P.I. and M.G.A. provided supervision, reviewed, and edited the manuscript.

Chapter 5: The work included in this chapter is published: A. Ahmadi⁺, A. Yip⁺, J. Chalaturnyk, M.G. Aucoin, B.P. Ingalls. A pipeline for calibrating agent-based models of microbial populations: From image collection to model parameterization. Proceedings of the 10th IFAC International Conference on Foundations of Systems Biology in Engineering, (2024).

I wrote the initial draft of the Introduction and Conclusion, performed the experiments and single-cell measurements, implemented code for calculating the order parameter and growth-related summary statistics, performed the parameter inference, and wrote the corresponding sections. Atiyeh Ahmadi (PhD Candidate, Department of Biology, University of Waterloo) wrote the Trackrefiner software, demonstrated that the CellProfiler and Omnipose software combination was the best option for cell segmentation and tracking, performed the image processing, cell tracking, implemented code for the convexity and aspect ratio calculations, performed the sensitivity analysis, and wrote the corresponding sections. Jonathan Chalaturnyk (Master's student, Department of Applied Mathematics, University of Waterloo) assisted with early phase testing of PyABC with agent-based models and implemented the algorithm for the colony density calculation. B.P.I conceptualized the research and provided supervision along with M.G.A.; B.P.I. revised the manuscript. Primary authorship was shared between A.A. and me.

Chapter 6: Matthew Courtney (PhD Candidate, Department of Mechanical and Mechatronics Engineering, University of Waterloo) designed and manufactured molds for the microfluidic devices using facilities and equipment provided by Carolyn Ren (Professor, Department of Mechanical and Mechatronics Engineering, University of Waterloo). M.C. wrote the method for designing and manufacturing the molds for the microfluidic devices. Atiyeh Ahmadi wrote the Trackrefiner software. This chapter has not been previously published.

Abstract

Microbial communities are ubiquitous across the planet and play essential roles in numerous aspects of society including human health, agriculture, food production, and biodegradation. Controlled manipulation of natural microbial communities by introduction of biological agents is a promising option for precise, *in situ* microbiome editing. Bacterial conjugation is a well-studied mechanism for gene exchange (horizontal gene transfer) between bacteria in physical proximity to one another. Conjugation can be repurposed as a gene delivery technique into microbial communities and holds potential for addressing problems of environmental and clinical relevance, including the biodegradation of pollutants such as microplastics and delivery of precision antimicrobials.

This thesis begins with a proof-of-concept study in which recombinant DNA is delivered to bacteria in wastewater via conjugation to enable the degradation of polyethylene terephthalate (PET), a major component of global plastic pollution. Using a broad-host-range conjugative plasmid, we enabled expression of FAST-PETase in various bacterial species from municipal wastewater, achieving substantial degradation of both commercial PET film and post-consumer PET products under laboratory conditions.

Next, the thesis builds computational tools and methodologies to support model-based design of *in situ* gene delivery into microbial communities through conjugation. Prior to starting model development, a review of practices for calibrating spatio-temporal models to microscopy data was conducted, which demonstrated the need for a more formal process for generating predictive models. Drawing on practices from ecology, new strategies for systematic model validation were proposed. Some of these strategies were then implemented into a model calibration pipeline based on Pattern-Oriented-Modelling and Bayesian parameter inference, and the pipeline was demonstrated by fitting biophysical parameters in an agent-based model to time-lapse microscopy data. The calibrated model was able to reproduce several patterns in microcolony formation that were observed experimentally, but did not fully replicate patterns associated with colony shape.

Finally, a single-cell-based approach to characterizing conjugation in microfluidic environments was developed to investigate spread of the previously developed conjugative plasmid that enabled expression of FAST-PETase. Although the plasmid could spread through conjugation, cells bearing the conjugative plasmid tended to get outcompeted for space by the healthier recipient population. The tendency of cells to self-orient in the direction towards the exit of the microfluidic traps through biomechanical processes also reduced conjugation efficiency. The previously developed model calibration pipeline was then applied to calibrate an agent-based model for conjugation to the microscopy data collected. Afterwards, the calibrated model was used to characterize how initial conditions and spatial factors influenced spread of conjugative plasmids in enclosed microenvironments.

Collectively, these studies enhance understanding of engineering microbial communities through conjugation, offering novel solutions for plastic waste degradation and advancing model-based design of gene delivery into microbial communities.

Acknowledgements

I knocked on the door of Brian Ingalls with nothing but a couple chemical engineering degrees and a desire to make the world a better place. Despite my almost non-existent biological toolkit prior to starting this PhD, Brian agreed to work with me, and I cannot thank him enough for taking this chance on me. I am thankful for his guidance, perspectives, and encouragement that led me to this moment. I am also grateful for all the opportunities he gave me to learn, lead, and to connect with the world.

I am also thankful to have had Marc Aucoin on my team. Marc was a great coach who demonstrated the highest level of professionalism one could ask for. I thank him for all the scholarly discussions, his mentorship, and for providing me with the tools that I needed to succeed.

I thank the following professors for fruitful discussions and their feedback on my work during my time at Waterloo: Valerie Ward, Matthew Scott, Christian Euler, Yilan Liu, Ellie Prince, and Dave Edgell. I am especially thankful to Valerie for giving me life-saving molecular cloning advice in times when I most needed it.

I am infinitely grateful to Owen McArthur for his help in turning my vision for turning environmental bacteria into plastic-eating machines into a reality. My project would not have taken off without his passion for research, his prowess in the lab, and his consistent work ethic. I also want to thank him for just being an overall wonderful person to be around, both in and out of the lab. I also thank Kalista Ho for her time volunteering in the lab doing countless protein purifications and hole-punching endless amounts of plastic garbage for science.

I am equally grateful to my colleague and good friend Atiyeh Ahmadi. Ati was always there for me, to give perspective, and to remind me of what truly matters in life. I am so thankful for having a colleague and true friend who understood me, who walked with me through the highest peaks and the lowest valleys. My journey would not have been the same without her.

I want to thank Matthew Courtney for introducing me to the fascinating world of microfluidics. Matt's character and thought process are that of a true engineer, and I am thankful that I got to have such a role model during my time in Waterloo.

I am grateful to all the genuine people who helped ease me into working in a biological lab, specifically, Thomas Hamilton, Anirban Banerjee, and Mark Bruder. They were always willing to answer my questions about basic lab protocols (which were not so basic for me at the time), as well as share their own advice on how to navigate the PhD journey based on their personal experiences.

I am grateful for the short and sweet two months I had with Vincent Demont in 2022. I'll never forget the adventures we had during the COVID rollercoaster, learning more about French culture, and just experiencing research at it's best, together. I am also grateful for getting to know Sultan Nazir, for showing me the beautiful things that arise from doing real math and biology.

I thank all of the excellent undergraduate students that I had the pleasure collaborating with, for all their hard work, dedication, willingness to learn, and for teaching me many lessons about what it means to be a good mentor: Luke Zhang, Joyce Yang, Aamna Tahir, Henry Zhang, Melissa Meighoo, Sohaib Nadeem, Sheng Fang, Adam Yuen Miles, Taha Farooqui, Siddharth Krishnamurthy, Julie Kim, and Nitya Yadav.

Thanks to Scott Boegel and Madhuj Chakraborty for helping me place equipment/reagent orders in E6, as well as for their good company in the office. Special thanks to Scott for having my back when the lights went off unexpectedly whilst I was working in the fume hood. I also want to thank Sanniv Ganguly, Wiweka Dharmasiddhi, and Abhishek Mishra for all the good times working in E6 and beyond.

Thank you to folks from CEGSA for all their work in organizing so many fun social activities that I would always look forward to during my time at Waterloo.

I genuinely thank my colleagues who promoted diversity, inclusion, and most of all, respect in the spaces that I worked in; your character and efforts don't go unnoticed.

I want to give a special thank you to the Waterloo Warriors Tennis Team for allowing me to fulfill my dream of competing alongside the best players at the university Varsity level. I am so grateful to have had this second family in Waterloo for all their support on and off the court. An extra special thanks to coach Ned Lazoja, for creating an amazing and fun practice environment, for fixing my serve, and for just being a great guy.

I am also grateful for the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) over my studies, which gave me more freedom to pursue my own projects.

Finally, I want to acknowledge my family for always letting me know that they are with me. I would not have been able to achieve everything I have without their love and support.

Dedication

To Maggie and Charlotte, for their unconditional love and the joy they brought into my life.

Table of Contents

Author's Declaration.....	iii
Statement of Contributions	iv
Abstract	vii
Acknowledgements	ix
Dedication	xii
List of Figures	xviii
List of Tables	xxv
List of Abbreviations	xxvii
Quote.....	xxix
Chapter 1 Introduction	1
1.1 Background	1
1.1.1 Microbiome engineering through biological augmentation	1
1.1.2 Methods for delivering DNA into microbiomes.....	4
1.1.3 Predicting efficacy of genetic bioaugmentation mediated through conjugation	5
1.1.4 Conjugation in context of the global plastic pollution problem	6
1.2 Thesis Outline	8
Chapter 2 Literature Review	10
2.1 Global plastic waste problem	10
2.2 Biological degradation of plastics	11
2.2.1 PET degradation mechanism in <i>Ideonella sakaiensis</i>	12
2.2.2 Enhancements to PET degradation using protein engineering	14
2.2.3 Engineering microbes to degrade PET	15
2.2.4 Engineering microbial communities to degrade PET	16
2.3 Bacterial Conjugation.....	18
2.3.1 Conjugation mechanism	18
2.3.2 Factors influencing spread of conjugative plasmids.....	20
2.3.3 Quantifying bacterial conjugation using end-point methods.....	22

2.3.4 Observations of bacterial conjugation using microscopy	24
2.4 Agent-based modelling of microbial communities	25
Chapter 3 Degradation of polyethylene terephthalate (PET) plastics by wastewater bacteria engineered via conjugation	29
3.1 Abstract	29
3.2 Introduction	30
3.3 Results	32
3.3.1 pFAST-PETase-cis conjugates to bacteria from wastewater	32
3.3.2 pFAST-PETase-cis is lost in absence of selection	37
3.3.3 Purified FAST-PETase expressed from transconjugant isolates degrades PET.....	38
3.3.4 Culture supernatant containing FAST-PETase degrades post-consumer PET without enzymatic purification	41
3.4 Discussion	44
3.5 Methods	49
3.5.1 Bacterial strains	49
3.5.2 Plasmid construction.....	49
3.5.3 Filter-mating conjugation and selection of engineered isolates	50
3.5.4 Conjugation efficiency in wastewater conditions.....	51
3.5.5 Plasmid stability	52
3.5.6 Protein expression and purification	52
3.5.7 PET sample preparation.....	53
3.5.8 PET degradation assay with purified FAST-PETase	53
3.5.9 PET degradation with culture supernatant.....	54
3.5.10 Scanning electron microscopy	55
3.5.11 Differential scanning calorimetry	55
3.5.12 Statistical analysis.....	55
3.5.13 Imaging plates.....	56
3.5.14 Confirming auxotrophy of <i>E. coli</i> donor strain	56
3.5.15 Identification of isolates	56

3.6 Acknowledgements	57
3.7 References	59
Chapter 4 Calibrating spatiotemporal models of microbial communities to microscopy data: A review.....	65
4.1 Abstract	65
4.2 Introduction	66
4.3 Single-Cell Level Observations of Microbial Communities.....	70
4.4 Calibration of Spatial Mathematical Models of Microbial Communities Against Single-cell Measurements.....	72
4.4.1 A Survey of Calibration Approaches for Spatiotemporal Models	74
4.4.2 A Catalogue of Spatiotemporal Summary Statistics for Microbial Community Dynamics	76
4.5 Outlooks	86
4.5.1 Pattern-Oriented Modelling as a Guideline for Standardizing Microbiological Models	87
4.5.2 Feature Identification through Topological Data Analysis	88
4.5.3 Machine Learning Algorithms to Accelerate Model Calibration	91
4.6 Conclusions	93
4.7 References	95
Chapter 5 A pipeline for calibrating agent-based models of microbial populations: From image collection to model parameterization.....	106
5.1 Abstract	106
5.2 Introduction	107
5.3 Methods.....	109
5.3.1 Microscopy	109
5.3.2 Cell tracking and measurements.....	109
5.3.3 Agent-based modelling.....	110
5.3.4 Summary statistics.....	111
5.3.5 Parameter inference	113

5.4 Results and Discussion.....	114
5.5 Conclusion.....	120
5.6 Acknowledgements	121
5.7 References	122
Chapter 6 Characterizing bacterial conjugation dynamics in microfluidic environments....	124
6.1 Abstract	124
6.2 Introduction	125
6.3 Results and Discussion.....	130
6.3.1 Delay between conjugation events and evidence of conjugation	130
6.3.2 Competition and reduced intermixing limit spread of conjugative plasmids	132
6.3.3 Fitness burden associated with the conjugative plasmid	136
6.3.4 Formulating an agent-based model for bacterial conjugation	141
6.3.5 Calibration of ABM reproduces plasmid conjugation dynamics	145
6.3.6 Initial conditions influence gene delivery outcomes	149
6.4 Concluding Remarks	153
6.5 Methods.....	155
6.5.1 Strains, growth conditions, and plasmid construction.....	155
6.5.2 Preparation of microfluidic devices.....	155
6.5.3 Microscopy	156
6.5.4 Microfluidic experiments	156
6.5.5 Image analysis	157
6.5.6 Validation of automated conjugation detection.....	158
6.5.7 Converting tracking data to lineage tree data structure	158
6.5.8 Measurement of single-cell properties	159
6.5.9 Intermixing	159
6.5.10 Isotropy	160
6.5.11 Analysis of conjugation events.....	160
6.5.12 Conjugation rate	161
6.5.13 Death rate.....	161

6.5.14 Agent-based model	162
6.5.15 Model calibration.....	165
6.5.16 Simulation case studies on initial conditions.....	166
6.6 Acknowledgements	166
Chapter 7 Discussion	168
7.1 Engineering microbiomes to degrade plastics.....	168
7.2 Gene delivery into microbial communities	170
7.3 Agent-based modelling of microbial communities	172
7.4 Conclusions	175
References	177
Appendix A Supplementary Data for Chapter 3.....	192
Appendix B Supplementary Data for Chapter 6.....	193
Appendix C Copyright Permission for Published Chapters	199
Appendix D Description of Supplementary Files.....	200
Appendix E CellModeller Biophysical Model Description.....	201

List of Figures

Figure 1.1. Operations on microbiomes enabled by bioaugmentation depicted using a cartoon microbial interaction network. Nodes represent microbial species or guilds of microbes performing similar functions. Arrows indicate a positive ecological interaction (e.g., cross-feeding through metabolite secretion), while barred lines indicate a negative ecological interaction (e.g., bacteriocin production). Microbiomes can be manipulated by adding in new microbial species or by delivering new genes to resident members, removing certain species or genes, and modulating ecological interactions.	3
Figure 2.1 Two-enzyme pathway for degradation of polyethylene terephthalate (PET) into its constituent molecules in <i>Ideonella Sakaiensis</i> 201-F6. Depolymerization of PET by PETase primarily yields mono-(2-hydroxyethyl) terephthalate (MHET). Smaller amounts of bis-(2-hydroxyethyl) terephthalate (BHET) and terephthalic acid (TPA) are produced in the depolymerization step. BHET is further hydrolyzed into MHET and ethylene glycol (EG) by PETase. MHET is hydrolyzed into TPA and ethylene glycol (EG) by MHETase.....	13
Figure 3.1. Plasmid map of pFAST-PETase-cis. The corresponding GenBank formatted file of the nucleotide sequence is provided in Supplementary Data 3.4.....	33
Figure 3.2. Conjugation of pFAST-PETase-cis into wastewater bacteria. a Schematic map of pFAST-PETase-cis (not to scale): <i>oriT</i> , RK2/RP4 conjugative origin of transfer; FAST-PETase, gene for FAST-PETase enzyme; Amp ^R , ampicillin resistance gene (<i>bla-TEM₁</i>); <i>mCherry</i> , gene for fluorescent protein mCherry; Conjugation genes, encoding the RK2/RP4 conjugation system; Gm ^R , gentamicin resistance gene; pBBR1 <i>oriV</i> , plasmid origin of replication. b FAST-PETase coding region (not to scale) highlighting the arabinose-inducible promoter (P _{BAD}), signal peptide (SP _{stu}), and 6×His-tag. c Experimental procedure for conjugation of pFAST-PETase-cis into bacteria from a wastewater sample. d Experimental procedure for measuring conjugation efficiency in wastewater suspension.....	34
Figure 3.3. Selection plate for wastewater sample without donor strain added (diluted 10-fold). No fluorescent colonies are present, confirming that wild-type bacteria in the sample do not express mCherry or other red fluorescent proteins.	36

Figure 3.4. Donor strain incubated at 30 °C on M9 glucose plate with antibiotics for 6 days. No colonies are present, confirming auxotrophy of the donor strain. 37

Figure 3.5. Cell density of wastewater transconjugants plated on selective (ampicillin and gentamicin) and non-selective plates over 14 serial passages in LB. Bars show the mean of three biological replicates. Error bars: $\pm$ s.d. 38

Figure 3.6. Enzymatic degradation of aPET film (7 mm diameter, 0.25 mm thick, ~11 mg) by purified FAST-PETase solution obtained from isolated transconjugants. SDS-PAGE gels of purified FAST-PETase are shown for **a**) cell lysates and **c**) culture supernatant. L: protein ladder (PageRuler Unstained Protein Ladder, ThermoFisher, no. 26614); BSA: bovine serum albumin; numbers indicate isolate ID. Mass loss from exposure to FAST-PETase purified from **b**) cell lysates and **d**) supernatant was measured after 96 h incubation at 50 °C. Bars show mean of three technical replicates for each isolate and control. Error bars: $\pm$ s.d. Statistical analyses were performed using Welch's one-sided *t*-test to compare mass loss of each sample to the negative control that contained no FAST-PETase (Buffer) (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). **e** Scanning electron microscopy (SEM) image of an untreated PET film. **f** SEM image of a PET film incubated at 50 °C for 96 h with FAST-PETase purified from WW21 culture supernatant. Scale bars: 50 μ m. 40

Figure 3.7. Degradation of a coffee cup lid using spent supernatant from engineered wastewater isolates diluted in 0.1 M KH₂PO₄-NaOH buffer (pH 8.0). **a** A₂₄₀ measurements for diluted supernatant incubated with a 6 mm hole punch of pcPET for 120 h at 50 °C. ara+: protein expression induced with 0.2% arabinose; glu+: protein expression suppressed with 0.2% glucose; ara-: no arabinose added. Bars show mean of three technical replicates. Error bars: $\pm$ s.d. **b** Photos (inset) and SEM images of PET samples prior to treatment (left) or incubated in culture supernatant from WW5 (1:40) for 120 h. Scale bars: 50 μ m. **c** Coffee cup lid (bottom) shows macroscopic signs of degradation after 168 h incubation at 50 °C with spent supernatant (ara+) from WW2 (1:80). An untreated lid (top) and lid incubated with spent supernatant (glu+) from WW2 (1:80) (middle) are shown for comparison. **d** Difference between initial and final mass of coffee cup lids incubated in supernatant from WW2 (1:80) after 168 h incubation at 50 °C. Bars show mean of three technical replicates. Error bars: $\pm$ s.d. All statistical analyses

were performed using Welch's one-sided t -test to compare ara+ samples to glu+ and ara- samples ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$).	42
Figure 3.8. A_{240} measurements for M9 glycerol supernatant incubated with pcPET for 120 h at 50 °C. ara+: protein expression induced with 0.2% arabinose. Bars show mean of three technical replicates. Error bars: $\pm$ s.d.	43
Figure 4.1. Modelling frameworks commonly used for capturing the behaviour of microbial communities, with associated spatial scales. ODE: ordinary-differential equation; PDE: partial differential equation; ABM: agent-based model.	68
Figure 4.2. Single-cell measurement techniques for microbial communities. (a) Flow cytometers can measure cell fluorescence and morphology in large sample sizes. (b) Time-lapse microscopy allows for direct visualization of physical cell-cell interactions and quantitative measurement of single-cell characteristics over time. (c) End-point microscopy can illustrate large-scale spatial patterns and features of cell arrangements in 2 or 3 dimensions.	70
Figure 4.3. Model calibration techniques for spatiotemporal models of microbial communities. Manual fitting involves direct adjustment of parameter values to achieve qualitative agreement between model predictions and observations. Non-spatial calibration is often systematic (based on a goodness-of-fit function) but is based on experiments that do not incorporate the spatial features of the system. Spatial calibration, against spatially-distributed data, can be systematic (SSE-based), but must rely on summary statistics collected from the data.	75
Figure 4.4. Calculating spatiotemporal summary statistics for microbial communities. (a) Population counts over time capture the overall dynamics in a multi-species community. (b) The frequency of adjacent species in physical contact, determined by a contact network, provides a measure of intermixing between different species. (c) Summary statistics can be calculated from data averaged within a particular cell's neighbourhood. (d) Single-species patch metrics, such as patch width and number of sectors, are useful for quantifying spatial patterns on a larger colony scale.	84
Figure 4.5. Persistence homology. Panels (a-c) illustrate the topology of a dataset changing as the length scale, L is varied. (a) For small values of L , the balls (disks) are mostly disconnected;	

only two of the nine intersect. (b) At an intermediate scale, all nine balls intersect, forming a single connected component, giving rise to a loop. (c) At larger scales, there is a single connected component and no loop. (d) The progression illustrated in (a-c) is documented in the persistence barcode; the blue bars correspond to separate connected components, the ends of which corresponds to intersection (merge) events, e.g., at $L = L_1$. The red bar corresponds to the loop which forms at $L = L_2$ and which becomes filled in at $L = L_3$. (e) The same information can be represented in persistence diagram in which the (x,y) coordinates of points correspond to the right and left ends, respectively, of each bar in the barcode. 90

Figure 5.1. Monolayer microcolony of *E. coli* (left) and corresponding agent-based model simulation (right). Scale bar: 10 μm 109

Figure 5.2. Distribution of summary statistics for experimental observations of 24 microcolonies with end-point populations of 140-225 cells. Lower box boundary: 25th percentile; upper box boundary: 75th percentile; line inside box: median; whiskers: 1.5 $\times$ interquartile range. 114

Figure 5.3. Global sensitivity analysis across γ and α values. Red bars represent the standard deviation of each feature value determined from an ensemble of 50 simulations for each parameter pair. 116

Figure 5.4. Posterior probability densities for calibrated parameter values of γ and α after six rounds of ABC calibration. 118

Figure 5.5. Summary statistics for experimental data and simulations sampled from prior distributions and final posterior distributions ($n = 24$). Lower box boundary: 25th percentile; upper box boundary: 75th percentile; line inside box: median; whiskers: 1.5 $\times$ interquartile range. 119

Figure 5.6. KL divergence between experimental data and simulations sampled from prior distributions and final posterior distributions shown in Figure 5.5. 120

Figure 6.1. Methodology for high-throughput single-cell analysis of plasmid spread in microfluidic devices and agent-based model (ABM) development. (A) Microfluidic chip loaded onto an inverted microscope. Nutrients are fed through the inlet into the main flow channel. Bacteria grow in a monolayer inside traps on the sides of the main flow channel. (B)

Overview of the analysis pipeline. Time-lapse microscopy: fluorescence images are recorded to monitor donor, recipient, and transconjugant populations. Image processing: individual cells are distinguished through a segmentation algorithm. Cell tracking: cell lineages are tracked through time. Classify populations: cells are classified as donors, recipients, or transconjugants based on fluorescence. Contact tracing: all physical contacts between cells are monitored in each image. Infer conjugation events: timing of conjugation events is inferred based on the most recent common ancestor of transconjugants in the same family. Summary statistics: quantitative measures of population dynamics, conjugation efficiency, and intermixing are calculated. Model formulation & testing: an ABM is formulated followed by sensitivity analysis. Parameter inference: Model parameter values are calibrated to experimentally measured summary statistics using an Approximate Bayesian Computation-Scheduled Monte Carlo (ABC-SMC) algorithm built into pyABC ³⁵ .	129
Figure 6.2. Maximum delay between occurrence and detection of conjugation events.	132
Figure 6.3. Total cells per trap. Solid line represents the mean over $n = 17$ traps; shaded area represents $\pm$ s.d. All traps are from the same experiment.	133
Figure 6.4. Population-level features relevant to conjugation in microfluidic traps. Red, green, and yellow populations represent donors, recipients, and transconjugants, respectively. (A) Images of donors, recipients, and transconjugants growing in a representative microfluidic trap at times 0, 5, and 16 h. (B) Population dynamics for each cell type. (C) Intermixing between populations quantified by an entropy-based intermixing score. (D) Conjugation events measured in each frame summed over all traps. (E) Conjugation frequency measured as the ratio of the cumulative sum of conjugation events and D-R plus D-T contacts per hour, per trap. Conjugation efficiency: ratio of transconjugants and contacts between D-R plus T-R contacts. Solid lines represent the means over $n = 17$ traps in graphs with shaded areas; shaded areas represent $\pm$ s.d. All traps are from the same experiment.	134
Figure 6.5. Orientation of cells over time measured by a global order parameter. Solid line represents the mean over $n = 17$ traps; shaded area represents $\pm$ s.d. All traps are from the same experiment.....	135

Figure 6.6. Minimum windowed-averaged elongation rate recorded for each population in $n = 17$ traps over 16 h (donors: 24127; recipients: 55475; transconjugants: 9305).....	137
Figure 6.7. Probability density for single-cell growth rate, radius, and division length. Solid lines indicate probability density function fits for a normal distribution (growth rate, radius) or lognormal distribution (division length).	138
Figure 6.8. Instantaneous death rates for each population over all traps. The boxed area on the left panel is shown in higher detail in the right panel.	139
Figure 6.9. Analysis of death rates in transconjugants arising from conjugation events that occurred in the first 4 h of the experiment. (A) Minimum window-averaged elongation rate of cells involved in lineages receiving the conjugative plasmid. Transconjugant points indicate that the cell has been identified as a transconjugant by the image processing algorithm. (B) Histogram for transconjugant cells with windowed-averaged elongation rates below $1 \mu\text{m}/\text{h}$ with a gamma distribution fit. (C) Cumulative distribution of the gamma distribution with shape and rate parameters.	140
Figure 6.10. Parametric sweep over the transfer gene activation frequency for the basic conjugation model M0. Bar plots show simulated endpoint measurements for: (A) Conjugation rate, (B) intermixing score, and (C) Transconjugant fraction $[T/(T+R)]$. (D) Population fractions for the experiment (Exp.; solid lines) and simulations (Sim.; dashed lines) for $f_{\text{conj}} = 0.2 \text{ h}^{-1}$. Error bars represent 95% confidence interval for $n = 17$ simulations with initial cell states set to initial conditions from the experiment.	143
Figure 6.11. Parametric sweep over the death rate parameter for the basic conjugation model with an additional parameter governing death rates of both donors and transconjugants (M1). Bar plots show simulated endpoint measurements for: (A) Conjugation rate, (B) intermixing score, and (C) Transconjugant fraction $[T/(T+R)]$. (D) Population fractions for the experiment (Exp.; solid lines) and simulations (Sim.; dashed lines) for $f_{\text{death}} = 0.2 \text{ deaths} \cdot \text{cell}^{-1} \cdot \text{h}^{-1}$. f_{conj} was held constant at 0.2 h^{-1} . Error bars represent 95% confidence interval for $n = 17$ simulations with initial cell states set to initial conditions from the experiment.....	145
Figure 6.12. Kernel density estimates for ABM parameters after each calibration round from the ABC-SMC algorithm.	147

Figure 6.13. Verification of predictions from M2 with calibrated parameter values. $n = 17$ simulations with initial cell states set to initial conditions from the experiment. Means for (A) conjugation rate, (B) intermixing score, (C) transconjugant fraction $[T/(T+R)]$, and (D) population fractions for the experiment (Exp.; solid lines) and simulations (Sim.; dashed lines). Dotted lines show experimental summary statistics plotted when recipients that have received the conjugative plasmid are considered as transconjugants prior to detecting fluorescence.	149
Figure 6.14. Simulated case studies for how different initial conditions in microfluidic traps influence end-point transconjugant fraction $[T/(T+R)]$, intermixing, and total conjugation events. Conditions that were tested varied were: (A) Initial D:R ratio, (B) initial number of cells loaded in the trap, and (C) sequential loading of different populations. Summary statistics presented are at 16 h after the simulation start time ($n = 5$ simulation replicates per condition).	151
Figure 6.15. Dimensions for bounding boxes of conjugative plasmid hosts (donors) and recipients.	164
Figure 7.1. An example of how a calibrated agent-based model of horizontal gene transfer in microbial communities could be used to make predictions about the efficacy of genetic bioaugmentation for plastic degradation. The degraded piece of PET from herbs packaging degraded by FAST-PETase (top left) has a pitted surface structure (bottom left) which could lead to local micro-environments for biofilm formation, conjugation, and additional plastic degradation. Scale bar: 10 μm .	175
Figure B.1. Window-averaged elongation rates for select cells after receiving the conjugative plasmid. Green dots indicate the cell was classified as a recipient; yellow dots indicate classification as a transconjugant. Number annotations indicate cell ID at the end of the lineage.	197
Figure B.2. Trackrefiner error tracking logs for pFAST-PETase-cis mating experiment. (A) Number of tracking errors that occurred at each time point. (B) Number of tracking errors that occurred at each depth in microfluidic traps. Errors are summed over all traps ($n = 17$). Any tracking errors that occurred within ~ 1 cell length (5 μm) of the trap exit were filtered out of the tracking error logs outputted by Trackrefiner.	198

List of Tables

Table 2.1. PETase variants characteristics and performance relative to wild-type PETase. ..	14
Table 2.2. ABM software available for modelling microbial communities.	26
Table 3.1. Selected transconjugant isolates identified by bITS sequencing.	36
Table 4.1. Summary Statistics for Quantifying Features in Microbial Communities	77
Table 4.2. Requirements for Generating Training Data for Microbial Community Models..	92
Table 5.1. CellModeller settings.	111
Table 5.2. ABC-SMC settings.	113
Table 6.1. Death rates for each population averaged over 16 h (192 time steps) across 17 traps.	139
Table 6.2. Parameters for probability distributions fit to single-cell properties.	142
Table 6.3. Statistics for third round posterior distributions from ABC-SMC calibration. ...	148
Table 6.4. Strains and plasmids used in this study.	155
Table 6.5. Microscopy imaging settings.	156
Table 6.6. CellModeller simulation settings.	164
Table 6.7. Prior distributions used in ABC-SMC calibration.	165
Table 6.8. Initialization settings in simulation case studies.	166
Table A.1. Primers used in Chapter 3.	192
Table B.1. Accuracy for detection of transconjugants with Ilastik.	193
Table B.2. Manual measurements of fluorescence intensity for recipients (<i>E. coli</i> DH5 α pEB1- gfpmut2) in trap # 12.	194
Table B.3. Manual measurements of fluorescence intensity for donors (<i>E. coli</i> NEB10- β pFAST-PETase-cis) in trap # 12.	195
Table B.4. Manual classification of cell growth state after receiving the conjugative plasmid.	196
Table E.1. State Variables in CellModeller.	202
Table E.2. CellModeller Outputs.	205

Table E.3. Default Attributes of the <i>CLBacterium</i> Class.....	206
--	-----

List of Abbreviations

Abbreviation	Meaning
A.U.	Arbitrary units
A ₂₄₀	Absorbance at 240 nm wavelength
ABC	Approximate Bayesian Computation
ABM	Agent-based model
aPET	Amorphous PET
BHET	bis-(2-hydroxyethyl) terephthalate
bITS	Bacterial internal transcribed spacer
CRISPR	Clustered regularly interspaced short palindromic repeats
D	Donor
DNA	Deoxyribonucleic acid
EG	Ethylene glycol
gRNA	Guide RNA
LB	Luria-Bertani
MHET	mono-(2-hydroxyethyl) terephthalate
OD ₆₀₀	Optical density at wavelength of 600 nm
ODD	Overview, Design concepts, and Details
ODE	Ordinary differential equation
PBS	Phosphate-buffered saline
PCA	Principle component analysis
pcPET	Post-consumer PET
PCR	Polymerase chain reaction
PDE	Partial differential equation
PE	Polyethylene
PET	Polyethylene terephthalate
POM	Pattern-oriented modelling
PP	Polypropylene
PS	Polystyrene
PU	Polyurethane
PVC	Polyvinyl chloride
R	Recipient
RM	Restriction modification
RNA	Ribonucleic acid
s.d.	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEM	Scanning electron microscopy
SMC	Scheduled Monte Carlo
SSE	Sum of squared errors
T	Transconjugant

T4P	Type IV pili
T4SS	Type IV secretion system
T6SS	Type VI secretion system
TDA	Topological data analysis
TPA	Terephthalic acid
WT	Wild-type

Quote

“If you don’t lose, you cannot enjoy the victories. So, you have to accept both things.”

-Rafael Nadal

Chapter 1

Introduction

1.1 Background

The discovery of methods to incorporate recombinant DNA into bacteria in 1973 ¹ sparked interest in genetically engineering single microbial species to perform novel tasks, such as engineering *E. coli* to produce insulin in 1979 ², and more recently, engineering baker's yeast to produce a precursor to an antimalarial drug in 2006 ³. Metabolic engineering of individual microbial strains remains a popular approach in industry and academic research, but the field is expanding in scope to engineer microbial consortia through assembling multiple strains or species of dynamically interacting microbes, or through *in situ* manipulation of pre-existing microbial consortia. More generally, design and manipulation of microbial communities to perform desired functions is known as microbiome engineering ⁴. This expansion was motivated by the benefits of having tasks executed by a mixed microbial community, including reduced metabolic burden associated with division of labour, isolation of biochemical pathways, extended metabolic capabilities, and improved robustness to environmental and biological perturbations ^{5,6}.

1.1.1 Microbiome engineering through biological augmentation

There are two complementary approaches to engineering microbial consortia: bottom-up, and top-down ⁴. The bottom-up approach uses mechanistic models of interactions between microbes to predict the overall functionality of the consortium. Consortia designed using the bottom-up approach tend to be relatively simple and are typically composed of only 2-4 distinct microbial strains; these tend to be suitable in controlled environments employed in industrial biotechnology applications. The bottom-up approach is constrained by the limited availability of experimental data on metabolism of individual strains and interactions between microbes. In contrast, the top-down approach relies on observations of more complex consortia. Experiments typically involve perturbations in environmental variables (e.g., substrate

availability, temperature, pH) to discern how changes in physical and chemical features impact the desired function of the consortium. This approach neglects the metabolic intricacies associated with individual consortium members, which restricts the improvements that can be achieved through mechanistic insight. This approach is applicable in situations where there is a microbiome already present in the environment at hand; relevant applications include agriculture, wastewater treatment, bioremediation, and healthcare.

One approach that falls under the top-down category is biological augmentation (bioaugmentation), whereby a biological agent is introduced to change the composition and/or function of the microbial consortium under investigation. Bioaugmentation can be classified into two types: cell bioaugmentation and genetic bioaugmentation. Cell bioaugmentation involves introduction of exogenous microbes to perform the desired task, whereas genetic bioaugmentation involves delivering genetic material to native microbes that will express genes that are carried on the genetic payload. With cell bioaugmentation, the function of interest persists only as long as the introduced organisms survive, which is a common source of failure in cell bioaugmentation interventions ⁷. The genetic bioaugmentation approach allows the function of interest to exist in multiple native species that are already adapted to their environment. This helps the introduced function persist through bypassing establishment barriers that exogenous microorganisms would otherwise face ⁸. Bioaugmentation provides a targeted approach for manipulating microbial consortia by enabling addition, removal, or modulation of functionalities in the consortium at hand (Figure 1.1). This approach provides more flexibility in comparison to less-specific microbiome manipulation methods such as modulating abiotic conditions, adding new substrates, or using antibiotics.

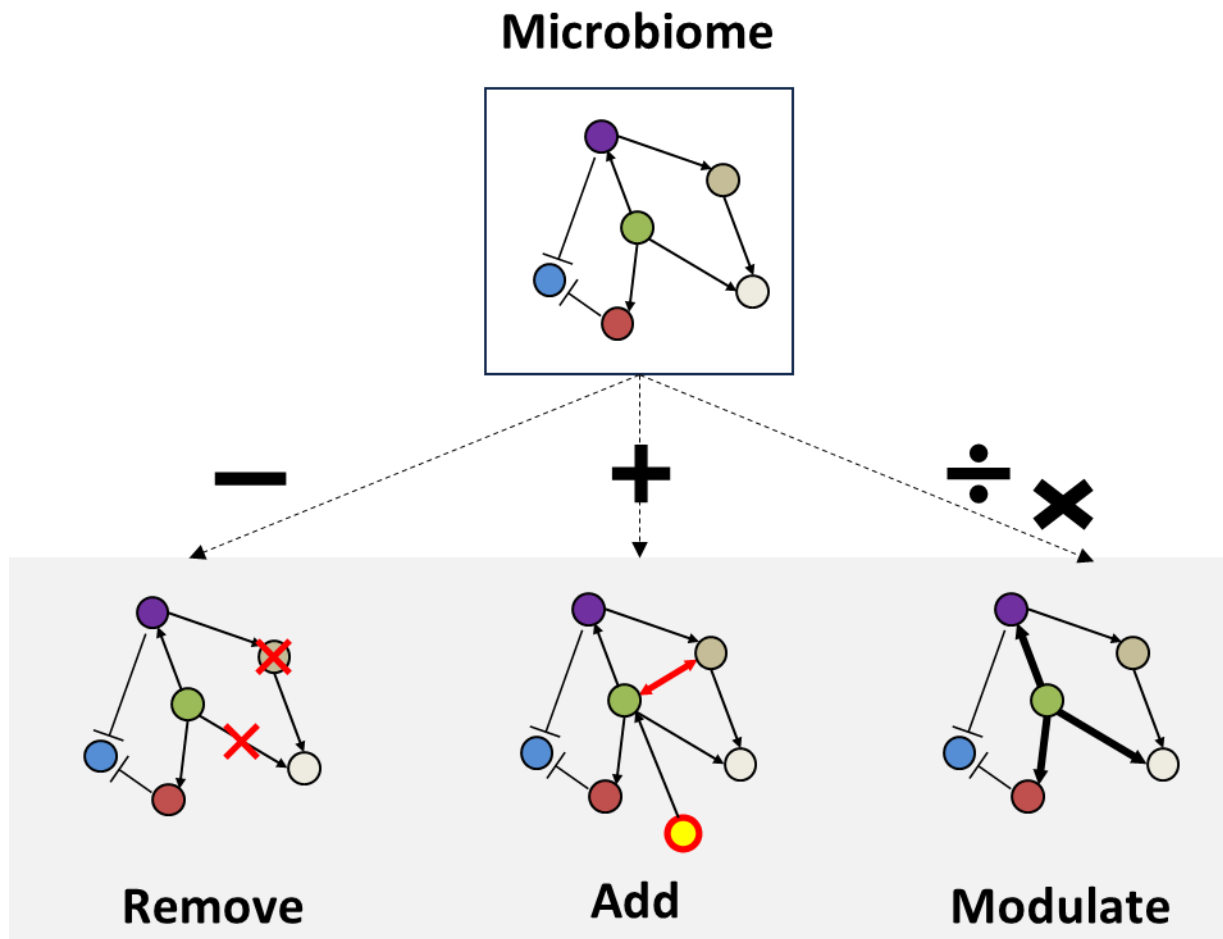


Figure 1.1. Operations on microbiomes enabled by bioaugmentation depicted using a cartoon microbial interaction network. Nodes represent microbial species or guilds of microbes performing similar functions. Arrows indicate a positive ecological interaction (e.g., cross-feeding through metabolite secretion), while barred lines indicate a negative ecological interaction (e.g., bacteriocin production). Microbiomes can be manipulated by adding in new microbial species or by delivering new genes to resident members, removing certain species or genes, and modulating ecological interactions.

Regarding addition of novel functionalities, bioaugmentation has often been applied to environmental remediation ^{9,10}. Some examples include the addition of *Dehalococcoides* organisms to contaminated waters to accelerate degradation of chlorinated hydrocarbons ¹¹, use of genetic bioaugmentation to degrade oil spills in soil ¹², and use of engineered microbes to degrade phenol released from mine tailings ¹³. Other examples include delivering genetic

material containing metabolic pathways to enhance nitrogen fixation and enabling consumption of substrates that a microbial consortium could not previously consume ^{14,15}. More recently, development of RNA-guided integrases (also known as CRISPR-associated transposon systems, or CASTs) has enabled sequence-specific genomic integration of genetic cargo into bacteria ^{15,16}, which enables precise editing of bacteria in a community context. Removal and modulation of organisms or genes with sequence-specificity can be achieved by delivering RNA-guided nucleases into microbial communities. Recent developments in using RNA-guided nucleases for microbiome engineering applications have been reviewed elsewhere ^{17,18}.

Genetic bioaugmentation requires deliberate release of engineered microorganisms and/or transgenic DNA into the environment, which raises concerns surrounding protection of the environment, biodiversity, and human health. More discussion about how to address these concerns is provided at the end of Section 1.1.4.

1.1.2 Methods for delivering DNA into microbiomes

There are three canonical mechanisms for delivering genetic material into bacterial hosts: transformation, transduction, and conjugation. Transformation is the direct uptake of DNA molecules from the host's environment. Transformation relies on stability of DNA in the environment, so it is more suitable for routine molecular cloning procedures than for *in situ* DNA delivery into microbiomes. Transduction uses viruses (i.e., bacteriophages) that infect bacteria. These bacteriophages can integrate into the host genome during the lysogenic cycle and produce more bacteriophage during the lytic cycle. Bacterial conjugation is a physical contact-based mechanism of transferring circular pieces of extra-chromosomal DNA (i.e. plasmids) between a donor and recipient cell. Additional background about conjugation will be given in Chapter 2.

The suitability of conjugation and transduction for *in situ* microbiome manipulation depends on the size of the genetic payload, the target species, and the physical environment. In bacteriophages, the size of the genetic payload is constrained by the size of the capsid ¹⁹,

whereas conjugative plasmids do not have a known size constraint for the payload—there are known conjugative plasmids that are larger than 420 kilobase-pairs ²⁰. Bacteriophages tend to have narrower host-ranges compared to plasmids, whereas the host-range for conjugative plasmids is highly flexible and can be tuned by selecting one or more origins of replication and replication initiation proteins ²¹. Some classes of conjugative plasmids are known to be able to conjugate to multiple classes of bacteria ²¹, and some broad host-range conjugation systems have also been reported to transfer between Gram-negative and Gram-positive bacteria ²². Some plasmids are also able to conjugate between kingdoms ^{23,24}. Conjugation is typically less efficient than bacteriophage delivery in planktonic conditions due to requiring a stable mating-pair to occur, but it is more efficient in biofilms where bacteria grow in close proximity ²⁵, which often occurs in relevant environments like the human gut and soil. For applications where engineering the largest diversity of bacteria is desirable (e.g., production of a therapeutic, environmental remediation), conjugative plasmids are likely better-suited. Furthermore, the metabolic load of conjugative plasmids pose a lower fitness cost to the host compared to bacteriophages, which hijack the host's cellular machinery to produce more bacteriophage at the expense of the host's viability. Transduction does have applications in microbiome engineering such as for phage therapy and delivery of precision antimicrobials ^{26,27}, but considering the above discussion, conjugation will remain the focus of this thesis.

1.1.3 Predicting efficacy of genetic bioaugmentation mediated through conjugation

The impact of a delivered function scales with the number of cells executing the function. Hence, high conjugation frequency and persistence of the transconjugants for the intended duration of a bioaugmentation treatment are desirable for introducing engineered functions into microbial communities. Conjugation frequencies for a given plasmid are most often measured *in vitro* between a donor and recipient strain, either through filter-mating or mating in suspension. While such measurements give some indication of how often a plasmid will transfer between two different hosts, these mating conditions do not capture the physical and biological complexities of real environments. Notably, it has been shown that conjugation

frequencies measured *in vitro* are considerably different than *in vivo* ²⁸, highlighting the challenge with predicting conjugation dynamics in more complex environments.

Computational models offer an alternative means to test *in situ* microbiome treatments by enabling rapid experimentation over large design spaces and exploring quantitative hypotheses. Although computer models are also an abstraction of real processes, the utility of validated models built on first principles and empirical measurements has been demonstrated in many areas of engineering, such as designing chemical unit operations, designing bridges, and predicting fluid flows. By the same note, having computational tools to predict gene delivery into microbiomes could facilitate with plasmid design, bacterial host selection, and reduce experimental testing times and costs by predicting efficacy of proposed genetic bioaugmentation treatments. A validated mechanistic model that predicts conjugation in spatially complex environments will help characterize mechanisms driving plasmid spread through a microbial community. Such a model would also create opportunities to rapidly explore questions about the impact of spatial heterogeneity and other cellular processes on the spread of conjugative plasmids.

1.1.4 Conjugation in context of the global plastic pollution problem

Thus far, I have discussed about conjugation through a microbiome engineering perspective. Here, I conjecture how conjugation could be leveraged to help alleviate the global plastic pollution problem. To provide some context about the problem, plastic waste only continues to accumulate in the environment due to the ubiquity of plastics in consumer products and their recalcitrance to biodegradation. Currently, plastic removal in the environment relies on government and volunteer efforts, which cannot compete with the rate of plastic waste production and is infeasible for large-scale collection. Furthermore, these efforts are ineffective at removing microplastics (defined as plastic particles less than 5 mm in diameter), which pose threats to wildlife and human health. More background on this subject is provided in Chapter 2.

To connect the topics of plastic pollution and conjugation, plastic pollution provides surfaces for biofilms to form on, which promotes conjugation by creating proximity between cells in aqueous environments ^{29,30}. Reactive oxygen species and other chemicals that adsorb onto plastics also invoke stress responses that increase conjugation frequencies ^{31,32}. What if we could leverage these heightened conjugation frequencies to deliver plastic-degrading functionalities into microbiomes for rapid, autonomous biodegradation of plastics? Numerous organisms and microbial communities have been discovered to be capable of consuming various commodity plastics ³³, suggesting that there are biological pathways to convert plastics into other products. The idea of engineering microbiomes for directly removing plastics in the environment was recently explored by Nguyen et al. ³⁴, where they discuss the idea of performing this feat in open oceans. Evidently, implementing this strategy in the open environment would require predictive models to estimate the impacts of the treatment, ecological risk analysis, and numerous controlled pilot studies. However, if the idea were to work, it would help prevent accumulation of plastic waste and microplastics in the environment. It could also move society towards a circular bioeconomy, where plastic waste would be upcycled into more valuable products through metabolic engineering.

Bioaugmentation requires deliberate release of microorganisms or other genetic material into the environment. There are two major concerns with this: 1) the fate of the engineered microorganisms and transgenic DNA in the environment, and 2) the ecological and environmental impacts of the released organisms. Regulatory frameworks and risk assessment strategies for handling environmental release differ by country, but generally, they require assessing the biology of the microorganism (e.g., taxonomy, pathogenicity, antimicrobial susceptibility, etc.) and the nature of the genetic modification, understanding the fate of the organism and its genome in the environment, and assessing the environmental and ecological impacts caused by the release ³⁵. One possible pathway for responsibly collecting data is through staged, controlled releases: in a process analogous to how new drugs are approved through the clinical trial process, the size and complexity would scale incrementally from petri dishes to defined laboratory habitats to field studies ³⁶. A standardized pipeline does not seem to exist yet for engineered microbes intended for environmental release, which poses additional

technical and financial risks to bringing bioaugmentation technologies to market. Collaborative efforts between regulatory agencies, academics, Indigenous communities, and the general public will be necessary to develop frameworks and standards for deployment of engineered microorganisms in the environment.

1.2 Thesis Outline

Discoveries of biological pathways for plastic consumption have opened opportunities for engineering microbiomes to degrade or upcycle plastics. Furthermore, there is no doubt that use of bacterial conjugation for *in situ* microbiome engineering will benefit from the availability of predictive models that can simulate conjugation in complex environments representative of those encountered in the field and clinic. Such models will assist with designing manipulations to microbial communities through genetic bioaugmentation. Therefore, the objectives of this thesis are to deliver plastic-degrading capabilities into an environmental microbiome through conjugation and to develop a model that characterizes conjugation in spatially heterogeneous environments, allowing for prediction of outcomes when manipulating microbial populations using conjugative plasmids. More specifically, this work aims to: 1) demonstrate manipulation of bacterial isolates from an environmental sample to degrade PET plastic, 2) develop a pipeline to calibrate spatio-temporal computational models to experimental data, and 3) characterize and predict gene transfer dynamics of conjugative plasmids in structured micro-environments.

In Chapter 2, an overview is given for biological degradation of plastics, bacterial conjugation, and agent-based modelling of microbial communities. For plastic degradation, an overview of advances in PETase engineering and applications is provided. For bacterial conjugation, molecular mechanisms will be described for canonical plasmids, followed by methods for measuring conjugation efficiencies.

In Chapter 3, a proof-of-concept for engineering environmental microbiomes to degrade PET plastics is demonstrated. The chapter addresses the feasibility of using engineered

bacteria from the environment to degrade PET plastic. This work appeared as a peer-reviewed article in the journal *Microbial Biotechnology*.

In Chapter 4, model calibration methods for spatio-temporal models of microbial communities are surveyed. This survey addresses the motivation for developing more standardized methods of calibrating computational models of microbial communities, provides a catalogue of relevant summary statistics describing spatio-temporal patterns commonly observed in microbial communities, and serves as motivation for Chapter 5. This work appeared as a peer-reviewed article in the journal *PLoS Computational Biology*.

Chapter 5 presents a pipeline to calibrate an agent-based model of bacterial micro-colony growth to microscopy data using Approximate Bayesian Computation, a reproducible and quantitative model calibration technique highly suited for stochastic models. This work appeared in the peer-reviewed proceedings of the 10th IFAC International Conference on Foundation of Systems Biology in Engineering.

Chapter 6 investigates effects of spatial constraints on conjugation dynamics. A mating experiment in a microfluidic environment is demonstrated to assess conjugation efficiency in a controlled and geometrically constrained environment. This experiment is then used to calibrate an agent-based model for conjugation. The calibrated model is then used to characterize how initial conditions such as donor-to-recipient ratio, cell density, and priority effects influence gene delivery outcomes. This chapter addresses the need for tools that can predict conjugation dynamics in more complex environments.

Finally, Chapter 7 discusses applications and limitations of modelling bacterial conjugation dynamics and biodegradation of PET, highlights opportunities for future work in advancing the tools developed in herein, and summarizes the overall contributions of the thesis.

Chapter 2

Literature Review

This chapter provides background information about the core topics in this thesis. First, the global plastic waste problem is introduced, followed by a review of biotechnological approaches for degrading plastics with a focus on PET. Then, an overview of the mechanisms behind bacterial conjugation is provided, followed by a review of techniques for measuring conjugation rates. To finish, a short introduction of modelling microbial communities is given, with a focus on agent-based modelling.

2.1 Global plastic waste problem

The global plastic waste problem stems from the linear economy *status quo* combined with the recalcitrance of petroleum-based plastics to biodegradation. Approximately 79% of all plastics ever made have been discarded into landfills or the environment ³⁷, and these plastics are expected to persist for hundreds of years due to their recalcitrance to biodegradation ³⁸. Plastic pollution poses an ecological and health problem due to their tendency to fragment into smaller, microscopic pieces known as microplastics, which are defined as plastic particles with a diameter between 1 μm and 5 mm; particles smaller than 1 μm in size are classified as nanoplastics ³⁹. Microplastics can originate from direct release of small plastic particles into the environment (e.g., plastic beads in cosmetic products) or through weathering and photodegradation of larger plastic items into smaller pieces. *In vitro* studies have shown that microplastics can cause oxidative stress and inflammation in human cells and may also be cytotoxic and carcinogenic ⁴⁰. A plethora of research has also demonstrated that microplastics have numerous detrimental effects on aquatic organisms like impaired growth, developmental defects, increased inflammation, and neurotoxicity ^{41,42}. Microplastics also act as carriers that adsorb chemicals and heavy metals ⁴³. Hence, microplastics are contaminants of concern as they bioaccumulate throughout the food chain, and have now been found even in humans ^{44,45}.

Numerous efforts are being made to move towards a circular plastics economy, where plastic waste is used to regenerate new products. Recycling is one of the most well-known strategies to return plastics back into their product life cycle, but only ~9% of all plastics have been recycled globally ³⁷ (the rate is low due to problems such as inefficiencies in collection, engineering challenges in separating different plastic types, and additives like dyes and flame retardants that don't separate during melting ⁴⁶). Using alternative materials that are biodegradable is another potential solution to addressing plastic waste accumulation. Biodegradable plastics made up about 1% of total plastic production in 2023 ⁴⁷, and although bioplastics are projected to take up a larger share of the market in the future, the timeline for when they will become economically and functionally competitive with petroleum-based plastics is uncertain. Government policy can impact both plastic use and waste management. It has been projected that reductions in plastic pollution from improvements to collection and disposal, recycling, and substitution in materials (in addition to meeting all global commitments announced prior to June 2019) will reduce global rates of plastic entering the environment by 78% compared to a “business as usual” scenario by 2040 ⁴⁸.

2.2 Biological degradation of plastics

Biotechnology offers potential solutions for sustainable conversion of plastic waste into inert and/or more valuable products. Microbes that can use paraffin as a carbon source were first reported by Fuhs in 1961 ⁴⁹. Since then, numerous bacteria and fungi have been discovered that can consume petroleum-based plastics, several of which are reviewed by Chattopadhyay ³³. To date, 753 microorganisms and 219 proteins with plastic-degrading capabilities have been documented in the PlasticDB database ⁵⁰. Out of all petroleum-based plastics, biodegradation of PET is currently the most advanced due to the discovery of a two-enzyme system in *Ideonella sakaiensis* 201-F6 that is active at mesophilic temperatures ⁵¹, and discovery of leaf compost cutinase that is active at thermophilic temperatures ⁵². Further developments in enzymatic PET degradation have led to the current construction of the world's first industrial-scale enzymatic PET recycling plant by Carbios, which is expected to be operational by 2025. Next, I provide an overview for how *I. sakaiensis* biodegrades PET, applications of protein

engineering for improving PET-degrading enzymes, and advances in genetically engineering microbes to biodegrade PET.

2.2.1 PET degradation mechanism in *Ideonella sakaiensis*

Ideonella sakaiensis is a gram-negative, aerobic bacterium that can use PET as a carbon and energy source. The organism was discovered in 2016 in a PET-contaminated soil sample from a PET recycling facility in Sakai, Japan ⁵¹. The PET degradation mechanism in *I. sakaiensis* strain 201-F6 is a two-enzyme pathway that breaks down PET into its constituents, terephthalic acid (TPA) and ethylene glycol (EG) (Figure 2.1), at mesophilic temperatures ^{51,53}. Depolymerization of PET begins by hydrolysis from PETase, which cleaves ester bonds in the polymer chain and yields mono-(2-hydroxyethyl) terephthalate (MHET) (the major product) and smaller quantities of TPA and bis-(2-hydroxyethyl) terephthalate (BHET). BHET is also hydrolyzed by PETase, yielding MHET and EG. MHET is further hydrolyzed into TPA and EG by the MHETase enzyme. *I. sakaiensis* is able to use TPA as a carbon source by first converting TPA into protocatechuic acid ⁵¹, followed by conversion into succinyl-CoA and acetyl-CoA, which are used in the tricarboxylic acid cycle ^{51,54}.

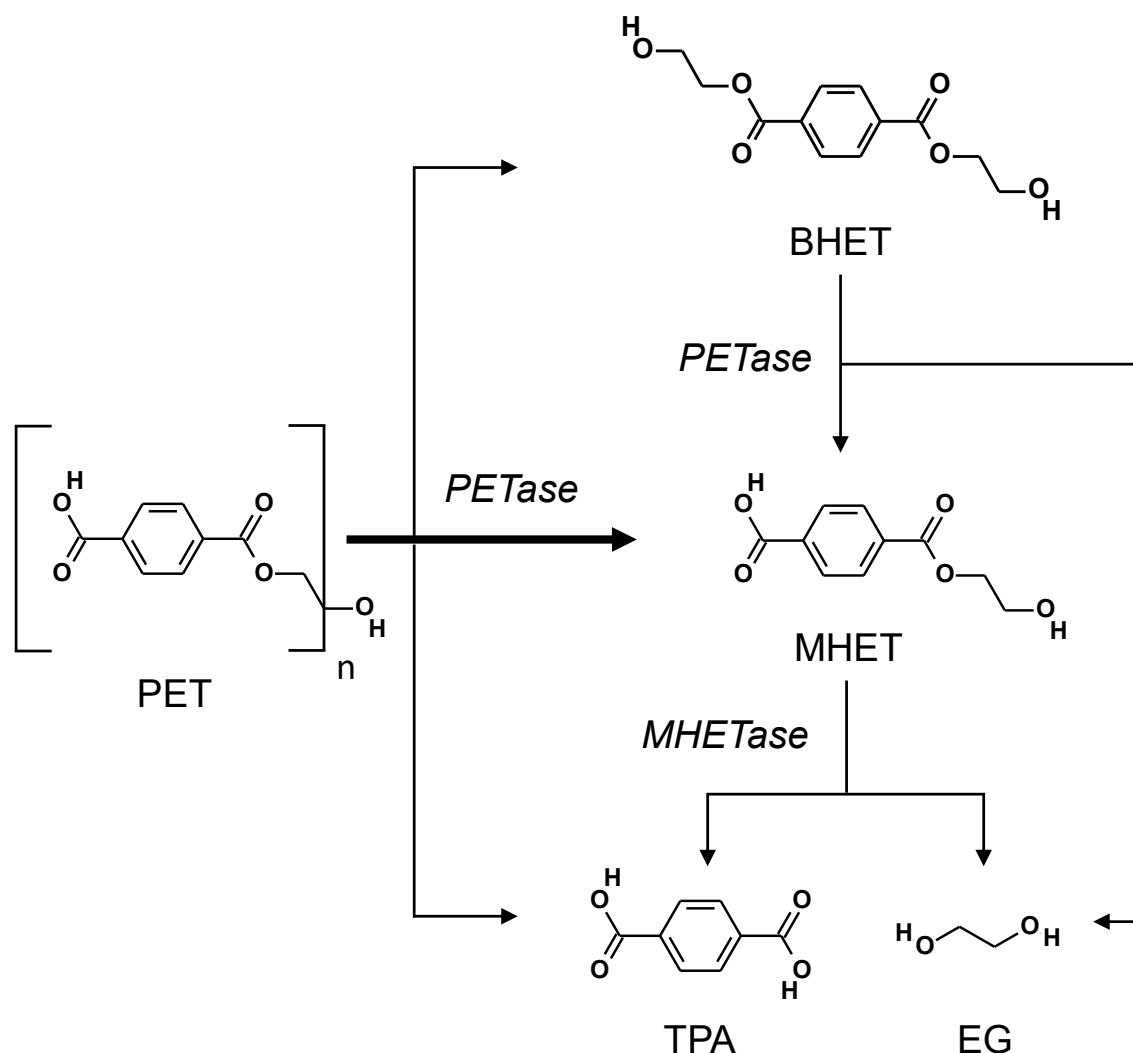


Figure 2.1 Two-enzyme pathway for degradation of polyethylene terephthalate (PET) into its constituent molecules in *Ideonella Sakaiensis* 201-F6. Depolymerization of PET by PETase primarily yields mono-(2-hydroxyethyl) terephthalate (MHET). Smaller amounts of bis-(2-hydroxyethyl) terephthalate (BHET) and terephthalic acid (TPA) are produced in the depolymerization step. BHET is further hydrolyzed into MHET and ethylene glycol (EG) by PETase. MHET is hydrolyzed into TPA and ethylene glycol (EG) by MHETase.

BHET, TPA, and EG are produced in the PETase reaction due to the different binding arrangements that PETase can take on the ends of the polymer chain⁵³. When a PET polymer chain is cleaved by PETase, two PET chains are formed: one with a TPA-terminal, and one with a HE-terminal. The favored conformation of PETase at the TPA-terminal produces one TPA molecule and an HE-terminal PET_{n-1} molecule. Similarly, the favored conformation on

the HE-terminal produces one MHET molecule and another HE-terminal PET_{n-1} molecule. PETase binding in the reverse direction (the cleavage site facing inwards to the polymer chain) on either terminal results in production of various PET dimers and monomers 2-HE(MHET)₂, (MHET)₂, MHET, and BHET, which are further hydrolyzed into MHET, TPA, and EG by PETase. More specifically, EG is only produced from hydrolysis of BHET ⁵³.

2.2.2 Enhancements to PET degradation using protein engineering

The landmark discovery of PETase in *Ideonella sakaiensis* 201-F6 enabled the use of PETase for PET depolymerization at mesophilic temperatures. Since the discovery, several variants of the PETase enzyme have been created through protein engineering for increased activity and thermostability (Table 2.1). Protein engineering involves finding optimal combinations of amino acid mutations that increase enzyme activity and stability. This design space was initially explored through rational design (ThermoPETase, DuraPETase). More recently, directed evolution (HotPETase and DepoPETase) and artificial intelligence-guided approaches (FAST-PETase, TurboPETase) have been employed to accelerate exploration of the design space for engineering PETase. The best performing PETase variants exhibit activities that are two to three orders of magnitude higher than the wild-type PETase at their optimal operating temperatures. Protein engineering has also been used to develop a MHETase variant (KL-MHETase) with improved thermostability, where a 96-fold improvement in activity was achieved relative to the wild-type MHETase at 50 °C ⁵⁵.

Table 2.1. PETase variants characteristics and performance relative to wild-type PETase.

Name	Increase in Activity w.r.t. WT PETase	Optimal Temperature (°C)	Melting Temperature(°C)	Reference
ThermoPETase	14-fold	40	57.62 (+8.81 to WT)	56
DuraPETase	25-fold ^a	50	77	58
FAST-PETase	143-fold	50	67.1	57

^a Relative activity from ⁵⁷

HotPETase	49-fold ^b	40	82.5	59
DepoPETase	1407-fold	50	72	60
TurboPETase	272-fold ^c	50	84	61

A different strategy that increases the conversion of PET into TPA and EG is to create a dual enzyme system by using a peptide linker to connect the PETase and MHETase enzymes. The depolymerization rate is improved in the dual enzyme system for two reasons. First, MHET inhibits PETase activity through product inhibition; MHETase converts MHET into TPA and EG, thereby reducing this inhibition effect ⁶². Second, linking the two enzymes increases proximity between active sites of the two enzymes and can promote channeling of intermediate molecules, thereby reducing diffusion limitations ⁶³. This strategy was originally found to increase depolymerization activity 6-fold compared to the PETase enzyme alone ⁶⁴. More recently, a dual enzyme system consisting of FAST-PETase and KL-MHETase was developed to enable operation at 50 °C, resulting in a 2.6-fold improvement in PET depolymerization rate compared to the FAST-PETase enzyme alone ⁵⁵.

2.2.3 Engineering microbes to degrade PET

Advances in enhancing the catalytic activity of PETases has enabled enzymatic recycling of PET at industrial-scale (e.g., Carbios). Costs could be further reduced by removing the need for protein purification, which is achievable by conducting the depolymerization reaction simultaneously with the microbial culturing step. Engineering of microbes to depolymerize plastics without the need for protein purification also opens opportunities for autonomous bioremediation of plastic pollution in the environment. Aside from expression in *E. coli* ⁶⁵, PETases have been expressed in other model organisms like *Bacillus subtilis* ⁶⁶, *Saccharomyces cerevisiae* ⁶⁷, *Chlamydomonas reinhardtii* ⁶⁸, and *Phaeodactylum tricornutum* ⁶⁹. The caveat with using the aforementioned organisms for biodegradation applications is that they grow at mesophilic temperatures, whereas the most efficient PET-degrading enzymes

^b 3.5-fold increased activity compared to ThermoPETase ⁵⁹.

^c 1.9-fold increased activity compared to FAST-PETase ⁶¹.

perform optimally at thermophilic temperatures. Yan *et al.* ⁷⁰ engineered the model thermophilic organism *Clostridium thermocellum* to express Leaf-Branch Compost Cutinase (LCC), which depolymerizes PET optimally at 68 °C ⁷¹. When they grew their engineered *C. thermocellum* with commercial PET film at 60 °C, they reported PET degradation rates almost double than rates reported for mesophilic organisms, which they attribute to the increased enzyme activity at the elevated operating temperature ⁷⁰. Engineering of thermophilic microbiomes for PET degradation (and other plastics) could be a promising avenue to improve efficiency of plastic degradation in environments with elevated temperatures such as in compost, which can contain thousands of microplastic particles per kilogram of compost dry weight ⁷².

Anchoring the enzyme on the cell surface improves enzyme stability and enables re-use of the cells as whole-cell biocatalysts ⁷³, which is desirable for reducing costs of enzymatic recycling at industrial-scale. Cell surface display has been previously used to immobilize enzymes involved in plastic degradation, with studies reporting 5- to 36-fold increases in depolymerization efficiency compared to the free enzyme ^{67,74,75}. Co-displaying PETase with hydrophobic adhesive proteins has been shown to further improve depolymerization rates due to increasing the adsorption rate to the hydrophobic plastic surface ⁷⁶. The cell surface display and co-display systems also showed improved efficiency towards higher crystallinity PET ^{74,76}, which is normally resistant to depolymerization from free PETases ⁷⁷. Another display system that immobilized PETase to bacterial curli fibers also showed similar improvements and maintained functionality in wastewater conditions ⁷⁸.

2.2.4 Engineering microbial communities to degrade PET

There are few studies that have investigated engineering of microbial communities to biodegrade PET. Qi *et al.* ⁷⁹ engineered a four-strain synthetic microbial community to convert PET into TPA and EG. The consortium was composed of *Rhodococcus jostii*, *Pseudomonas putida*, and two *B. subtilis* strains. The two *B. subtilis* strains were each engineered to secrete PETase and MHETase, *R. jostii* was added to degrade TPA, and *P. putida* was added to degrade EG. They reported improved PET degradation rates with their four-member consortium

compared to the two strain *B. subtilis* consortium in culture conditions. They attribute this improvement to removal of hydrolysate products TPA and EG, thus alleviating the inhibitory effects on PETase. It should be noted that the PET degradation experiments performed by Qi *et al.*⁷⁹ were in nutrient-rich media, so it would be interesting to investigate whether their artificial consortium could grow in a minimal medium with PET as the sole carbon source. Gao *et al.*⁸⁰ constructed a three-member consortium from marine bacteria isolated from the Huiquan Bay (China) and demonstrated that the consortium grew better in minimal media with PET as the sole carbon source compared to minimal media without PET.

Findings from other studies have also shown that dividing TPA and EG metabolism between microbial community members yields improvements in cell growth. Roberts *et al.*⁸¹ screened an environmental sample for lipase and esterase activity and isolated strains with the highest activity measured through screening on different substrates. They constructed a five-member consortium of *Pseudomonas* and *Bacillus* species that could degrade PET and BHET and use their monomers as a sole carbon source. They confirmed the presence of synergy between consortium members by screening for metabolism of TPA and EG and they found that this synergy yielded improvements in cell growth compared to individual isolates. *Bacillus* species were converting BHET into TPA and EG and were able to use EG as a carbon source but not TPA. Meanwhile, only the *Pseudomonas* strains contained genes for TPA metabolism, and two of the *Pseudomonas* strains also contained genes for EG metabolism that were genotypically similar to the well-known strain KT2400. Enhanced cell growth was also observed. A different study by Bao *et al.*⁸² focused on synergistic TPA and EG degradation through metabolic engineering of two *Pseudomonas putida* strains to consume either EG or TPA. They demonstrated that their co-culture grew better in comparison to a single strain that consumed both EG and TPA, highlighting the improvements in PET monomer consumption arising from using a division of labour approach. Their study also proposed a two-step chemical/biological method for converting PET to biomass, where PET depolymerization was performed by alkaline hydrolysis and the resulting monomers were mixed into a minimal medium to provide a carbon source for metabolism. This method enables bypassing biological

depolymerization of PET (the slowest step in PET biodegradation) to generate utilizable carbon sources, which could have applications in industrial biological upcycling of PET.

2.3 Bacterial Conjugation

Bacterial conjugation is a horizontal gene transfer process where a donor transfers genetic material to a recipient through physical contact mediated by a sex pilus. This process was discovered by Tatum and Lederberg in 1946⁸³, which won them the Nobel Prize in Physiology or Medicine in 1958. Their discovery meant that some organisms had not evolved simply from genetic information being passed from parent to offspring (i.e. vertical gene transfer) as depicted in the “Tree of Life”, but that bacteria had evolved through simultaneous vertical and horizontal gene transfer processes. It also implies that bacteria can evolve and adapt to challenging environmental conditions through horizontal gene transfer. This mechanism has notably contributed to increased spread of antibiotic resistance genes and motivated fundamental research for understanding and predicting spread of conjugative plasmids⁸⁴. Next, we delve into an overview of conjugation mechanisms in Gram-negative bacteria, factors affecting conjugation efficiency, techniques for measuring conjugation efficiency, and applications in microbiome engineering.

2.3.1 Conjugation mechanism

Conjugation begins with expression of the *tra* genes (which encode all factors necessary to form a conjugative pilus), the Type IV secretion system (T4SS), and proteins to unwind and transport DNA from the host to recipient⁸⁵. DNA is nicked at the origin of transfer (*oriT*) by a conjugative relaxase. The strand to be transferred (i.e. T-strand) binds to a protein complex called the relaxosome. The relaxosome then binds to a Type IV coupling protein (T4CP) and initiates ssDNA transfer through the T4SS. The transfer process occurs simultaneously with rolling circle replication in the donor. In the recipient, a single-stranded promoter on the T-strand is recognized by the host RNA polymerase to initiate production of RNA primers. The strand is circularized, followed by conversion into dsDNA by the host DNA polymerase. Once converted into dsDNA, the new host can begin expressing genes carried on the plasmid.

The RP4 plasmid (also known as RK2) is a canonical IncP conjugative plasmid that is known to have a broad-host range and has been used for several *in situ* genetic engineering applications^{15,86,87}. Conjugation genes on RP4 are located in two regions that are approximately 10-15 kb each in size: Tra1 and Tra2. Genes located in Tra1 encode functions necessary for DNA transfer and replication, including the relaxosome, while genes located in Tra2 are responsible for mating-pair formation⁸⁸.

Regulation of conjugation expression in conjugative IncP plasmids relies on plasmid-encoded factors that repress expression of transfer genes. The global regulators are KorA, -B, -C, and TrbA. The gene regulation network is rather complex and described in detail by Zatyka and Thomas⁸⁹. To summarize, KorA and KorC are involved in regulating plasmid replication and partitioning and help coordinate them with transfer processes^{90,91}. Meanwhile, expression of transfer genes is tightly regulated by KorB and TrbA⁹². Together, KorB and TrbA amplify repression of transfer gene expression, and small changes in these regulators can have big effects on change in expression⁹³. A key factor enabling low basal expression of transfer genes during tight regulation is that the sequences of TrbA binding sites are imperfect palindromes, resulting in reduced affinity of TrbA and therefore reducing sensitivity to the repressor. The impact of these binding sites on regulating conjugation was demonstrated by Bingle *et al.*⁹³, who converted these binding sites into perfect palindromes, resulting in increased sensitivity to repression by TrbA and a drop in 2-3 orders of magnitude in conjugation frequency compared to the wild-type plasmid⁹³. This opens the possibility of achieving an upregulated conjugation state by modifying TrbA binding sites to be less palindromic. Overall, the gene regulatory network results in increased expression of conjugation genes during plasmid establishment in a new host due to lack of regulatory proteins present (i.e., transitory de-repression), followed by steady repression to low levels of expression as regulatory proteins accumulate^{89,94}. This regulation mechanism reduces the metabolic burden associated with over-expression of transfer genes, which are lethal to the host when expressed in excess⁹⁵.

2.3.2 Factors influencing spread of conjugative plasmids

The rate at which a conjugative plasmid spreads is a key factor determining the efficacy of genetic bioaugmentation. This is determined by how often the donor expresses *tra* genes in presence of a recipient, how often stable mating pairs can be formed with the recipient, and the ability of the recipient to take up and maintain the plasmid. Here, we review relevant factors associated with gene regulation in the donor, cellular metabolism, taxonomic relatedness of the donor and recipient, and environmental conditions.

Based on the mechanism of transitory de-repression of conjugation genes in IncP conjugative plasmids, it was initially hypothesized that a conjugative plasmid could spread through an entire population in a wave-like manner ⁹⁶. However, evidence from experiments mating donors and recipients in colonies showed that a conjugative plasmids only transferred to the outer layer of cells ⁹⁷. Numerous works then demonstrated that conjugation rates for certain plasmids were correlated to the growth rate of donors and recipients ^{98–101}, suggesting that increased cellular growth made donors more likely to transfer the plasmid. In the F-plasmid, this occurs due to transcriptional and post-translational regulation in the transfer genes ^{91,102}; this regulation is associated with general regulators in the host metabolism (e.g., cAMP) that modulate expression of genes involved in conjugation ^{103,104}. Growth stage-dependent transcriptional regulation in the donor was also found to occur in other conjugative plasmids belonging to classes IncF, I, N, and X ¹⁰³. In contrast, the RP4 plasmid does not experience growth stage-dependent regulation of conjugation genes ¹⁰³, which could be due to regulation of transfer genes solely from plasmid-encoded factors.

For a conjugative plasmid to spread, the recipient bacteria must be capable of maintaining and replicating the plasmid. Features enabling plasmids to replicate in a broad range of hosts (i.e. broad-host-range plasmids) include having multiple origins of replication and replication initiation being independent of host factors, and avoidance of restriction modification (RM) systems ²¹. Empirical evidence indicates that RM systems can play a substantial role in defense against conjugative plasmids ^{105,106}, with recent reports showing between 0 to 10⁵-fold protection against uptake of diverse conjugative plasmids ¹⁰⁷. Evolution

of anti-restriction functions on conjugative plasmids also suggests that it is advantageous for conjugative plasmids to evade RM systems ¹⁰⁸, providing some guidance on how to design broad-host-range conjugative plasmids. Indeed, plasmid maintenance can be improved by avoiding restriction modification systems of the host, which can be bypassed by removing restriction sites *in vitro* ¹⁰⁹ and/or by using a donor strain that has a similar methylation profile to the recipient of interest ¹¹⁰.

While restriction modification systems are one reason why conjugation may be more efficient in species that are closely related, Alderliesten *et al.* ¹¹¹ performed a meta-analysis on conjugation efficiencies and found that conjugation efficiency decreased with lower taxonomic relatedness in liquid matings, but not in filter matings. The observed negative correlation between conjugation efficiency and taxonomic relatedness could be due to differences in cell wall structure arising from diverse lipopolysaccharide pathways between species and strains, in turn reducing the recipient's susceptibility to conjugation ^{112,113}. Alderliesten *et al.* ¹¹¹ suggest that stable mating-pair formation is a likely cause for the discrepancy between matings in suspension and filter matings, since on the filter, the cells remain approximately fixed in space and have more time to form a stable mating-pair with a given neighbour.

Further complicating the task of estimating plasmid spread, abiotic factors also play a role in conjugation efficiency. As mentioned previously, stable mating-pair formation is essential for conjugation to occur, and the physical conditions can determine the effectiveness of a conjugation system. Long, flexible pili facilitate formation of mating-pairs in suspension, whereas thick, rigid pili favour mating-pair formation on solid surfaces ¹¹⁴. Production of Type IV Pili (T4P) has also been shown to improve conjugation efficiency *in vivo* ²⁸ due to their role in attachment between bacteria ^{115,116}. Other environmental factors like temperature, pH, oxygen, salinity, total organic carbon content, cations, and heavy metals have also been shown to influence conjugation efficiencies ^{28,117–119}. The optimum conditions seem to depend on factors associated with the conjugation system and the host, suggesting that in general, abiotic conditions affect conjugation frequency through the physiology of the donor and the donor's response to environmental cues.

2.3.3 Quantifying bacterial conjugation using end-point methods

Conjugation frequency is a measure of how often a donor will transfer a plasmid on a per donor and recipient basis (units of donor⁻¹ recipient⁻¹ h⁻¹). This is different from conjugation efficiency, which is reported as a ratio of donor, recipient, and/or transconjugant population densities at a specific point in time. More than 10 different population density metrics for conjugation efficiency have been used in the literature ¹²⁰, which makes it difficult to draw comparisons across experiments. Furthermore, such measures do not distinguish plasmid conjugation from plasmid propagation due to cell division of already formed transconjugants (i.e. vertical gene transfer). Levin *et al.* ¹²¹ developed the first method to decouple horizontal and vertical gene transfer, which used a differential equation model to estimate the increase in the transconjugant population due to cell growth:

$$\frac{dN_R}{dt} = \psi N_R - \gamma(N_D + N_T)N_R \quad (2.1)$$

$$\frac{dN_D}{dt} = \psi N_D \quad (2.2)$$

$$\frac{dN_T}{dt} = \psi N_T + \gamma(N_D + N_T)N_R \quad (2.3)$$

where N is the population of recipients (R), donors (D), and transconjugants (T), t is time, ψ is the population growth rate, and γ is the conjugation rate. The model is relatively simple and captures essential features of conjugation dynamics. Based on this model formulation, Simonsen *et al.* ¹²² devised a formula to measure γ using initial and end-point population density measurements:

$$\gamma = \frac{\psi}{N_{tot} - N_0} \ln \left(1 + \frac{N_T N_{tot}}{N_R N_D} \right) \quad (2.4)$$

where N_0 is the initial population density, and N_{tot} is the end-point population density. The Simonsen *et al.* end-point-method has been the standard for measuring conjugation frequencies since its conception in 1990. The model from Levin *et al.* assumes that mass-action kinetics apply to mating pair formation, plasmid loss rate is negligible, all populations share the same growth rate, there is no delay between the time a transconjugant takes up the plasmid

and the time it can act as a new donor, and that the conjugation frequency is constant and is the same for donors and transconjugants. The assumptions of mass-action kinetics and a uniform growth rate across different populations are the most notable limitations of the Simonsen *et al.* method. Mating assays performed in structured environments (such as agar, filters, soil, and mammalian gut) violate the assumption of mass-action kinetics. Mating between bacterial species with different growth rates violates the assumption that all populations have the same growth rate. Additionally, the assumption of an equal conjugation frequency between donors and transconjugants does not account for transitory de-repression of conjugation in newly formed transconjugants ^{98,123}.

Recently, newer methods and guidelines have been proposed to address some assumptions that the Simonsen *et al.* method is based on ^{120,124,125}. The approximate Simonsen method proposed by Huisman *et al.* ¹²⁰ extends the classical Simonsen method by accounting for different growth rates in donor, recipient, and transconjugant populations. The Luria-Delbrück method proposed by Kosterlitz *et al.* ¹²⁴ treats conjugation as a stochastic process, and centers around experimentally measuring the probability that at least one conjugation event has happened after a fixed time, allowing decoupling of HGT and VGT of a plasmid in the transconjugant population ¹²⁴. In addition to these newer experimental approaches, Malwade *et al.* ¹²⁶ performed a comprehensive model-based assessment of more complex variations of Levin *et al.*'s model, finding that models of intermediate complexity were better suited for reproducing batch filter mating experiments. Nonetheless, these newer methods still rely on a spatially homogeneous experiment and assumptions on cell metabolic state ¹²⁵, which limits the experimental conditions where the conjugation frequency can be measured in. Furthermore, macroscopic methods for quantifying conjugation frequency cannot account for mechanisms occurring at single-cell level, such as the influence of contact area between the donor and recipient, state in the cell cycle, and transitory de-repression of conjugation ⁹⁸.

2.3.4 Observations of bacterial conjugation using microscopy

Uptake of a conjugative plasmid in recipients can be observed using fluorescence microscopy if the plasmid is tagged with a fluorescent marker. Transconjugants can be distinguished from the original donor cells by either tagging donors with a unique fluorescent marker or by repressing expression of the fluorescent marker on the plasmid in donor cells ¹²⁷. This technique has been used primarily for quantifying transconjugant populations ^{128–130}. Information on cell lineages becomes available when populations are enumerated in a time-series (i.e., time-lapse microscopy), which enables distinguishing between horizontal gene and vertical gene transfer. The ability to identify individual conjugation events opens opportunities to investigate how single-cell properties influence the likelihood of conjugation ⁹⁸.

Information about intracellular processes involved in conjugation can be visualized by binding fluorescent markers to the conjugative plasmid itself. For example, the localization of conjugative plasmids has been visualized using a LacO/LacI-GFP system, where a LacO cassette is introduced onto the plasmid and the LacI-GFP fusion protein binds to the lactose operator ¹³¹. In a similar manner, proteins involved in conjugation can be tagged with a fluorescent marker to investigate their intracellular localization throughout the conjugation process ¹³². ssDNA transfer has also been visualized using a chromosomally encoded single-strand-binding protein ¹³³, enabling immediate detection of conjugation events *in situ* and identification of the donor cell. This technique overcomes some shortcomings in the more conventional approach to measuring conjugation *in situ* using fluorescence microscopy, where conjugation is identified from recipients expressing a plasmid-encoded fluorescent protein. By marking ssDNA, it becomes possible to observe both the donor and recipient involved in a conjugation event, allowing clear identification of the donor strain. It also eliminates delays in observing when the conjugation event occurs by circumventing the need to wait for plasmid transfer, recircularization, fluorescent protein expression, and fluorescent protein maturation. This relatively novel technique will be beneficial for future studies of conjugation at single-cell level.

Aside from observing plasmid acquisition and intracellular processes involved in conjugation, fluorescence microscopy has also been used to visualize expression of conjugative pili. Clarke *et al.*¹³⁴ used a fluorescently marked R17 bacteriophage that binds to F-pili to determine that pili randomly extend and retract to scan for recipients around the host cell. Moreover, they observed the pili bringing donors and recipients together during mating-pair formation, highlighting the role of flexible conjugative pili in stabilizing mating-pairs. Some years later, Goldlust *et al.*¹³³ combined the approach of fluorescently labelling pili and ssDNA to observe transport of ssDNA through the F-pilus in scenarios where the donor and recipient were connected by the pilus at a distance, demonstrating that conjugative pili also serve as a transport channel for ssDNA to travel between the donor and recipient. Their study also yielded quantitative insights into how frequently pili are expressed, their length, and their localization on the cell, which could be useful for formulating computational models of the conjugation process.

2.4 Agent-based modelling of microbial communities

Models of microbial communities can be helpful for testing hypotheses to validate experimentally, or for understanding mechanisms that explain experimental results. Models can also be used for predictive purposes if they have been validated against experimental data. The three main classes of mechanistic models for modelling population dynamics in microbial communities are: ordinary differential equations, partial differential equations, and agent-based models (ABMs). A summary of the utility for each model class is given in Chapter 4, along with methods for calibrating spatiotemporal models of microbial communities to experimental data. Here, the focus is on providing an overview of how ABMs are formulated in the context of microbial communities and explaining the selection of the ABM framework that is used in subsequent chapters. A high-level explanation of the general architecture and implementation of ABMs of microbial communities is also available elsewhere¹³⁵.

ABMs have been defined as “models where individuals or agents are described as unique and autonomous entities that usually interact with each other and their environment locally”¹³⁶. In the context of microbial communities, ABMs are temporally discrete

simulations where individual cells are treated as agents that act according to a set of rules. The properties of each cell are given by state variables. State variables are the properties that are required to completely describe an agent. For simulating bacteria, relevant state variables include position, orientation, growth rate, morphology, intracellular molecule concentrations, etc. At each time step, state variables for each cell are updated according to rules chosen by the user, which can be based on mechanistic or empirical models. For example, cell elongation can be modelled by the exponential growth equation, whereas cell division can be modelled using a probability distribution derived from experimental observations.

Several tools have been developed for agent-based modelling of microbial communities and are summarized in Table 2.2. Cellular automata have not been included for the purposes of this work as they do not model individual cell behaviours. Each tool provides different features that others may not, which makes them somewhat specialized in simulating certain scenarios (this does not mean that they cannot be modified to expand their functionality). For example, CellModeller is useful for simulating large numbers of cells during colony expansion due to its compatibility with graphics processing units (GPUs), which are faster at performing parallel computation compared to central processing units (CPUs). Meanwhile, BSim 2.0 offers a toolbox for simulating quorum-sensing in constrained environments but is not compatible with computation with GPUs. Comparisons between some of these simulators have been tabulated elsewhere ^{137–139}.

Table 2.2. ABM software available for modelling microbial communities.

Software	Year Released	Language	Unique Features	Reference
iDynoMiCS	2011	Java	Biofilms	140
CellModeller	2012	Python, PyOpenCL	GPU computing	141
DiSCUS	2013	Python	Bacterial conjugation	142,143
BSim 2.0	2017	Java	Microfluidic environments	138

Simbiotics	2017	Java	Agent initialization from microscopy data, bacterial aggregation	139
Gro	2017	C++	Bacterial conjugation, genetic circuits	144
NUFEB	2019	C++	Biofilms, parallel computing	145
Vivarium	2022	Python	Flux balance analysis, parallel computing	146

In this work, CellModeller will be used for some practical reasons. The compatibility of CellModeller with GPU computing makes it an attractive candidate given the computational expense of spatiotemporal simulations of microbial communities. Additionally, in comparison to other software packages, CellModeller has been used extensively in spatio-temporal modelling of microbial communities, including antagonistic interactions between populations^{147–151}, biofilm development^{152,153}, cooperation and antibiotic resistance¹⁵⁴, horizontal gene transfer^{155–158}, cellular adhesion¹⁵⁹, and synthetic microbial communities^{160,161}. Although the original version of CellModeller lacks certain features to simulate bacterial conjugation and environments like microfluidic devices, these features have been developed since the original release. Details about ABMs investigating bacterial conjugation will be reviewed in Chapter 6.

Given that fundamental laws governing basic cellular behaviour such as rigid body dynamics, exponential growth, and molecular transport are universal, it is surprising that there is such a wide variety of ABM frameworks available. Currently, a “standard” ABM framework does not exist for modelling microbial communities, which scatters efforts in model and software development across the various platforms. This has likely occurred due to historical reasons, where sub-fields of microbiology developed tools with specific features relevant to the problems at hand¹³⁷. It would be beneficial for the scientific community to come together and develop a smaller set of more comprehensive modelling tools, which would reduce the need for expertise in software development, reduce software development time, minimize

repeating model validation, and improve reproducibility of simulated results. A set of desirable features would be:

1. Ability to simulate complex geometries and environments
2. Ability to model mass transport in the environment
3. Rules for cell behaviour and interactions, such as adhesion, signalling, contact-dependent inhibition, and horizontal gene transfer
4. Differential equations to model intracellular states
5. Control over stochasticity in cellular behaviour and interactions
6. Cell motility
7. Chemotaxis
8. Control over cell morphology
9. Parallel computing capabilities
10. Modular software architecture that enables usage of custom models
11. Documented validation for all sub-models against experimental data

The need for rigorous validation is especially important for using ABMs for design and predictive purposes. Methods to improve calibration of ABMs to spatio-temporal data are reviewed in further detail in Chapter 4, and Chapter 5 provides a pipeline to streamline calibration of ABMs to microscopy data.

Chapter 3

Degradation of polyethylene terephthalate (PET) plastics by wastewater bacteria engineered via conjugation

The work included in this chapter is published: A. Yip, O.D. McArthur, K.C. Ho, M.G. Aucoin, B.P. Ingalls. Degradation of polyethylene terephthalate (PET) plastics by wastewater bacteria engineered via conjugation. *Microbial Biotechnology*, 17 (2024), e70015.

There is a pressing need to develop technologies for removing microplastics from the environment, which pose threats to ecosystems and human health. In light of this problem, this chapter demonstrates delivery of enhanced plastic-degrading capabilities to an environmental microbiome using conjugation. This Chapter also introduces a gene delivery vector, pFAST-PETase-cis, that will serve as a model conjugative plasmid for investigating conjugation dynamics in structured microenvironments later in the thesis.

3.1 Abstract

Wastewater treatment plants are a major pathway for microplastics to enter the environment. In general, microplastics are contaminants of global concern that pose risks to ecosystems and human health. Here, we present a proof-of-concept for reduction of microplastic pollution emitted from wastewater treatment plants: delivery of recombinant DNA to bacteria in wastewater to enable degradation of polyethylene terephthalate (PET). Using a broad-host-range conjugative plasmid, we enabled various bacterial species from a municipal wastewater sample to express FAST-PETase, which was released into the extracellular environment. We found that FAST-PETase purified from some transconjugant isolates could degrade about 40% of a 0.25 mm thick commercial PET film within four days at 50 °C. We then demonstrated partial degradation of a post-consumer PET product over 5-7 days by exposure to conditioned

media from isolates. These results have broad implications for addressing the global plastic pollution problem by enabling environmental bacteria to degrade PET.

3.2 Introduction

Plastics are versatile materials that are lightweight, strong, and chemically-resistant. These desirable properties have made them an integral part of daily activities, infrastructure, and economies since their invention in the 20th century (Andrady and Neal, 2009). Although plastics have been an economic boon—particularly in packaging and construction applications—their use has produced unsustainable levels of waste. From 1950 to 2015, about 4900 megatons, accounting for 59% of all plastics ever produced, have been discarded into landfills and the environment (Geyer *et al.*, 2017). Ninety percent of plastics produced—consisting of polyethylene (PE), polypropylene (PP), polyvinyl chloride (PVC), polyethylene terephthalate (PET), polyurethane (PU), and polystyrene (PS) (Geyer *et al.*, 2017)—are projected to persist for hundreds of years; they are not biodegradable on timescales relative to their end of use (Chamas *et al.*, 2020). Consequently, their bioaccumulation poses threats to ecosystems and, potentially, to human health (Wang *et al.*, 2019; Xu *et al.*, 2020; Yuan *et al.*, 2022). Annual plastic waste generation is projected to almost triple by 2060 compared to 2019 levels, if trends in current plastic usage continue (OECD, 2022). Even with immediate implementation of ambitious strategies for waste management, environmental recovery, and reduction of plastic production, conservative estimates suggest that plastic waste entering terrestrial and aquatic ecosystems will exceed or remain close to 2016 levels between 2030-2040 (Borrelle *et al.*, 2020; Lau *et al.*, 2020).

Although PE, PP, PVC, PET, PU, and PS are referred to as non-biodegradable, some microbes and insects have ways to metabolize these plastics into their constituent molecules which can then be used as carbon and energy sources (Jeon and Kim, 2016; Yoshida *et al.*, 2016; Brandon *et al.*, 2018; Chattopadhyay, 2022; Maheswaran *et al.*, 2023). These natural processes are relatively slow, taking weeks to months for significant biodegradation to occur (Shah *et al.*, 2008; Sekhar *et al.*, 2016; Yoshida *et al.*, 2016; Gao and Sun, 2021; Kumar *et al.*, 2023; Maheswaran *et al.*, 2023). Among the most produced plastics, PET is a hydrolysable

polymer that is widely used in packaging and textile industries. Compared to other common plastics, PET is more amenable to biodegradation due to the presence of hydrolysable ester bonds. Substantial progress has been made in enzymatic PET depolymerization for resource recovery (Liu *et al.*, 2023). For example, PET hydrolase (PETase) from *Ideonella sakaiensis* has been subject to numerous protein engineering efforts that have resulted in PETase variants with depolymerization rates orders of magnitude higher than the wild-type enzyme, and with greater thermo- and chemostability (Son *et al.*, 2019; Cui *et al.*, 2021; Bell *et al.*, 2022; Lu *et al.*, 2022; Zurier and Goddard, 2023). Efforts have also been made to optimize expression and secretion of PET-degrading enzymes in some bacterial species other than *E. coli* (Wang *et al.*, 2020; Yan *et al.*, 2021; Li *et al.*, 2023), and developments are ongoing for employing naturally occurring and designer microbial communities to degrade PET (Roberts *et al.*, 2020; Gao and Sun, 2021; Qi *et al.*, 2021; Maheswaran *et al.*, 2023).

Effluent and sludge from wastewater treatment plants are major pathways by which microplastics enter the environment (Sun *et al.*, 2019; Liu *et al.*, 2021; Zurier and Goddard, 2021). Notably, a recent study reported that 6% of plastic processed at a recycling facility was found as microplastics in the facility's wash water post-filtration (Brown *et al.*, 2023). Secondary treatment processes and digestion of sludge offer opportunities for removal of microplastics through bioaugmentation. As reviewed by Zurier and Goddard (Zurier and Goddard, 2021), microbial communities could be supplemented with engineered bacteria that secrete plastic-degrading enzymes; this can be more cost effective than continuously supplementing a cocktail of purified enzymes into a secondary treatment unit.

A major challenge with the cell bioaugmentation approach is persistence of the introduced microorganisms, which are often not adapted to the environment at hand and are therefore likely to be quickly eliminated (Albright *et al.*, 2022; Ma *et al.*, 2022). As an alternative strategy that would bypass these barriers to establishment, we propose genetically engineering microbes native to the environment to express metabolic pathways for biodegradation of plastic waste. This approach, known as genetic bioaugmentation, can be achieved through *in situ* delivery of broad-host-range conjugative plasmids. Genetic

bioaugmentation has been shown to improve persistence of the introduced functionality when using donor strains native to their environment (Ronda *et al.*, 2019). Genetic bioaugmentation has not yet been applied for bioremediation of plastic waste, but this technique has been used to introduce catabolic genes into microbial communities for bioremediation of non-plastic-based pollutants in laboratory and pilot scale settings (Top *et al.*, 2002; Bathe *et al.*, 2004; Venkata Mohan *et al.*, 2009; Garbisu *et al.*, 2017; Ren *et al.*, 2018; French *et al.*, 2020; Ke *et al.*, 2022; Chettri *et al.*, 2023).

In this proof-of-concept study, we lay a foundation for the use of genetic bioaugmentation for bioremediation of PET. We genetically engineered bacteria in a municipal wastewater sample to degrade PET plastics by delivering a broad-host-range, conjugating plasmid carrying the FAST-PETase gene (Lu *et al.*, 2022), which codes for an engineered PET hydrolase that is more robust to pH and temperature ranges and is orders of magnitude more efficient than the wild-type enzyme (Lu *et al.*, 2022). We assessed degradation of commercial PET film using purified FAST-PETase from several engineered isolates. We then measured degradation of post-consumer PET using conditioned media from some of the engineered isolates. This report provides a foundation to support the next steps in assessing the performance and efficacy of such a genetic bioaugmentation approach.

3.3 Results

3.3.1 pFAST-PETase-cis conjugates to bacteria from wastewater

We constructed an RK2-based (Strand *et al.*, 2014) conjugative plasmid, pFAST-PETase-cis (Figure 3.1, Figure 3.2a), which carries the FAST-PETase gene (Lu *et al.*, 2022) under control of an arabinose-inducible promoter (P_{BAD}) (details in Methods). Our design of pFAST-PETase-cis retains the signal peptide SP_{pstu} and 6×His-tag from the backbone pBTK522::FAST-PETase (Lu *et al.*, 2022) (Figure 3.2b). These facilitate secretion and purification of FAST-PETase. We chose to drive FAST-PETase expression from the P_{BAD} promoter to achieve tunable gene expression for testing purposes. Moreover, this promoter has been shown to drive expression in a diverse range of gram-negative bacteria (Prior *et al.*, 2010). The plasmid carries a

fluorescent marker (*mCherry*) and two antibiotic resistance markers (Amp^R, Gm^R) to facilitate selection and plasmid maintenance.

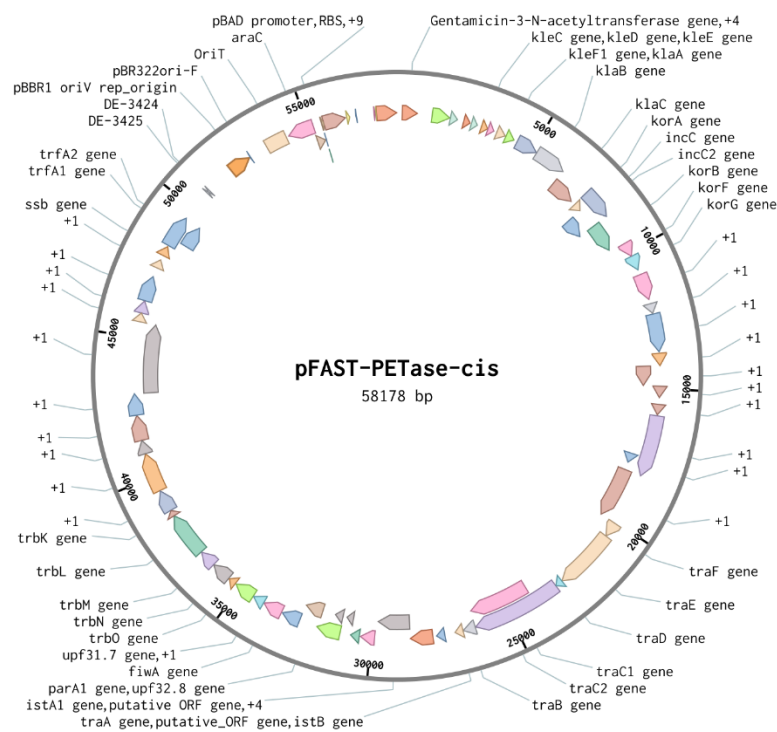


Figure 3.1. Plasmid map of pFAST-PETase-cis. The corresponding GenBank formatted file of the nucleotide sequence is provided in Supplementary Data 3.4.

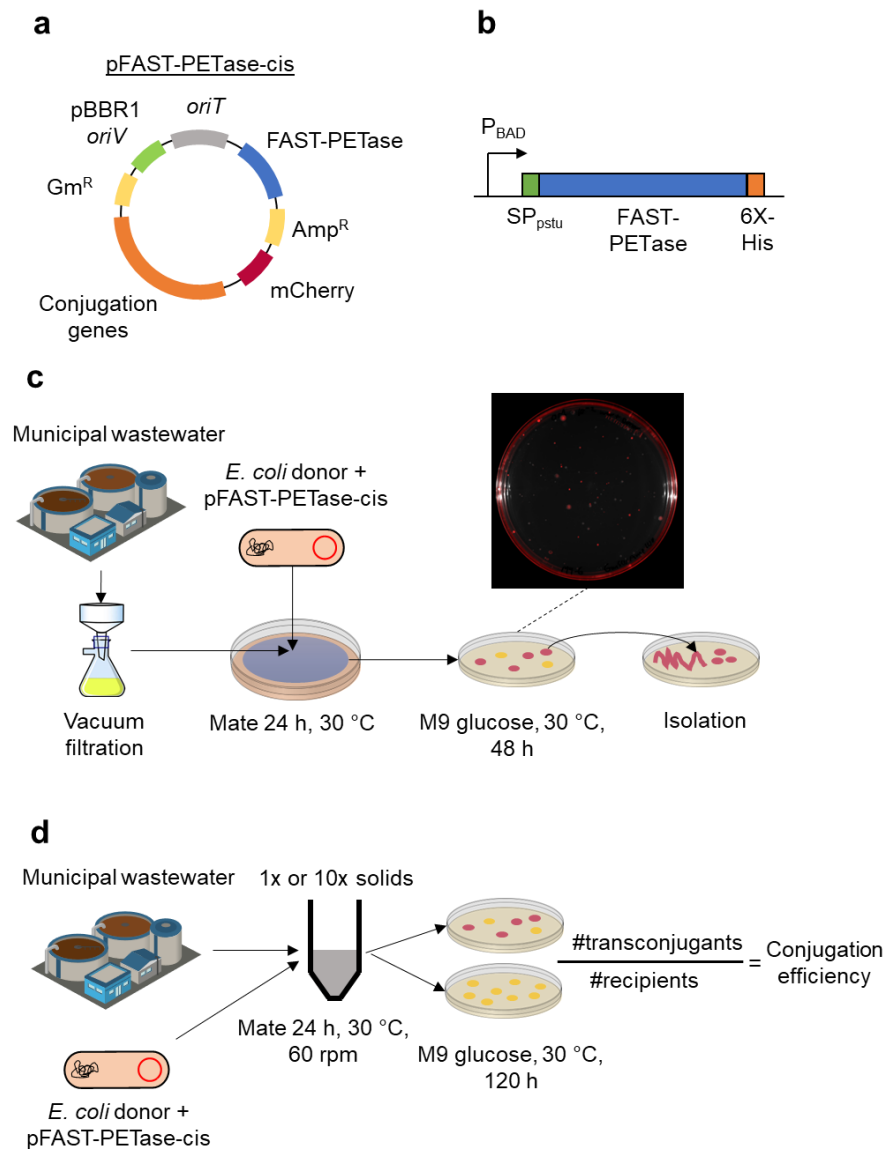


Figure 3.2. Conjugation of pFAST-PETase-cis into wastewater bacteria. **a** Schematic map of pFAST-PETase-cis (not to scale): *oriT*, RK2/RP4 conjugative origin of transfer; FAST-PETase, gene for FAST-PETase enzyme; Amp^R , ampicillin resistance gene (*bla-TEM₁*); *mCherry*, gene for fluorescent protein mCherry; Conjugation genes, encoding the RK2/RP4 conjugation system; Gm^R , gentamicin resistance gene; pBBR1 *oriV*, plasmid origin of replication. **b** FAST-PETase coding region (not to scale) highlighting the arabinose-inducible promoter (P_{BAD}), signal peptide (SP_{pstu}), and 6×His-tag. **c** Experimental procedure for conjugation of pFAST-PETase-cis into bacteria from a wastewater sample. **d** Experimental procedure for measuring conjugation efficiency in wastewater suspension.

We used *E. coli* K-12 ($\Delta ilvD::FRT$, $\Delta galK::cfp-bla$, pSAS31, pTGD, pFAST-PETase-cis) as conjugative donor. This strain is auxotrophic to isoleucine, leucine, and valine, allowing for counter-selection on minimal media (to facilitate isolation of transconjugants). Untreated wastewater influent was sourced from the City of Waterloo, Ontario, Canada. This water source was chosen because wastewater influent is often a major source of microplastics which enter the environment through wastewater treatment plants (Sun *et al.*, 2019; Liu *et al.*, 2021). In addition, wastewater influent is representative of bacterial diversity inside wastewater treatment plants (Ye and Zhang, 2013). We filter-mated the donor strain with the wastewater sample and picked 32 transconjugant colonies expressing *mCherry* on M9 glucose agar plates supplemented with ampicillin and gentamicin (Figure 3.2c, details in Methods). No *mCherry* expression was detected in a wastewater sample incubated without donors (Figure 3.3). Based on bacterial internal transcribed spacer (bITS) sequencing (Milani *et al.*, 2020) on all 32 colonies in a pooled sample (performed by Metagenom Bio Life Science, details in Methods), we observed that most isolates were *Gammaproteobacteria* from the *Enterobacteriaceae* family (Supplementary Data 3.1), with one exception: *Pusillimonas* *ssp.*, a gram-negative bacterium from the *Burkholderiaceae* family. From these picked colonies, we selected seven isolates with distinct colony morphologies to identify using bITS sequencing on each individual colony (Table 2.1, Supplementary Data 3.2) and to investigate for PET degradation activity.

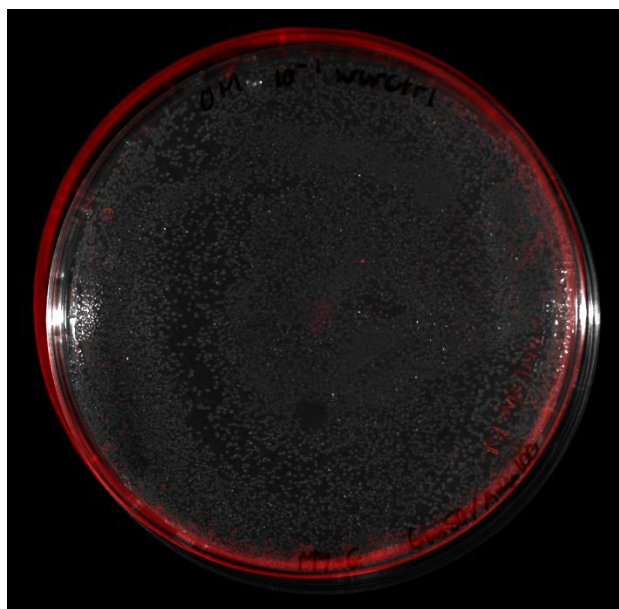


Figure 3.3. Selection plate for wastewater sample without donor strain added (diluted 10-fold). No fluorescent colonies are present, confirming that wild-type bacteria in the sample do not express mCherry or other red fluorescent proteins.

Table 3.1. Selected transconjugant isolates identified by bITS sequencing.

Isolate Number	Species
WW2	<i>Pusillimonas</i> ssp.
WW5	<i>Enterobacter</i> ssp.
WW7	<i>Escherichia</i> ssp.
WW17	<i>Enterobacter</i> ssp.
WW21	<i>Enterobacter</i> ssp.
WW23	<i>Citrobacter</i> ssp.
WW24	<i>Citrobacter</i> ssp.

To confirm *in situ* delivery in wastewater conditions, we performed a liquid-mating assay where donors were mixed directly into municipal wastewater influent samples (1× or 10× solids) and allowed to mate for 24 h (Figure 3.2d). After washing the mating mixture twice in 1× PBS, the mixture was serially diluted and plated onto selective (ampicillin and gentamicin) and non-selective M9 glucose plates, which were incubated for five days at 30 °C.

After counting colonies showing *mCherry* activity, we found that the conjugation efficiency (defined as the ratio of transconjugants to transconjugants plus recipients) was on the order of 1% (1.2% with 10x solids, 1.9% with 1x solids). To confirm that observed colonies were not the donor strain, an exponentially growing donor culture was washed twice in 1× PBS and incubated at 30 °C for up to seven days on selective M9 glucose plates—no growth was observed (Figure 3.4).

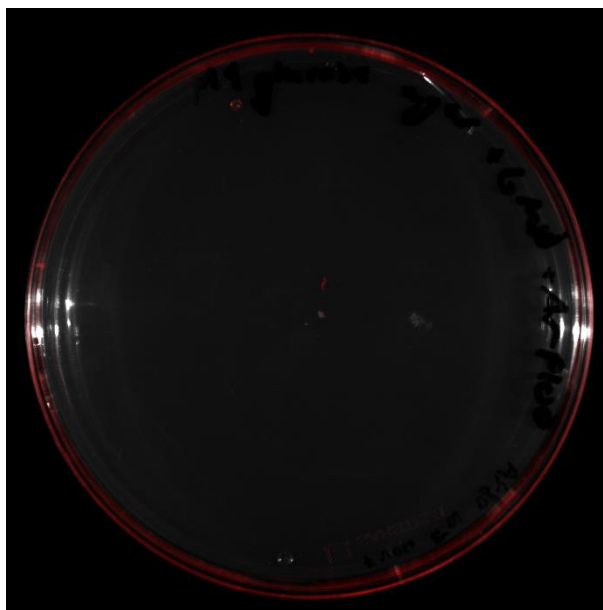


Figure 3.4. Donor strain incubated at 30 °C on M9 glucose plate with antibiotics for 6 days. No colonies are present, confirming auxotrophy of the donor strain.

3.3.2 pFAST-PETase-cis is lost in absence of selection

To assess the stability of pFAST-PETase-cis, we serially passaged four different transconjugant species once daily in LB media over the course of 14 days. The number of cells expressing ampicillin- and gentamicin-resistance (which are carried on pFAST-PETase-cis) decreased by 4-6 logs for WW2 (*Pusillimonas ssp.*) and WW7 (*Escherichia ssp.*), whereas it decreased by about one log for WW17 (*Enterobacter ssp.*) and WW23 (*Citrobacter ssp.*) (Figure 3.5). These results demonstrate that the plasmid is lost in the absence of selection, which highlights the need for additional measures to maintain the FAST-PETase gene in transconjugants.

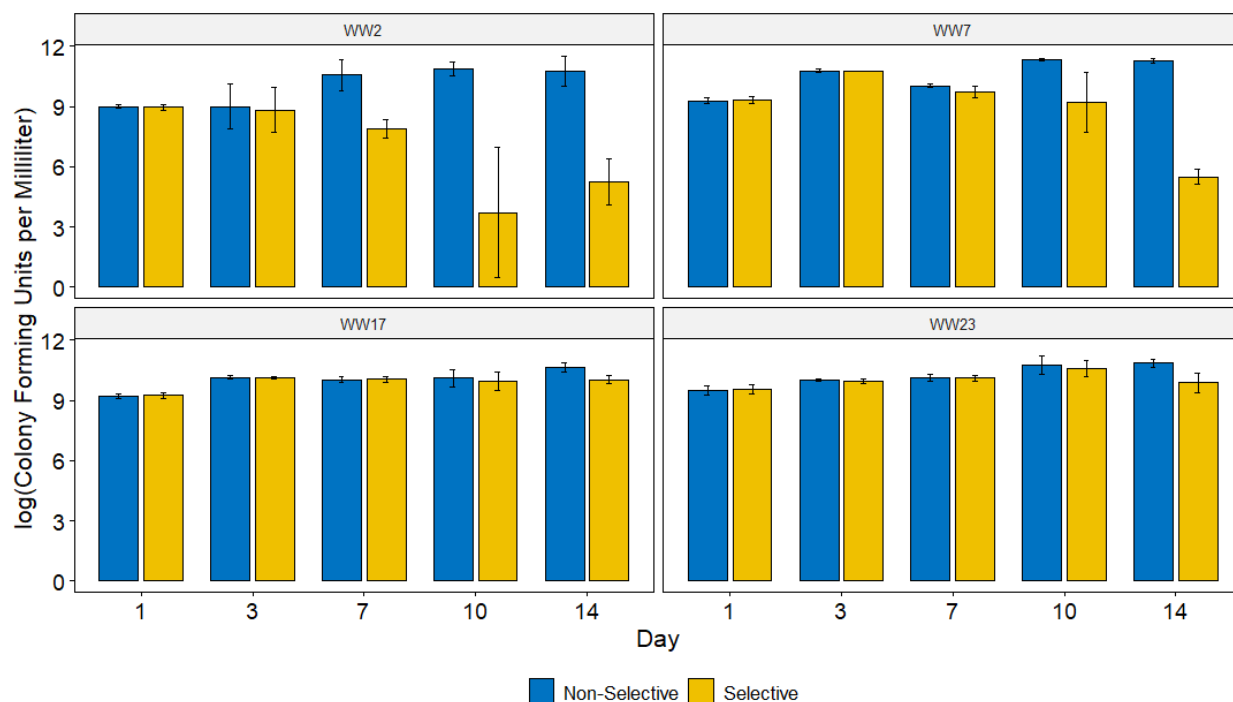


Figure 3.5. Cell density of wastewater transconjugants plated on selective (ampicillin and gentamicin) and non-selective plates over 14 serial passages in LB. Bars show the mean of three biological replicates. Error bars: $\pm$ s.d.

3.3.3 Purified FAST-PETase expressed from transconjugant isolates degrades PET

Before beginning our activity assays, we used SDS-PAGE to characterize intracellular expression of FAST-PETase (Figure 3.6a). We then incubated commercial, amorphous PET (aPET) film (crystallinity $8.8 \pm 1.0\%$ determined by differential scanning calorimetry, Methods) with $3.6 \mu\text{g}$ of purified protein in $0.1 \text{ M KH}_2\text{PO}_4\text{-NaOH}$ (corresponding to 200 nM FAST-PETase) from cell extracts. As a negative control, we incubated aPET film with the same buffer without FAST-PETase. Significant mass loss occurred for several isolates (Figure 3.6c) in comparison to the negative control. We then characterized presence of FAST-PETase in the supernatant by SDS-PAGE (Figure 3.6d) and incubated an aPET film with $3.6 \mu\text{g}$ of purified protein from culture supernatant. Again, significant mass loss occurred (Figure 3.6d) compared to the negative control. Inspection of aPET film samples with scanning electron

microscopy (SEM) showed that samples treated with FAST-PETase (Figure 3.6f) had a pitted surface, which is characteristic of enzymatic degradation (Yoshida *et al.*, 2016; Lu *et al.*, 2022). In contrast, an untreated aPET sample had a smooth surface (Figure 3.6e).

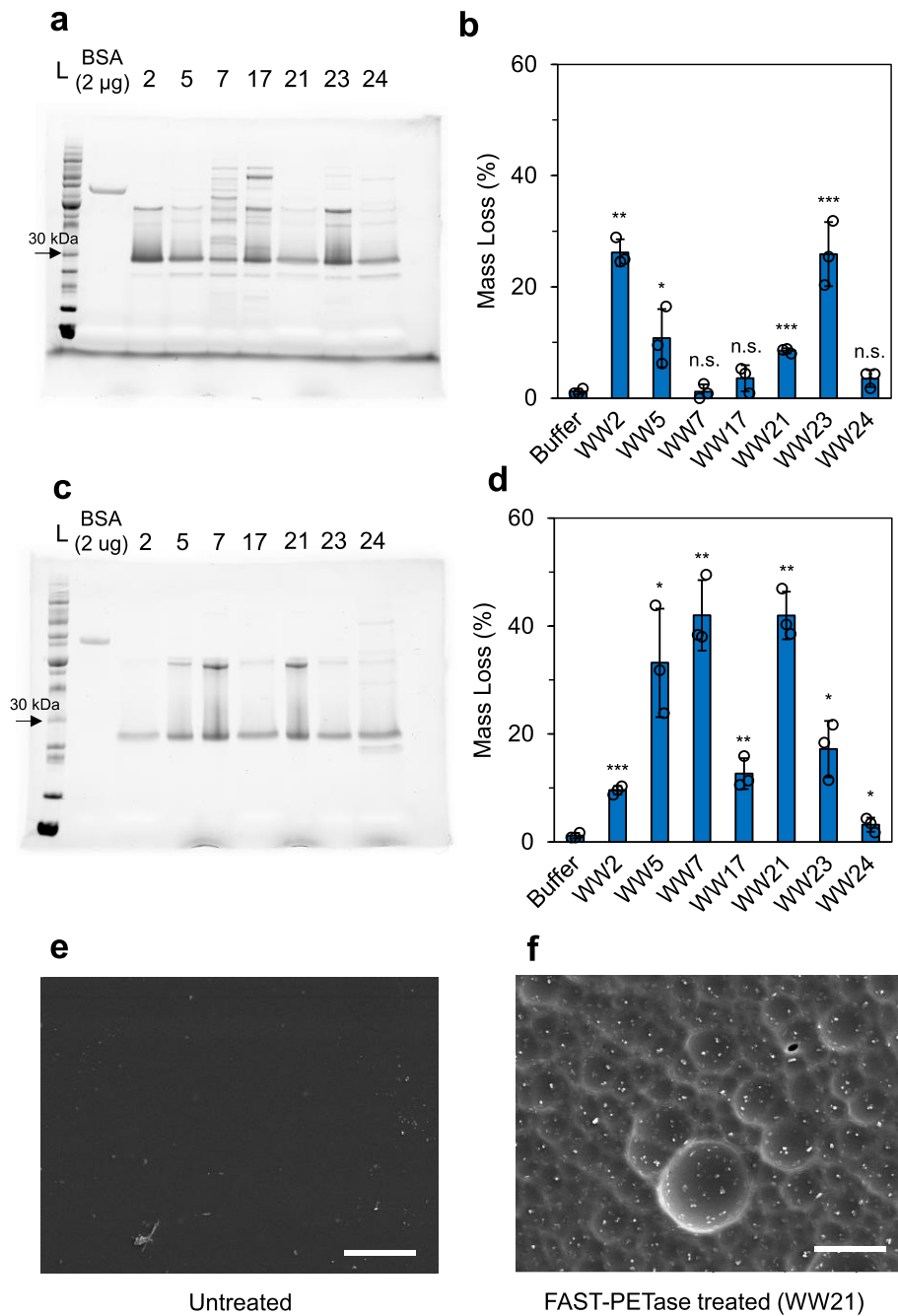


Figure 3.6. Enzymatic degradation of aPET film (7 mm diameter, 0.25 mm thick, ~11 mg) by purified FAST-PETase solution obtained from isolated transconjugants. SDS-PAGE gels of purified FAST-PETase are shown for **a**) cell lysates and **c**) culture supernatant. L: protein ladder (PageRuler Unstained Protein Ladder, ThermoFisher, no. 26614); BSA: bovine serum albumin; numbers indicate isolate ID. Mass loss from exposure to FAST-PETase purified from **b**) cell lysates and **d**) supernatant was measured after 96 h incubation at 50 °C. Bars show mean of three technical replicates for each isolate and control. Error bars: $\pm$ s.d. Statistical analyses

were performed using Welch's one-sided *t*-test to compare mass loss of each sample to the negative control that contained no FAST-PETase (Buffer) (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). **e** Scanning electron microscopy (SEM) image of an untreated PET film. **f** SEM image of a PET film incubated at 50 °C for 96 h with FAST-PETase purified from WW21 culture supernatant. Scale bars: 50 µm.

3.3.4 Culture supernatant containing FAST-PETase degrades post-consumer PET without enzymatic purification

After confirming the effectiveness of FAST-PETase isolated from transconjugants against commercial PET film, we next sought to confirm whether transconjugants could degrade PET without the need for enzymatic purification. FAST-PETase has been previously demonstrated to be effective against numerous post-consumer PET (pcPET) products (Lu *et al.*, 2022), so we switched to using post-consumer PET to simulate a source of PET that is more representative of PET waste encountered in the field. To test whether FAST-PETase produced from the isolates would degrade pcPET without enzymatic purification, we incubated 6 mm hole punches of pcPET from a coffee cup lid (crystallinity $11.4 \pm 0.2\%$ determined from differential scanning calorimetry, Methods) with culture supernatant (M9 glycerol medium) from WW2 (*Pusillimonas ssp.*), WW5 (*Enterobacter ssp.*), and WW7 (*Escherichia ssp.*) that was diluted in 0.1 M KH₂PO₄-NaOH buffer. We measured degradation using a spectrophotometric technique (Methods), which is more sensitive than gravimetric measurements. The PET degradation byproducts, terephthalic acid (TPA) and mono-(2-hydroxyethyl)terephthalate (MHET), have both been shown to have an absorbance peak around 240 nm (*A*₂₄₀) (Zhong-Johnson *et al.*, 2021).

In comparison to samples where FAST-PETase expression was either repressed or not induced, a significant increase in *A*₂₄₀ occurred in samples exposed to supernatant from cultures where FAST-PETase expression was induced (Figure 3.7a), indicating enzymatic degradation of the PET sample. Samples showing significant increase in *A*₂₄₀ also showed signs of degradation observable with the naked eye (Figure 3.7b). Further inspection by SEM showed roughening of the plastic surface of PET samples incubated with cultures induced to

express FAST-PETase (Figure 3.7b), which is characteristic of enzymatic degradation of PET (Yoshida *et al.*, 2016; Lu *et al.*, 2022).

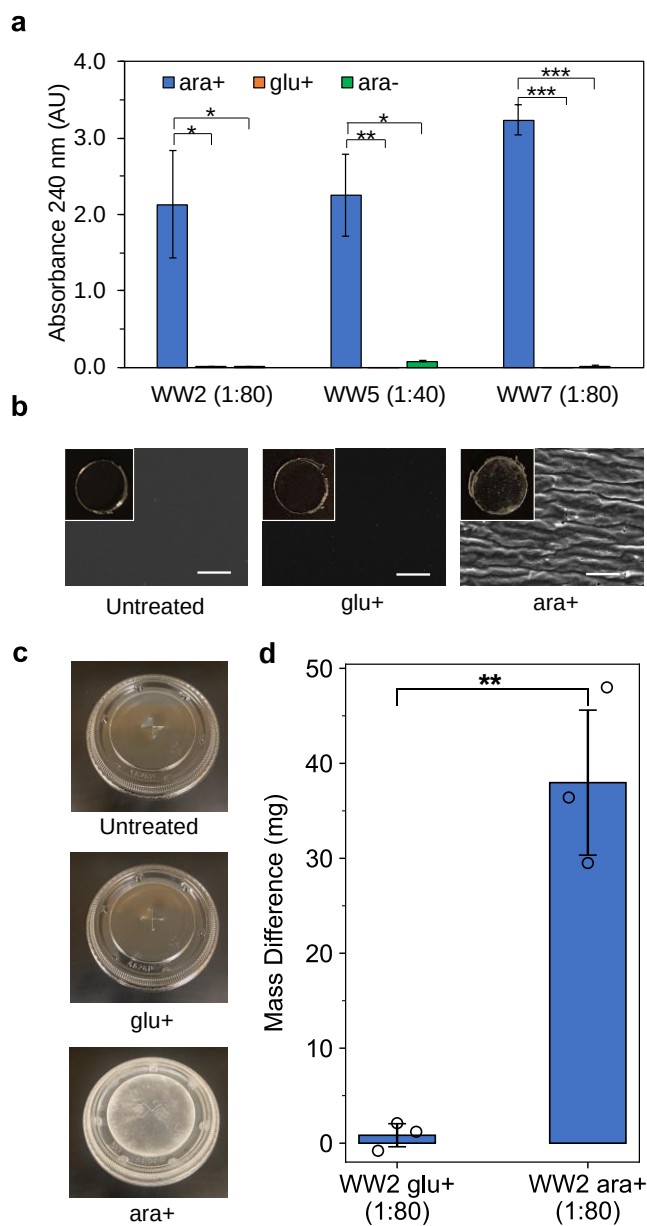


Figure 3.7. Degradation of a coffee cup lid using spent supernatant from engineered wastewater isolates diluted in 0.1 M KH_2PO_4 -NaOH buffer (pH 8.0). **a** A_{240} measurements for diluted supernatant incubated with a 6 mm hole punch of pcPET for 120 h at 50 °C. ara+: protein expression induced with 0.2% arabinose; glu+: protein expression suppressed with 0.2% glucose; ara-: no arabinose added. Bars show mean of three technical replicates. Error bars: $\pm$

s.d. **b** Photos (inset) and SEM images of PET samples prior to treatment (left) or incubated in culture supernatant from WW5 (1:40) for 120 h. Scale bars: 50 μm . **c** Coffee cup lid (bottom) shows macroscopic signs of degradation after 168 h incubation at 50 °C with spent supernatant (ara+) from WW2 (1:80). An untreated lid (top) and lid incubated with spent supernatant (glu+) from WW2 (1:80) (middle) are shown for comparison. **d** Difference between initial and final mass of coffee cup lids incubated in supernatant from WW2 (1:80) after 168 h incubation at 50 °C. Bars show mean of three technical replicates. Error bars: $\pm$ s.d. All statistical analyses were performed using Welch's one-sided *t*-test to compare ara+ samples to glu+ and ara- samples (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

We hypothesized that higher enzyme concentration in undiluted culture supernatant would improve the PET degradation rate. We thus followed the same protocol with culture supernatant that had not been diluted in KH_2PO_4 -NaOH buffer. Contrary to our hypothesis, we observed minimal activity: there was little to no increase in A_{240} relative to media where FAST-PETase expression was suppressed (Figure 3.8). We observed the same negative result when we repeated this experiment with an additional pH adjustment step (addition of 1 M NaOH) to modify the pH of the supernatant from 6.5 to 8.0, which is optimal for FAST-PETase activity (Lu *et al.*, 2022) (Figure 3.8).

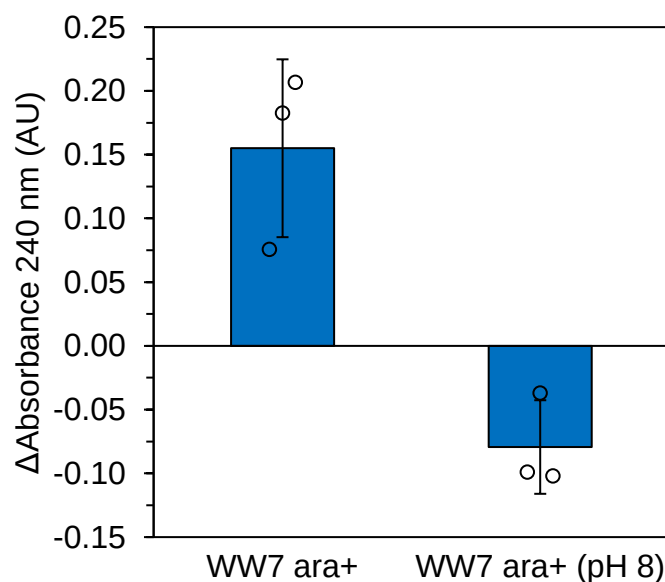


Figure 3.8. A_{240} measurements for M9 glycerol supernatant incubated with pcPET for 120 h at 50 °C. ara+: protein expression induced with 0.2% arabinose. Bars show mean of three technical replicates. Error bars: $\pm$ s.d.

We tested scaling of pcPET degradation by increasing the total reaction volume, allowing an entire coffee cup lid (same pcPET product used in Figure 3.7a) to incubate in diluted supernatant from WW2 (*Pusillimonas ssp.*) for 168 h at 50 °C. A lid that was incubated in supernatant where FAST-PETase expression was induced showed signs of degradation visible to the naked eye (Figure 3.7c) and significant mass loss (38 ± 9 mg, corresponding to $1.3 \pm 0.3\%$ of the initial mass of the lid; Figure 3.7d). In comparison, a lid incubated in supernatant where FAST-PETase expression was suppressed did not show any visible signs of degradation (Figure 3.7c) nor appreciable mass loss (Figure 3.7d).

3.4 Discussion

Engineering microbiomes to enable *in situ* degradation of plastics in diverse ecological niches remains a key milestone for deploying biological solutions for plastic waste at scale (Engineering Biology Research Consortium, 2022). Biotechnological plastic degradation and upcycling can offer a means to achieve a circular life cycle for plastic waste in a sustainable and cost-effective manner, facilitating a broader move towards a circular economy. Developments in biotechnological approaches to plastic waste management will ideally complement related strategies for reduction of single-use plastics and innovations in biodegradable plastics.

In this work, we demonstrated a proof-of-concept for genetically engineering bacteria from municipal wastewater *in situ* to degrade PET. We achieved this by using a broad-host-range conjugative plasmid carrying the FAST-PETase gene. Compared to cell bioaugmentation efforts, genetic bioaugmentation bypasses the adaptation barrier of introducing exogenous microbes into a new environment, and thus could be used to achieve long-term bioremediation solutions. In particular, an approach focused on degradation of specific plastic compounds could be useful for removing microplastics from the waste streams of plastic manufacturing and recycling facilities, which are a notable source of microplastic output into wastewater (Brown *et al.*, 2023).

The approach demonstrated here may generalize to a variety of ecological niches: several studies have utilized the RK2/RP4 conjugation machinery coupled with a broad-host-

range origin of replication (or a library of such origins) to engineer microbes from diverse environments: soil (Klümper *et al.*, 2015), mammalian guts (Ronda *et al.*, 2019), and marine ecosystems (Goodman *et al.*, 1993). A broad-host-range genetic engineering platform could facilitate construction of microbial consortia for degradation of individual types of plastic polymers that have known biodegradation pathways (e.g., PET). Such libraries of consortia could eventually be further expanded for degradation of other common plastic polymers, potentially resulting in modular tools for degradation of mixed-waste plastics.

Our results demonstrate that active FAST-PETase released by the engineered isolates can degrade PET without the need for enzymatic purification (Figure 3.7). The assays using purified enzyme (Figure 3.6) illustrate the plastic-degrading potential of the engineered isolates, which can guide future optimization efforts. However, we found that the solution conditions had a substantial effect on FAST-PETase activity. Our preliminary attempts to degrade commercial PET film in LB medium using purified enzyme solution showed no activity (purified from WW5; data not shown). Furthermore, our attempts at degrading PET using untreated and pH-adjusted supernatant also proved unsuccessful (Figure 3.8), highlighting that further developments are necessary for effective *in situ* degradation. It was necessary to dilute the supernatant to achieve significant PET degradation, consistent with previous observations that FAST-PETase activity is inhibited at high enzyme concentrations (Avilan *et al.*, 2023) and acidic conditions (Lu *et al.*, 2022). We found that the secreted enzyme fraction performed better than the intracellular enzyme fraction at similar dosages (e.g., WW5 and WW7 in Figure 3.6b and Figure 3.6d), with a lower abundance of contaminating proteins in the purified secreted fraction compared to the intracellular fraction (Figure 3.6a and Figure 3.6c), suggesting that the presence of other proteins in the medium may hinder FAST-PETase activity. Improved activity could be achieved by a design in which expressed FAST-PETase is anchored to the cell surface—immobilizing PETase enzymes has been shown to improve the stability and activity of the enzyme (Chen *et al.*, 2020), including under simulated wastewater conditions (Zhu *et al.*, 2022; Zurier and Goddard, 2022). Further optimization of enzyme activity could be achieved by addressing environmental factors such as pH, temperature, and biotic factors such as other secreted enzymes and metabolic waste products.

Alongside improving operating conditions, PET degradation efficiency could be enhanced by improving the genetic vector design. In pFAST-PETase-cis, FAST-PETase expression was regulated by the inducible P_{BAD} promoter for testing purposes. Regulating expression using a constitutive promoter could enable autonomous plastic degradation. Alternatively, expression induced by specific environmental conditions could be employed to minimize metabolic load and avoid unintended plastic degradation in e.g., products in use and infrastructure. Optimization of ribosome-binding sites, transcriptional terminators, and screening of signal peptides in members of the microbiome of interest could likewise prove beneficial (Low *et al.*, 2013; Xu *et al.*, 2019).

To achieve complete conversion of PET to its constituents, TPA and ethylene glycol (EG), the PETase activity investigated here could be complemented by a MHETase enzyme, which would enable conversion of the depolymerization product, MHET, into TPA and EG; alternatively, MHETase might be present in the environment (produced by, e.g. *I. sakaiensis* (Yoshida *et al.*, 2016)). The turnover rate of wild-type MHETase from *I. sakaiensis* ($k_{cat} \sim 25 \text{ s}^{-1}$ at 30 °C) (Knott *et al.*, 2020) is much higher than that of FAST-PETase ($k_{cat} \sim 0.016 \text{ s}^{-1}$ at 50 °C)^d, so depolymerization by PETase will likely be the rate-limiting step for the conversion of PET into TPA and EG. Moreover, MHETase has been observed to improve PETase activity by relieving product inhibition caused by MHET monomers (Knott *et al.*, 2020; Erickson *et al.*, 2022). Recently, a MHETase that is thermostable at the optimum operating temperature for FAST-PETase has been designed (Zhang *et al.*, 2023), offering an attractive dual-enzyme system for applications in PET biodegradation and upcycling. If production of large quantities of TPA and EG pose ecological or environmental concerns, bioaugmentation of additional consortia could be considered, using, for example, the consortium designed by Bao *et al.* (Bao *et al.*, 2023) for efficient conversion of TPA and EG into biomass.

Using *E. coli* as a donor strain, we found that pFAST-PETase-cis could conjugate on a filter-mating setup and in native wastewater conditions at efficiencies high enough to allow

^d Converted from the specific activity of $348 \mu\text{mol}_{\text{TPAeq}} \text{ h}^{-1} \text{ mg}_{\text{enzyme}}^{-1}$ (Arnal *et al.*, 2023).

isolation of transconjugants by plating. Improvements in conjugation efficiency could be achieved by exploring alternative donor strains, conjugation systems, and plasmid incompatibility classes; working toward applications, such efforts would need to account for environmental conditions (e.g., pH, temperature, seasonal variability). Moreover, the concentration of suspended solids (not measured) could affect conjugation efficiency by increasing the total surface area available for attachment, promoting spatial proximity between donor and recipient bacteria (Hamilton *et al.*, 2019). We observed low taxonomic diversity among the isolates, likely in part due to our reliance on *mCherry* expression, driven by the P_{bs} promoter whose activity has only been previously reported in *Escherichia coli*, *Bacillus subtilis*, and *Saccharomyces cerevisiae* (Yang *et al.*, 2018). Moreover, it is possible that conjugation efficiency from *E. coli* is higher towards closely-related species (Alderliesten *et al.*, 2020). Finally, the laboratory growth conditions likely had an impact on the results; reduced diversity of transconjugants has been reported in mating conditions that are less conducive to growth of diverse species (Ronda *et al.*, 2019). Additional improvements for selection of the plasmid could be made by using a promoter library for selection markers and origins of replication (e.g., Ronda *et al.* (2019)). Such efforts could target specific hosts that might be best suited to application (e.g. non-pathogenic strains known to be abundant in the conditions of interest).

We observed that pFAST-PETase-cis was lost in the absence of selection to varying extents based on the host species (Figure 3.5). This could be attributed to copy number variance arising from differences in host-encoded replication and regulatory proteins, plasmid oligomer formation in recombination-proficient hosts, and differences in metabolic load (Summers, 1991). Nonetheless, our observations highlight that long-term bioremediation applications will require efforts toward improving genetic stability. Improved maintenance of the function of interest could be achieved by incorporating plasmid maintenance systems such as toxin-antitoxin systems (Kroll *et al.*, 2010; Lin *et al.*, 2023), gene entanglement (Blazejewski *et al.*, 2019; Chlebek *et al.*, 2023), and plasmid partitioning systems (Danino *et al.*, 2015) into the plasmid design, or by integrating the gene of interest into the chromosome of the host (Ronda *et al.*, 2019). Chromosomal integration of the genetic cargo could be further constrained to

non-pathogenic species by using species- and site-specific gene-editing tools (Rubin *et al.*, 2022). Complementary selection efforts may be required to ensure long-term *in situ* maintenance of engineered strains (Raper *et al.*, 2018).

Deployment of genetically modified microorganisms for degrading plastics will require careful consideration of environmental and public safety, and adhere to regional regulatory policies. To date, there have been few pilot-scale or field studies utilizing genetic bioaugmentation for bioremediation purposes (Ripp *et al.*, 2000; Sayler and Ripp, 2000; Venkata Mohan *et al.*, 2009). One notable case was a field release in Estonia in 1989 of *Pseudomonas putida* cultures carrying a gene for metabolizing phenol into a watershed contaminated from a phenol release caused by a subterranean oil shale mine fire (Peters *et al.*, 1997). Six years after the release, the introduced operon was still found to be present in native microbiota (Peters *et al.*, 1997). This study suggests that an introduced functionality could persist long-term in the environment if under selection for metabolism of an abundant pollutant. It will be prudent to test the use of engineered microbes for plastic degradation in contained systems while simultaneously assessing ecological risk, including: characterizing the modified the microorganisms and nature of the genetic modification, tracking the fate of these microorganisms and the introduced genes of interest, assessing the environmental impact of the release, and monitoring effects of the release on non-target microorganisms (Gustafsson and Jansson, 1993). Genetic biocontainment of the genetic vector using conditional lethality, conditional genetic regulation, and conditional fitness control could help prevent unintended escape of the modified microorganisms (Lee *et al.*, 2018). In wastewater treatment plants, disinfection through tertiary treatment operations could also offer an additional means of biocontainment. Moreover, the controlled environments in wastewater treatment units offer opportunities to assess ecological risks and explore performance of *in situ* microplastic degradation at scale.

3.5 Methods

3.5.1 Bacterial strains

E. coli DH5 α (Matthew Scott Lab, University of Waterloo) and NEB10- β (New England Biolabs (NEB), Whitby, Ontario, Canada) were used for cloning steps. *E. coli* K-12 (Δ ilvD::\DeltagalK::cfp-bla, pSAS31, pTGD, pFAST-PETase-cis) was used as a conjugative donor strain for matings with wastewater samples. Engineered wastewater isolates (Table 1) were used for protein expression as indicated.

3.5.2 Plasmid construction

All polymerase chain reaction (PCR) amplification steps were performed using Phusion DNA polymerase (NEB, Whitby, Ontario, Canada). Primers used for PCR amplification were ordered from Integrated DNA Technologies (IDT, Coralville, Iowa, United States) and are listed in Appendix A. PCR products were assembled using NEB Hifi DNA Assembly Master Mix (NEB) according to the manufacturer instructions. pNuc-trans-mCherry was constructed by replacing *mRFP* under the *P_{bs}* promoter on pNuc-trans-mRFP (Hamilton *et al.*, 2019) (gift from Thomas Hamilton and David Edgell, Western University) with *mCherry*. The backbone of pNuc-trans-mRFP was amplified using primers AY-9 and AY-10, and *mCherry* from pBT1-proD-mCherry (Addgene plasmid #65823, gift from Xiaoxia Lin, University of Michigan) using primers AY-7 and AY-8, followed by heat shock transformation into *E. coli* DH5 α . pFAST-PETase-trans was assembled using PCR-amplified fragments from pNuc-trans-mCherry (primers OM-5 and OM-6), pFA6a-link-yoEGFP-SpHis5 (primers OM-8 and OM-9) (Addgene plasmid #44836), and pBTK522::FAST-PETase (Lu *et al.*, 2022) (gift from Hal Alper, University of Texas at Austin) (primers OM-1 and OM-2), and was subsequently transformed into *E. coli* DH5 α by heat shock. pFAST-PETase-cis was constructed by amplifying pFAST-PETase-trans (primers DE-3124 and DE-3125) (Hamilton *et al.*, 2019) with 60-bp homologous overlaps to the AvrII cut site in pTA-Mob (Strand *et al.*, 2014). The pTA-Mob plasmid (gift from Thomas Hamilton and David Edgell, Western University) was linearized using AvrII (NEB) and combined with the amplified fragment from pFAST-PETase-

cis using a modified NEB Hifi Assembly protocol with a two-hour assembly step at 50 °C, followed by electroporation into electrocompetent *E. coli* NEB10-β.

3.5.3 Filter-mating conjugation and selection of engineered isolates

An auxotrophic donor strain was constructed by filter-mating *E. coli* NEB10-β (pFAST-PETase-cis) with *E. coli* K-12 ($\Delta ilvD::FRT$, $\Delta galK::cfp-bla$, pSAS31, pTGD) (gift from Xiaoxia Lin, University of Michigan) using a procedure similar to that described in ref. (Malwade *et al.*, 2017). Transconjugants were selected on Luria-Bertani (LB) plates (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, 15 g/L agar) supplemented with ampicillin (100 µg/mL), gentamicin (50 µg/mL), and kanamycin (50 µg/mL), and were used as donors in filter-mating procedures with wastewater samples.

Untreated wastewater samples were collected from the City of Waterloo wastewater system access point (Wayne Parker lab, University of Waterloo, Ontario, Canada) and were processed the same day. 50 mL of wastewater was filtered through a sterile 0.40 µm-pore polycarbonate filter (Sigma-Aldrich, Oakville, Ontario, Canada), under vacuum. The filter and its contents were then vortexed in 1 mL of 1× phosphate-buffered saline (PBS; 2.56 g/L Na₂HPO₄·7H₂O, 8 g/L NaCl, 0.2 g/L KCl, 0.2 g/L KH₂PO₄) for 1 min. Three washes, consisting of centrifuging for 5 min at 4000×g followed by resuspension in 1× PBS, were then completed. The solution was resuspended to an OD₆₀₀ of 0.5. The donor strain, which was grown overnight from single colony in LB medium (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract) supplemented with appropriate antibiotics, was diluted 1:50 in LB with appropriate antibiotics, and was grown at 37 °C with shaking until exponential phase (OD₆₀₀ of ~0.5). The donor culture was then pelleted by centrifugation and resuspended to an OD₆₀₀ of 0.5 in 1× PBS. Filter-mating of the donor culture and collected solids from the wastewater sample was completed by adding 100 µL of each onto a sterile 0.4 µm-pore polycarbonate filter placed on R2A agar (0.5 g/L yeast extract, 0.5 g/L proteose peptone no. 3, 0.5 g/L casamino acids, 0.5 g/L glucose, 0.5 g/L soluble starch, 0.3 g/L sodium pyruvate, 2.2 mM KH₂PO₄, 0.2 mM MgSO₄, 15 g/L agar). Mating was allowed to occur for ~24 hours before the filters were vortexed in 1 mL of 1× PBS for 1 min, and serial dilutions were plated on M9 glucose agar (2

g/L glucose, 2 mM MgSO₄, 0.1 mM CaCl₂, 64 g/L Na₂HPO₄·7H₂O, 15 g/L KH₂PO₄, 2.5 g/L NaCl, 5.0 g/L NH₄Cl, 15 g/L agar) supplemented with 50 µg/mL gentamicin and 100 µg/mL ampicillin. Plates were incubated at 30 °C until transconjugants appeared, which were observed after 1-2 days as red/pink colonies. Colonies that were identified as expressing *mCherry* based on fluorescence at 670 nm (Cy5 channel) using a Chemidoc MP Imaging System (Bio-Rad, Mississauga, Ontario, Canada) were streaked onto M9 glucose plates supplemented with 50 µg/mL gentamicin and 100 µg/mL ampicillin until pure cultures were obtained.

3.5.4 Conjugation efficiency in wastewater conditions

Untreated wastewater samples were collected from the City of Waterloo wastewater system access point (Wayne Parker lab, University of Waterloo) and were processed the same day. The donor strain *E. coli* K-12 ($\Delta ilvD::FRT$, $\Delta galK::cfp-bla$, pSAS31, pTGD, pFAST-PETase-cis) was grown from a single colony overnight in LB supplemented with 0.2% glucose and appropriate antibiotics, then diluted 1:50 into fresh LB with glucose and antibiotics. Once the donor strain reached an OD₆₀₀ = 0.4-0.6, 50 mL of donor culture was washed twice in sterile 1× PBS and resuspended in 50 mL of sterile 1× PBS. Five mL of donor culture was added to either 20 mL of wastewater or 20 mL of wastewater with 10× concentrated solids (prepared by centrifuging 200 mL of wastewater at 2550×*g* for 15 min and decanting 180 mL of supernatant) in a 50 mL falcon tube, then incubated for 24 h at 30 °C with 60 rpm shaking. After 24 h, 1 mL of culture was washed twice in sterile 1× PBS. Serial dilutions of this mixture were then plated on selective M9 glucose plates containing gentamicin and ampicillin, and on non-selective M9 glucose plates. The plates were incubated at 30 °C for five days. Transconjugants were enumerated by counting colonies identified as expressing *mCherry* by visual inspection and by imaging under the Cy5 channel on a Chemidoc MP Imaging System. Total recipients were enumerated by counting colonies on non-selective plates. Conjugation efficiency was determined as the ratio of transconjugants to the total of recipients plus transconjugants.

3.5.5 Plasmid stability

A single colony was inoculated into 3 mL of LB media and grown at 30 °C with shaking. The culture was diluted 1:1000 into 3 mL of fresh growth medium once per day. At selected time points (Day 1, 3, 7, 10, 14), the culture was serially diluted in sterile 1× PBS and plated on selective (gentamicin at 50 µg/mL and ampicillin at 100 µg/mL) and non-selective LB plates. Plates were incubated at 30 °C for 16-24 h, and colony forming units were counted manually on each plate to quantify the cell density.

3.5.6 Protein expression and purification

A single colony was inoculated into LB supplemented with ampicillin (100 µg/mL) and gentamicin (50 µg/mL) and grown overnight at 30 °C with shaking. The overnight culture was then diluted 1:50 into 50 mL of LB supplemented with ampicillin and gentamicin. Protein expression was induced with 0.2% arabinose when the OD₆₀₀ of the culture reached 0.6-0.8, after which the cells were cultured for ~20 h at 30 °C with shaking. For purification of FAST-PETase from cell extracts, 50 mL of induced culture was centrifuged at 2550×*g* for 15 min at 4 °C and resuspended in 10 mL of cold lysis buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). The cells were incubated on ice with 0.1 mg/mL lysozyme for 30 min. The cells were further lysed by sonication in three 1 min intervals each followed by 1 min cool-down (Qsonica XL-2000, 3.2 mm, 8 W). Cell debris was pelleted at 2550×*g* for 30 min, and the 6×His-tagged FAST-PETase was purified from the supernatant by Ni-NTA affinity chromatography. 1 mL of 50% Ni-NTA agarose (Qiagen, Toronto, Ontario, Canada) was equilibrated using 10 bed volumes of wash buffer (same formulation as lysis buffer). The cell lysate was passed through the column, followed by washing twice with 10 bed volumes of wash buffer twice. The bound proteins were eluted using an imidazole step gradient of 50, 75, 100, and 250 mM in five elution steps using 2 bed volumes of elution buffer (20 mM sodium phosphate, 300 mM NaCl, pH 7.4. Imidazole concentrations in each elution step were 50, 75, 100, 250, 250 mM, respectively). For purification of FAST-PETase from culture supernatant, 50 mL of induced culture was centrifuged at 2550×*g* and 4 °C for 30 min, and FAST-PETase was purified from the supernatant using the same chromatography procedure described above.

Protein purity for each elution was checked by SDS-PAGE using stain-free imaging on a Chemidoc MP Imaging System. The purified protein from elutions containing the fewest contaminating proteins (typically elution steps 3-5) were concentrated in 1× PBS using a 5 kDa cut-off Vivaspin Centrifugal Concentrator (Sartorius, Oakville, Ontario, Canada). Protein concentrations were determined using the micro-BCA assay (Thermo Fisher Scientific, Mississauga, Ontario, Canada), where protein concentrations were correlated to bovine serum albumin standards at an absorbance of 562 nm in a plate reader.

3.5.7 PET sample preparation

Amorphous PET (aPET) film from GoodFellow (Pittsburgh, Pennsylvania, United States; product ES301445; specification: 0.25 mm thick, 1.300-1.400 g cm⁻³ density, 1.580-1.640 refractive index, 100×10⁻¹³ cm³ · cm cm⁻² s⁻¹ Pa⁻¹ permeability to water at 25 °C) and post-consumer PET (pcPET) (coffee cup lid No. A626P, Amhil North America, Mississauga, Ontario, Canada) were prepared in circular form (aPET: 7 mm diameter, pcPET: 6 mm diameter) using office hole punchers. The PET discs were washed for 30 min with 1% SDS, 20% ethanol, and distilled water, and dried overnight at 50 °C. Prior to degradation experiments, individual PET discs were weighed on an analytical balance with accuracy of ± 0.1 mg.

3.5.8 PET degradation assay with purified FAST-PETase

A single, pre-weighed aPET disc was placed in a 2 mL polypropylene microcentrifuge tube with 600 µL of 0.1 M KH₂PO₄-NaOH buffer (pH 8.0) with 3.6 µg of purified protein concentrate (corresponding to 200 nM FAST-PETase). The tube was capped to prevent volatilization. The sample was incubated at 50 °C for 96 h, replenishing with fresh buffer and enzyme solution every 24 h to maximize the degradation rate. The PET samples were subsequently washed with 1% SDS, 20% ethanol, and distilled water, dried for 24 h at 50 °C, then weighed on an analytical balance to determine weight loss.

3.5.9 PET degradation with culture supernatant

A single colony was inoculated into 10 mL of M9 glycerol (4 g/L glycerol, 2 mM MgSO₄, 0.1 mM CaCl₂, 64 g/L Na₂HPO₄·7H₂O, 15 g/L KH₂PO₄, 2.5 g/L NaCl, 5.0 g/L NH₄Cl) supplemented with ampicillin (100 µg/mL) and gentamicin (100 µg/mL) and grown until saturation at 30 °C with shaking. The saturated culture was then diluted 1:50 into 50 mL of M9 glycerol supplemented with ampicillin and gentamicin. Protein expression was induced with 0.2% arabinose, not induced, or repressed with 0.2% glucose when the OD₆₀₀ of the culture reached 0.3-0.5, after which the cells were cultured for 12-14 h at 30 °C with shaking (220 rpm). The supernatant for each culture was collected by centrifuging 50 mL of culture at 2550×*g* for 60 min at 4 °C and stored for the duration of the experiment at 4 °C. A single pcPET disc was placed in a 2 mL polypropylene microcentrifuge tube with 2 mL of supernatant. A replicate tube was prepared without PET for comparison to account for any change in absorbance from other compounds. All samples were incubated at 50 °C for 120 h. After incubation, samples were centrifuged at 14000×*g* for 3 min. The supernatant was transferred to a quartz cuvette, and absorbance at 240 nm was measured using UV-spectrophotometry (Pirillo *et al.*, 2021; Zhong-Johnson *et al.*, 2021), where the spectrophotometer was blanked to 0.1 M KH₂PO₄-NaOH buffer (pH 8). The mean absorbance of samples without PET was subtracted from the absorbance reading of samples that contained PET.

For degradation of whole coffee cup lids, the lids were washed, dried, and pre-weighed using the same procedure described in Section 3.5.7. The lids were submerged in individual sealed containers with 150 mL of 0.1 M KH₂PO₄-NaOH buffer (pH 8) containing 1:80 culture supernatant from WW2. The reaction was carried out for 168 h at 50 °C. The final mass of the lids were measured after washing and drying using the same procedure described in Section 3.5.7.

3.5.10 Scanning electron microscopy

PET samples were washed with 1% SDS, 20% ethanol, and distilled water, then dried overnight at 50 °C. The samples were attached to an aluminum stub using double-sided carbon tape, followed by sputter-coating (Polaron Instruments SEM Coating Unit E5100) with gold particles for 2 min at 20 mA in an argon atmosphere (layer thickness ~30 nm). The samples were then imaged by scanning electron microscopy (Tescan VEGA TS-5130) at 20 kV and 1000× magnification.

3.5.11 Differential scanning calorimetry

PET film crystallinity was determined using differential scanning calorimetry. PET film samples (9-11 mg) were loaded into aluminum Tzero pans (TA Instruments, Grimsby, Ontario, Canada; product 901683.901) with a Tzero lid (TA Instruments; product 901671.901) for solid samples. Samples were heated from 40 to 300 °C at 10 °C min⁻¹, held at 300 °C for 1 min, cooled from 300 to 30 °C at 10 °C min⁻¹, and held at 30 °C for 1 min in a Q2000 calorimeter (TA Instruments) with an RCS90 refrigerated cooling system. Percentage crystallinity was calculated as:

$$\%crystallinity = \left[\frac{\Delta H_f - \Delta H_c}{\Delta H_f^o} \right] \times 100\%$$

where ΔH_f is the enthalpy of fusion at the melting point, ΔH_c is the enthalpy of crystallization, and ΔH_f^o is the enthalpy of fusion for a completely crystalline sample at its equilibrium melting temperature, which is 140 J g⁻¹ (ref. (Mehta *et al.*, 1978)). TA Universal Analysis software was used to calculate ΔH_f and ΔH_c . ΔH_f and ΔH_c were measured by integrating from 110-120 °C to 150 °C and 210 °C to 260 °C, respectively, with a linear baseline. Crystallinity was measured in triplicate for each source of PET.

3.5.12 Statistical analysis

Experiments were performed in biological and technical replicates where mentioned in figure captions. All statistical analyses were performed in Python 3.9 using the SciPy (Virtanen *et*

al., 2020) library (v1.7.3). Welch's one-sided *t*-test was used to compare the means of two groups. $P < 0.05$ was considered as statistically significant.

3.5.13 Imaging plates

Plates were imaged on a Chemidoc MP Imaging System (Bio-Rad) on the Cy5 and Colorimetric channels with an exposure time of 5 ms and 0.035 ms, respectively. 16-bit images for each channel were overlayed using ImageJ and the minimum displayed pixel intensity for each channel was adjusted to enhance viewing (Cy5: 9994; Colorimetric: 60026).

3.5.14 Confirming auxotrophy of *E. coli* donor strain

An overnight culture grown in LB supplemented with ampicillin (100 µg/mL), gentamicin (50 ug/mL) and kanamycin (50 µg/mL) from a single colony and cultured at 37 °C with shaking. The overnight culture was diluted 1:50 into LB supplemented with the same antibiotics and was grown to early exponential phase ($OD_{600} \sim 0.5$). The culture was washed twice in sterile $1\times$ PBS and was resuspended in 1 mL of $1\times$ PBS. 100 µL of culture was plated on M9 glucose agar supplemented with ampicillin (100 µg/mL), gentamicin (50 ug/mL) and kanamycin (50 µg/mL) and incubated at 30 °C for up to six days.

3.5.15 Identification of isolates

Pure cultures were submitted to Metagenom Bio Life Science (Waterloo, Ontario, Canada) for identification by sequencing the bacterial internal transcribed spacer (ITS) region (Milani *et al.*, 2020). The company provided the following details about their protocol: "Genomic DNA was extracted using the Metagenom Bio Sox DNA Isolation Kit (No. 18011-50). Full length PCR primers as described below were designed and synthesized by IDT. Illumina adapter and 8-nt index sequences were added at the 5' ends of bacterial ITS forward [UNI_ITS_fw: 5' KRGGRYKAAGTCGTAACAAG] and reverse primers [bacITS189r: 5' TACTDAGATGTTTCASTTC] respectively (Hunt *et al.*, 2006; Milani *et al.*, 2020). In order to increase T_m , CGC and CGCG were added to the 5' ends of forward and reverse primers respectively. PCR was set up in duplicate for each DNA sample (25 µl each). Reaction mixture

consisted of 12.5 µl of Q5 High-Fidelity 2x Master Mix (NEB), 1.25 µl of 10 µM PCR forward primer, 1.25 µl of 10 µM PCR reverse primer, 5.0 µl eDNA, and 5.0 µl of PCR water. DNA was initially denatured at 98 °C for 3 min, followed by 35 cycles of 98 °C for 10 sec, 53 °C for 30 sec and 72 °C for 1 min, and then extended at 72 °C for 5 min. Duplicate PCR products were pooled and confirmed by 1.5% TAE agarose gel. Remaining product was purified by AMPure XP beads and quantified using the Qubit dsDNA assay kit. Average sizes of bITS amplicons were estimated based on the major band on the gel. Amplicons were sequenced with MiSeq Reagent v2 kit (2x250 cycles), and FASTQ data were generated. The 3' ends of 16S rRNA are in MiSeq read 1, and the 5' ends of 23S sequences are in Read 2 respectively. Primers were removed from raw reads using cutadapt v.3.5 (Martin, 2011). Reads were then quality filtered using DADA2 v.1.22 (Callahan *et al.*, 2016) with default parameters and removing all sequences with ambiguous nucleotides (N). Chimeric sequences were removed using the *dada2::removeBimeraDenovo()* method and due to the amplicon length the forward and reverse reads were processed separately and stored in separate amplified sequence variant (ASV) tables. ASVs constructed from forward reads were used in taxonomic assignment. Classification was performed using a naive Bayesian classifier implemented in DADA2, *dada2::assignTaxonomy()*, trained against the bacterial ITS region extracted from genomes in the Genome Taxonomy Database (GTDB) release 202 (Parks *et al.*, 2022). Species-level taxonomy was corrected by exact sequence classification using *dada2::assignSpecies()*. All analysis was performed using R 4.1.2.”

3.6 Acknowledgements

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC): RGPIN-03826-2018 (B.P.I.); 355513-2017 (M.G.A.); CGS-D (A.Y.). A.Y. was also supported by funding from the Ken O’Driscoll Graduate Scholarship in Polymer Engineering/Science, and Murray Moo Young Biochemical Engineering Scholarship. We are thankful to our lab assistants for their support in preparing materials and reagents (Melissa Meighoo, Julie Kim, Joyce Yang), cloning procedures (Melissa Meighoo), and CFU counts (Nitya Yadav). We thank Narasimman Lakshminarasimman and Wayne Parker

(University of Waterloo) for providing wastewater samples. We thank Michael Lynch (Metagenom Bio Life Science) for providing information about the bITS sequencing protocol. This work was carried out on the Haldimand Tract, the land granted to the Six Nations that includes six miles on each side of the Grand River.

3.7 References

- Albright, M.B.N., Louca, S., Winkler, D.E., Feeser, K.L., Haig, S.-J., Whiteson, K.L., et al. (2022) Solutions in microbiome engineering: prioritizing barriers to organism establishment. *ISME J* **16**: 331–338.
- Alderliesten, J.B., Duxbury, S.J.N., Zwart, M.P., de Visser, J.A.G.M., Stegeman, A., and Fischer, E.A.J. (2020) Effect of donor-recipient relatedness on the plasmid conjugation frequency: a meta-analysis. *BMC Microbiol* **20**: 135.
- Andrady, A.L. and Neal, M.A. (2009) Applications and societal benefits of plastics. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**: 1977–1984.
- Arnal, G., Anglade, J., Gavalda, S., Tournier, V., Chabot, N., Bornscheuer, U.T., et al. (2023) Assessment of four engineered PET degrading enzymes considering large-scale industrial applications. *ACS Catal* **13**: 13156–13166.
- Avilan, L., Lichtenstein, B.R., König, G., Zahn, M., Allen, M.D., Oliveira, L., et al. (2023) Concentration-dependent inhibition of mesophilic PETases on poly(ethylene terephthalate) can be eliminated by enzyme engineering. *ChemSusChem* **16**: e202202277.
- Bao, T., Qian, Y., Xin, Y., Collins, J.J., and Lu, T. (2023) Engineering microbial division of labor for plastic upcycling. *Nat Commun* **14**: 5712.
- Bathe, S., Mohan, T.V.K., Wuertz, S., and Hausner, M. (2004) Bioaugmentation of a sequencing batch biofilm reactor by horizontal gene transfer. *Water Science and Technology* **49**: 337–344.
- Bell, E.L., Smithson, R., Kilbride, S., Foster, J., Hardy, F.J., Ramachandran, S., et al. (2022) Directed evolution of an efficient and thermostable PET depolymerase. *Nat Catal* **5**: 673–681.
- Blazejewski, T., Ho, H.-I., and Wang, H.H. (2019) Synthetic sequence entanglement augments stability and containment of genetic information in cells. *Science* **365**: 595–598.
- Borrelle, S.B., Ringma, J., Law, K.L., Monnahan, C.C., Lebreton, L., McGivern, A., et al. (2020) Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. *Science* **369**: 1515–1518.
- Brandon, A.M., Gao, S.-H., Tian, R., Ning, D., Yang, S.-S., Zhou, J., et al. (2018) Biodegradation of polyethylene and plastic mixtures in mealworms (larvae of *Tenebrio molitor*) and effects on the gut microbiome. *Environ Sci Technol* **52**: 6526–6533.
- Brown, E., MacDonald, A., Allen, S., and Allen, D. (2023) The potential for a plastic recycling facility to release microplastic pollution and possible filtration remediation effectiveness. *Journal of Hazardous Materials Advances* **10**: 100309.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.
- Chamas, A., Moon, H., Zheng, J., Qiu, Y., Tabassum, T., Jang, J.H., et al. (2020) Degradation rates of plastics in the environment. *ACS Sustainable Chem Eng* **8**: 3494–3511.
- Chattopadhyay, I. (2022) Role of microbiome and biofilm in environmental plastic degradation. *Biocatalysis and Agricultural Biotechnology* **39**: 102263.

- Chen, Z., Wang, Y., Cheng, Y., Wang, X., Tong, S., Yang, H., and Wang, Z. (2020) Efficient biodegradation of highly crystallized polyethylene terephthalate through cell surface display of bacterial PETase. *Science of The Total Environment* **709**: 136138.
- Chettri, D., Verma, Ashwani Kumar, and Verma, Anil Kumar (2023) Bioaugmentation: an approach to biological treatment of pollutants. *Biodegradation* **35**: 117–135.
- Chlebek, J.L., Leonard, S.P., Kang-Yun, C., Yung, M.C., Ricci, D.P., Jiao, Y., and Park, D.M. (2023) Prolonging genetic circuit stability through adaptive evolution of overlapping genes. *Nucleic Acids Research* **51**: 7094–7108.
- Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y., et al. (2021) Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. *ACS Catal* **11**: 1340–1350.
- Danino, T., Prindle, A., Kwong, G.A., Skalak, M., Li, H., Allen, K., et al. (2015) Programmable probiotics for detection of cancer in urine. *Science Translational Medicine* **7**: 289ra84-289ra84.
- Engineering Biology Research Consortium (2022) Engineering Biology for Climate & Sustainability: A Research Roadmap for a Cleaner Future.
- Erickson, E., Shakespeare, T.J., Bratti, F., Buss, B.L., Graham, R., Hawkins, M.A., et al. (2022) Comparative performance of PETase as a function of reaction conditions, substrate properties, and product accumulation. *ChemSusChem* **15**: e202101932.
- French, K.E., Zhou, Z., and Terry, N. (2020) Horizontal ‘gene drives’ harness indigenous bacteria for bioremediation. *Sci Rep* **10**: 15091.
- Gao, R. and Sun, C. (2021) A marine bacterial community capable of degrading poly(ethylene terephthalate) and polyethylene. *Journal of Hazardous Materials* **416**: 125928.
- Garbisu, C., Garaiurrebaso, O., Epelde, L., Grohmann, E., and Alkorta, I. (2017) Plasmid-mediated bioaugmentation for the bioremediation of contaminated soils. *Frontiers in Microbiology* **8**.
- Geyer, R., Jambeck, J.R., and Law, K.L. (2017) Production, use, and fate of all plastics ever made. *Science* **3**: e1700782.
- Goodman, A.E., Hild, E., Marshall, K.C., and Hermansson, M. (1993) Conjugative plasmid transfer between bacteria under simulated marine oligotrophic conditions. *Applied and Environmental Microbiology* **59**: 1035–1040.
- Gustafsson, K. and Jansson, J.K. (1993) Ecological risk assessment of the deliberate release of genetically modified microorganisms. *Ambio* **22**: 236–242.
- Hamilton, T.A., Pellegrino, G.M., Therrien, J.A., Ham, D.T., Bartlett, P.C., Karas, B.J., et al. (2019) Efficient inter-species conjugative transfer of a CRISPR nuclease for targeted bacterial killing. *Nat Commun* **10**: 4544.
- Hunt, D.E., Klepac-Ceraj, V., Acinas, S.G., Gautier, C., Bertilsson, S., and Polz, M.F. (2006) Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. *Applied and Environmental Microbiology* **72**: 2221–2225.
- Jeon, H.J. and Kim, M.N. (2016) Isolation of mesophilic bacterium for biodegradation of polypropylene. *International Biodeterioration & Biodegradation* **115**: 244–249.

- Ke, Z., Wang, S., Zhu, W., Zhang, F., Qiao, W., Jiang, J., and Chen, K. (2022) Genetic bioaugmentation with triclocarban-catabolic plasmid effectively removes triclocarban from wastewater. *Environmental Research* **214**: 113921.
- Klümper, U., Riber, L., Dechesne, A., Sannazzarro, A., Hansen, L.H., Sørensen, S.J., and Smets, B.F. (2015) Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J* **9**: 934–945.
- Knott, B.C., Erickson, E., Allen, M.D., Gado, J.E., Graham, R., Kearns, F.L., et al. (2020) Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proceedings of the National Academy of Sciences* **117**: 25476–25485.
- Kroll, J., Klintner, S., Schneider, C., Voß, I., and Steinbüchel, A. (2010) Plasmid addiction systems: perspectives and applications in biotechnology. *Microbial Biotechnology* **3**: 634–657.
- Kumar, V., Sharma, N., Duhan, L., Pasrija, R., Thomas, J., Umesh, M., et al. (2023) Microbial engineering strategies for synthetic microplastics clean up: A review on recent approaches. *Environmental Toxicology and Pharmacology* **98**: 104045.
- Lau, W.W.Y., Shiran, Y., Bailey, R.M., Cook, E., Stuchtey, M.R., Koskella, J., et al. (2020) Evaluating scenarios toward zero plastic pollution. *Science* **369**: 1455–1461.
- Lee, J.W., Chan, C.T.Y., Slomovic, S., and Collins, J.J. (2018) Next-generation biocontainment systems for engineered organisms. *Nat Chem Biol* **14**: 530–537.
- Li, T., Menegatti, S., and Crook, N. (2023) Breakdown of polyethylene terephthalate microplastics under saltwater conditions using engineered *Vibrio natriegens*. *AIChE Journal* **69**: e18228.
- Lin, J., Guo, Y., Yao, J., Tang, K., and Wang, X. (2023) Applications of toxin-antitoxin systems in synthetic biology. *Engineering Microbiology* **3**: 100069.
- Liu, F., Wang, T., Yang, W., Zhang, Y., Gong, Y., Fan, X., et al. (2023) Current advances in the structural biology and molecular engineering of PETase. *Front Bioeng Biotechnol* **11**.
- Liu, W., Zhang, J., Liu, H., Guo, X., Zhang, X., Yao, X., et al. (2021) A review of the removal of microplastics in global wastewater treatment plants: Characteristics and mechanisms. *Environment International* **146**: 106277.
- Low, K.O., Muhammad Mahadi, N., and Md. Illias, R. (2013) Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Appl Microbiol Biotechnol* **97**: 3811–3826.
- Lu, H., Diaz, D.J., Czarnecki, N.J., Zhu, C., Kim, W., Shroff, R., et al. (2022) Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* **604**: 662–667.
- Ma, H., Zhao, Y., Yang, K., Wang, Y., Zhang, C., and Ji, M. (2022) Application oriented bioaugmentation processes: Mechanism, performance improvement and scale-up. *Bioresource Technology* **344**: 126192.
- Maheswaran, B., Al-Ansari, M., Al-Humaid, L., Sebastin Raj, J., Kim, W., Karmegam, N., and Mohamed Rafi, K. (2023) In vivo degradation of polyethylene terephthalate using microbial isolates from plastic polluted environment. *Chemosphere* **310**: 136757.

- Malwade, A., Nguyen, A., Sadat-Mousavi, P., and Ingalls, B.P. (2017) Predictive modeling of a batch filter mating process. *Frontiers in Microbiology* **8**.
- Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.
- Mehta, A., Gaur, U., and Wunderlich, B. (1978) Equilibrium melting parameters of poly(ethylene terephthalate). *Journal of Polymer Science: Polymer Physics Edition* **16**: 289–296.
- Milani, C., Alessandri, G., Mangifesta, M., Mancabelli, L., Lugli, G.A., Fontana, F., et al. (2020) Untangling species-level composition of complex bacterial communities through a novel metagenomic approach. *mSystems* **5**: e00404-20.
- OECD (2022) Global Plastics Outlook: Policy Scenarios to 2060, Paris: Organisation for Economic Co-operation and Development.
- Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., and Hugenholtz, P. (2022) GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* **50**: D785–D794.
- Peters, M., Heinaru, E., Talpsep, E., Wand, H., Stottmeister, U., Heinaru, A., and Nurk, A. (1997) Acquisition of a deliberately introduced phenol degradation operon, pheBA, by different indigenous *Pseudomonas* species. *Applied and Environmental Microbiology* **63**: 4899–4906.
- Pirillo, V., Pollegioni, L., and Molla, G. (2021) Analytical methods for the investigation of enzyme-catalyzed degradation of polyethylene terephthalate. *The FEBS Journal* **288**: 4730–4745.
- Prior, J.E., Lynch, M.D., and Gill, R.T. (2010) Broad-host-range vectors for protein expression across gram negative hosts. *Biotechnology and Bioengineering* **106**: 326–332.
- Qi, X., Ma, Y., Chang, H., Li, B., Ding, M., and Yuan, Y. (2021) Evaluation of PET degradation using artificial microbial consortia. *Frontiers in Microbiology* **12**.
- Raper, E., Stephenson, T., Anderson, D.R., Fisher, R., and Soares, A. (2018) Industrial wastewater treatment through bioaugmentation. *Process Safety and Environmental Protection* **118**: 178–187.
- Ren, C., Wang, Y., Tian, L., Chen, M., Sun, J., and Li, L. (2018) Genetic bioaugmentation of activated sludge with dioxin-catabolic plasmids harbored by *Rhodococcus* sp. strain p52. *Environ Sci Technol* **52**: 5339–5348.
- Ripp, S., Nivens, D.E., Ahn, Y., Werner, C., Jarrell, J., Easter, J.P., et al. (2000) Controlled field release of a bioluminescent genetically engineered microorganism for bioremediation process monitoring and control. *Environ Sci Technol* **34**: 846–853.
- Roberts, C., Edwards, S., Vague, M., León-Zayas, R., Scheffer, H., Chan, G., et al. (2020) Environmental consortium containing *pseudomonas* and *bacillus* species synergistically degrades polyethylene terephthalate plastic. *mSphere* **5**: e01151-20.
- Ronda, C., Chen, S.P., Cabral, V., Yaung, S.J., and Wang, H.H. (2019) Metagenomic engineering of the mammalian gut microbiome in situ. *Nat Methods* **16**: 167–170.

- Rubin, B.E., Diamond, S., Cress, B.F., Crits-Christoph, A., Lou, Y.C., Borges, A.L., et al. (2022) Species- and site-specific genome editing in complex bacterial communities. *Nat Microbiol* **7**: 34–47.
- Sayler, G.S. and Ripp, S. (2000) Field applications of genetically engineered microorganisms for bioremediation processes. *Current Opinion in Biotechnology* **11**: 286–289.
- Sekhar, V.C., Nampoothiri, K.M., Mohan, A.J., Nair, N.R., Bhaskar, T., and Pandey, A. (2016) Microbial degradation of high impact polystyrene (HIPS), an e-plastic with decabromodiphenyl oxide and antimony trioxide. *Journal of Hazardous Materials* **318**: 347–354.
- Shah, A.A., Hasan, F., Akhter, J.I., Hameed, A., and Ahmed, S. (2008) Degradation of polyurethane by novel bacterial consortium isolated from soil. *Ann Microbiol* **58**: 381–386.
- Son, H.F., Cho, I.J., Joo, S., Seo, H., Sagong, H.-Y., Choi, S.Y., et al. (2019) Rational protein engineering of thermo-stable PETase from *Ideonella sakaiensis* for highly efficient PET degradation. *ACS Catal* **9**: 3519–3526.
- Strand, T.A., Lale, R., Degnes, K.F., Lando, M., and Valla, S. (2014) A New and improved host-independent plasmid system for RK2-based conjugal transfer. *PLOS ONE* **9**: e90372.
- Summers, D.K. (1991) The kinetics of plasmid loss. *Trends in Biotechnology* **9**: 273–278.
- Sun, J., Dai, X., Wang, Q., van Loosdrecht, M.C.M., and Ni, B.-J. (2019) Microplastics in wastewater treatment plants: Detection, occurrence and removal. *Water Research* **152**: 21–37.
- Top, E.M., Springael, D., and Boon, N. (2002) Catabolic mobile genetic elements and their potential use in bioaugmentation of polluted soils and waters. *FEMS Microbiology Ecology* **42**: 199–208.
- Venkata Mohan, S., Falkentoft, C., Venkata Nancharaiah, Y., Sturm, B.S.M., Wattiau, P., Wilderer, P.A., et al. (2009) Bioaugmentation of microbial communities in laboratory and pilot scale sequencing batch biofilm reactors using the TOL plasmid. *Bioresource Technology* **100**: 1746–1753.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**: 261–272.
- Wang, N., Guan, F., Lv, X., Han, D., Zhang, Y., Wu, N., et al. (2020) Enhancing secretion of polyethylene terephthalate hydrolase PETase in *Bacillus subtilis* WB600 mediated by the SPamy signal peptide. *Letters in Applied Microbiology* **71**: 235–241.
- Wang, W., Gao, H., Jin, S., Li, R., and Na, G. (2019) The ecotoxicological effects of microplastics on aquatic food web, from primary producer to human: A review. *Ecotoxicology and Environmental Safety* **173**: 110–117.
- Xu, N., Wei, L., and Liu, J. (2019) Recent advances in the applications of promoter engineering for the optimization of metabolite biosynthesis. *World J Microbiol Biotechnol* **35**: 33.
- Xu, S., Ma, J., Ji, R., Pan, K., and Miao, A.-J. (2020) Microplastics in aquatic environments: Occurrence, accumulation, and biological effects. *Science of The Total Environment* **703**: 134699.

- Yan, F., Wei, R., Cui, Q., Bornscheuer, U.T., and Liu, Y.-J. (2021) Thermophilic whole-cell degradation of polyethylene terephthalate using engineered *Clostridium thermocellum*. *Microbial Biotechnology* **14**: 374–385.
- Yang, S., Liu, Q., Zhang, Y., Du, G., Chen, J., and Kang, Z. (2018) Construction and characterization of broad-spectrum promoters for synthetic biology. *ACS Synthetic Biology* **7**: 287–291.
- Ye, L. and Zhang, T. (2013) Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing. *Appl Microbiol Biotechnol* **97**: 2681–2690.
- Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., et al. (2016) A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* **351**: 1196–1199.
- Yuan, Z., Nag, R., and Cummins, E. (2022) Human health concerns regarding microplastics in the aquatic environment - From marine to food systems. *Science of The Total Environment* **823**: 153730.
- Zhang, J., Wang, H., Luo, Z., Yang, Z., Zhang, Z., Wang, P., et al. (2023) Computational design of highly efficient thermostable MHET hydrolases and dual enzyme system for PET recycling. *Commun Biol* **6**: 1–18.
- Zhong-Johnson, E.Z.L., Voigt, C.A., and Sinskey, A.J. (2021) An absorbance method for analysis of enzymatic degradation kinetics of poly(ethylene terephthalate) films. *Sci Rep* **11**: 928.
- Zhu, B., Ye, Q., Seo, Y., and Wei, N. (2022) Enzymatic Degradation of Polyethylene Terephthalate Plastics by Bacterial Curli Display PETase. *Environ Sci Technol Lett* **9**: 650–657.
- Zurier, H.S. and Goddard, J.M. (2021) Biodegradation of microplastics in food and agriculture. *Current Opinion in Food Science* **37**: 37–44.
- Zurier, H.S. and Goddard, J.M. (2022) Directed Immobilization of PETase on Mesoporous Silica Enables Sustained Depolymerase Activity in Synthetic Wastewater Conditions. *ACS Appl Bio Mater* **5**: 4981–4992.
- Zurier, H.S. and Goddard, J.M. (2023) PETase Engineering for Enhanced Degradation of Microplastic Fibers in Simulated Wastewater Sludge Processing Conditions. *ACS EST Water* **3**: 2210–2218.

Chapter 4

Calibrating spatiotemporal models of microbial communities to microscopy data: A review

The work included in this chapter is published: A. Yip, J. Smith-Roberge, S. Haghayegh-Khorasani, M.G. Aucoin, B.P. Ingalls. Calibrating spatiotemporal models of microbial communities to microscopy data: A review. *PLoS Computational Biology*, 18 (2022), e1010533.

There is considerable scientific and commercial interest in manipulating microbial communities to perform tasks in a predictable manner, and computational models are useful tools to predict the outcomes of such manipulations. Spatiotemporal models are frequently used to enhance understanding and prediction of community behaviour, and the model output is often compared against live-cell microscopy experiments for validation. This chapter surveys how spatiotemporal models are calibrated to live-cell microscopy observations of microbial communities. It also provides motivations for developing more rigorous model calibration practices in developing models of microbial communities, which are presented later in the thesis.

4.1 Abstract

Spatiotemporal models that account for heterogeneity within microbial communities rely on single-cell data for calibration and validation. Such data, commonly collected via microscopy and flow cytometry, have been made more accessible by recent advances in microfluidics platforms and data processing pipelines. However, validating models against such data poses significant challenges. Validation practices vary widely between modelling studies; systematic and rigorous methods have not been widely adopted. Similar challenges are faced by the (macro) ecology community, in which systematic calibration approaches are often employed to improve quantitative predictions from computational models. Here, we review single-cell observation techniques that are being applied to study microbial communities and

the calibration strategies that are being employed for accompanying spatiotemporal models. To facilitate future calibration efforts, we have compiled a list of summary statistics relevant for quantifying spatiotemporal patterns in microbial communities. Finally, we highlight some recently developed techniques that hold promise for improved model calibration, including algorithmic guidance of summary statistic selection and machine learning approaches for efficient model simulation.

4.2 Introduction

Microbial communities are ubiquitous [1]. They are responsible for life-sustaining planetary processes [2,3] and they maintain health in almost all metazoans, including humans [4]. Humanity has a long history of harnessing the power of natural microbial communities in, e.g., food fermentation [5], waste water treatment [6], and health [7]. Advances in sequencing and omics technologies have elucidated the roles of individual microbes within their communities, and how they contribute to the overall community function. This, in turn, has opened opportunities for manipulating and designing microbial communities to perform useful tasks across the bioeconomy [8].

Within microbial communities, species interact through, e.g., physical contact, competition for nutrients, metabolite exchange, toxin production, antibiotic inactivation, and quorum sensing. These interactions are shaped by a multitude of factors such as evolution [9] and abiotic features of the environment [10]. The network of interactions determines species abundances in a microbial community, thereby influencing the community's operation [11–13]. Growth of most communities involves attachment, and so cell-cell interactions influence the community's spatial arrangement [14]; the spatial structure may in turn influence the evolution of cooperative or competitive interactions [15,16]. To complicate things further, the community composition can also be impacted by the environment's colonization history [17,18]. The complex dependencies among cellular interactions, spatial dynamics, evolution, and community function make precision manipulation of microbiomes difficult. Mathematical models can be used to address this challenge by untangling the factors governing community behaviour.

To engineer microbial communities to suit our needs, we must first acquire a thorough understanding of how these communities operate [19]. Mathematical models can be used to guide rational manipulation and design of microbial communities, to predict how communities will behave, and to determine how well they will perform desired functions in, e.g. biotechnology, health and medicine, food and agriculture, and energy production [8]. Figure 4.1 depicts three classes of predictive models commonly used to describe microbial communities: ordinary differential equations (ODEs), partial differential equations (PDEs), and agent/individual-based models (henceforth referred to as ABMs) [20,21]. The primary distinction between these model types is their spatial resolution. ODE models are built on the assumption that the dynamics are not dependent on spatial distribution. Consequently, simulation and analysis of these models incurs a relatively low computational cost. PDE models explicitly account for spatial distribution, but describe local averages, rather than individuals. In contrast, ABMs can capture the spatiotemporal behaviour of individuals within populations, and so can account explicitly for heterogeneity amongst cells. This high degree of resolution comes at substantial computational cost, which scales (potentially nonlinearly) with the number of individuals and interactions in the population. ABMs are often combined with PDE and ODE sub-models to describe phenomena such as intracellular biomolecular network dynamics and extracellular diffusion. Recent applications of these modelling frameworks to microbial communities are reviewed in [22].

The choice of modelling framework is influenced by the system under consideration, the modelling objective, the data available, and the computational resources at hand. For each modelling framework, there are numerous open-source simulation packages available. The choice of software depends primarily on whether the built-in features are suitable for the application at hand. Reusability can be a challenge due to the diversity of programming languages and documentation formats employed. Some groups are developing packages with graphical user interfaces to facilitate reuse of their simulation software [23–25].

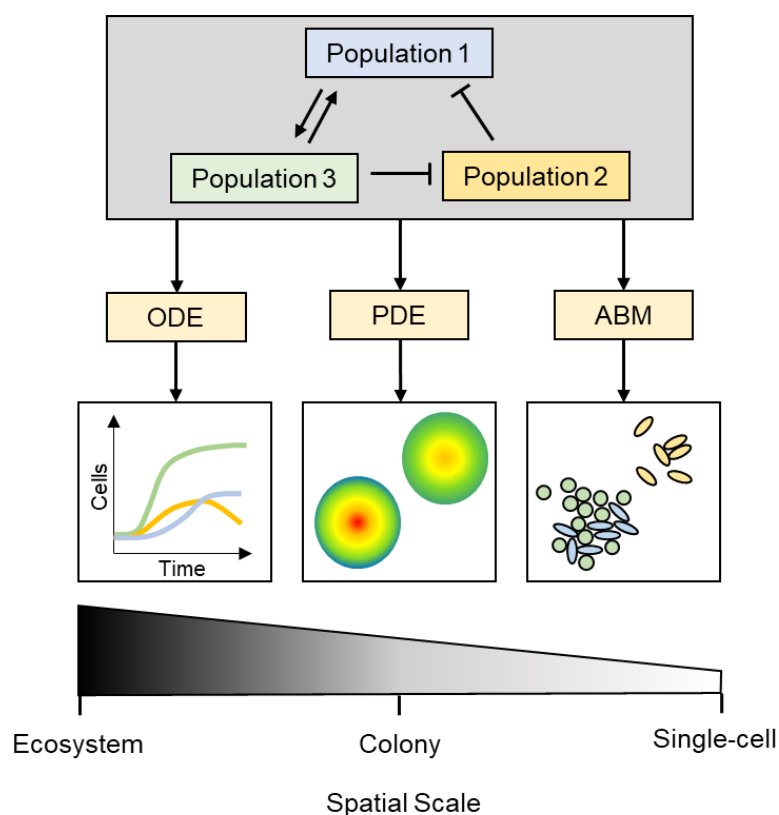


Figure 4.1. Modelling frameworks commonly used for capturing the behaviour of microbial communities, with associated spatial scales. ODE: ordinary-differential equation; PDE: partial differential equation; ABM: agent-based model.

Predictive models are necessary for applications that require precise design and manipulation of complex microbial communities [8]. These applications include human and animal health, food production, and environmental remediation. ABM and PDE models are suitable for modelling microbial growth in heterogeneous environments such as the mammalian gut and soil. To make accurate predictions, these models must be validated against experimental data, such as direct observations of populations of cells growing in spatially distributed environments. Observations at single-cell resolution can simultaneously provide data at the single-cell, population, and community scale. Such single-cell data is especially valuable for validating ABMs that aim to capture emergent population-level features by modelling single-cell behaviour [26,27]. Validation against independent patterns occurring at multiple scales generally improves model accuracy and predictive power [28].

To define the scope of the following discussion, we first establish a working definition of ‘microbial community’. Although a broad definition could be ‘microbes living together’, there is no consensus on how much variability is required to distinguish a microbial community from a microbial ‘monoculture’: even these exhibit some degree of genetic and phenotypic variability. For this review, we define the fundamental property of a community as the presence of at least two distinct characterized organism types, and thus we exclude monocultures that have developed some uncharacterized genetic heterogeneity. For details on the application of single-cell technologies to investigations of heterogeneity in such monocultures, the reader is referred to [29,30]. A number of distinct categories of communities have been investigated at the single-cell level [31]:

- **Communities of ‘isogenic mutants’** are co-cultures composed of at least two strains derived from the same parent that exhibit some genetic differences (due to either engineered or natural genetic alterations). Communities of isogenic mutants commonly serve as testing grounds for design and characterization of ecological interactions.
- **Designer laboratory communities** are composed of distinct species purposefully combined in a laboratory environment. The number of species is typically small compared to natural communities.
- **Natural communities** are sampled from natural or engineered environments (e.g., soil, animal guts, wastewater treatment plants, fermentation cultures).

Below, we provide a brief overview of current techniques for collecting single-cell level observations of microbial communities. We then survey how such measurements have been used to calibrate computational models of community dynamics, and we highlight systemic approaches that could be used to improve the rigour of model calibration procedures. Finally, we discuss techniques from (macro) ecology, topology, and data science that hold promise for facilitating efficient calibration.

4.3 Single-Cell Level Observations of Microbial Communities

Microbial communities are most commonly observed by flow cytometry and microscopy (Figure 4.2). Flow cytometry can be used to categorize cells by their morphology and physiological characteristics (e.g., fluorescence). Modern flow cytometers can process $\sim 10^4$ cells per second. The resulting large samples can provide robust statistics characterizing heterogeneous populations. The availability of several out-of-the-box software packages for flow cytometry [32,33] makes processing cytometry data straightforward in comparison to microscopy images, which usually call for customized image processing routines [34].

Flow cytometry is often used to determine the relative fraction of subpopulations within a community. Cells are distinguished by, e.g., fluorescent labeling, viability staining, or morphological differences. This approach has been used to measure short-term population dynamics in synthetic consortia [35], rates of plasmid propagation [36,37] and monitoring of eco-evolutionary feedback between cooperators and cheaters [38], among others.

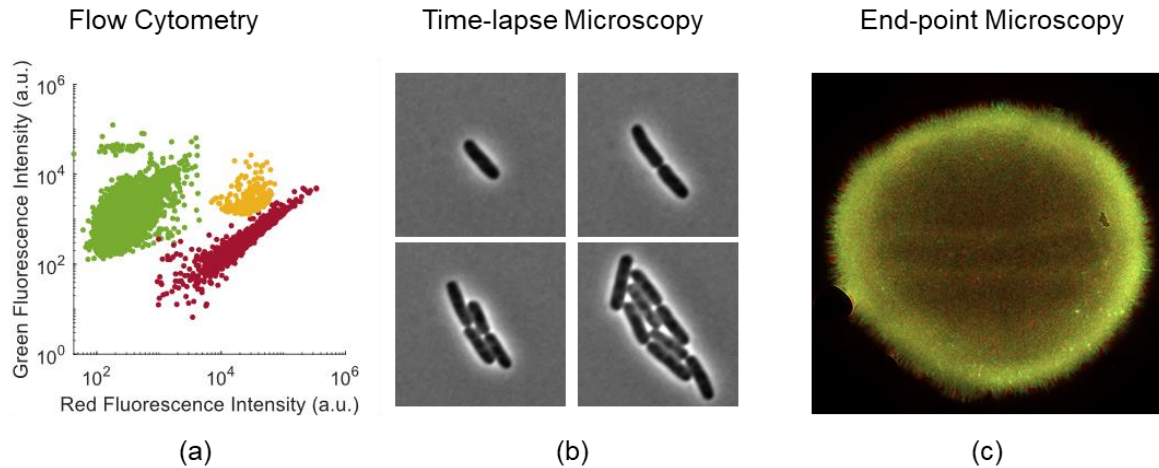


Figure 4.2. Single-cell measurement techniques for microbial communities. (a) Flow cytometers can measure cell fluorescence and morphology in large sample sizes. (b) Time-lapse microscopy allows for direct visualization of physical cell-cell interactions and quantitative measurement of single-cell characteristics over time. (c) End-point microscopy can illustrate large-scale spatial patterns and features of cell arrangements in 2 or 3 dimensions.

Time-lapse microscopy has been used to collect spatiotemporally resolved measurements of microbes in confined environments. Cells are typically observed growing in

monolayer using a widefield microscope, but multilayer growth can be resolved by confocal microscopy [39–41]. Fluorescence microscopy experiments can generate data on the spatiotemporal positions of cells and their fluorescence-associated phenotypes. Time-lapse images can be processed to reveal growth rates, lineages, gene expression levels, and to infer intercellular interactions, which together give rise to spatiotemporal features at the population level.

Time-lapse experiments generally involve observation of cells growing under agar pads or within microfluidic devices. Agar pads are simple to use, but suffer the limitations of batch cultures such as transient effects of nutrient depletion, desiccation, waste-product accumulation, and crowding [42,43]. In contrast, microfluidic devices offer more controlled and sustained environments, in which multiple cell generations can be observed through transient and steady-state growth conditions [44].

Quantitative analysis of time-lapse microscopy images demands the use of cell segmentation and cell tracking algorithms [45,46]. Analysis of time-lapse images reveals individual cell properties such as elongation rate, motion, and lineage, as well as population-level features such as population density and species abundance. Populations can sometimes be discriminated by morphology, but it is more common to use fluorescent markers. Moreover, fluorescence intensity can be used as a read-out of an internal genetic state [47]. Time-lapse microscopy has been used to obtain both single-cell and population measurements in both isogenic mutant [48–54] and laboratory designer communities [55,56].

By correlating individual cell elongation rates with counts of neighbouring cells, researchers have gained insight into cell-cell interactions such as metabolite exchange [57,58] and antibiotic efflux [59]. Such studies can take advantage of microfluidic device designs that constrain the proximity of neighbouring cells. For example, Moffitt et al. [60] and Gupta et al. [61] designed microfluidic devices permitting nutrient exchange between two physically-separated populations.

Contact-dependent interactions can be inferred by comparing changes in cell state to the presence of directly neighbouring cells. This approach has been employed by several groups studying type VI secretion (toxin-delivery) systems (T6SS). LeRoux et al. [62] and Smith et al. [63] measured the efficiency of target cell lysis as a function of contacts made. Steinbach et al. [64] investigated how the accumulation of dead cell debris reduces T6SS killing efficiency. Time-lapse microscopy studies of conjugation (contact-dependent horizontal gene transfer) have demonstrated the influence of contact mechanics on conjugation frequencies [65] and have also revealed enhanced gene transfer by transformation (uptake of DNA from the environment) in predator-prey communities [66,67].

Some investigations of community behaviour have relied on representative snapshots of spatial structure provided by single time-point (i.e., end-point) microscopy. This approach is useful when time-lapse approaches may not be feasible, such as in highly structured environments like biofilms and solid matrices [68], or in microdroplets [69]. (The recent time-lapse work of Hartmann et al. [70] and Nijjer et al. [71] characterizing biofilm growth is a notable exception and may represent a new paradigm for such measurements.) Data on three-dimensional arrangements within communities provide quantitative insights on how spatial distributions impact phenotype [72] and vice versa [73,74]. Co-occurrence networks in non-spatial environments can also be determined in microdroplets [10,69]. End-point cell arrangements constrained in two-dimensions have been used to measure interaction ranges of quorum-sensing mechanisms involved in horizontal gene transfer [75].

4.4 Calibration of Spatial Mathematical Models of Microbial Communities Against Single-cell Measurements

When mathematical models are employed to explore a range of possible behaviours, parameterizations need not accurately capture specific observations (e.g., [76–78]). In contrast, when models are used for predictive purposes (as in most engineering applications), models must be fit to observations. In such cases, descriptions of the formulation, calibration, and validation of the model are needed to specify the predictive strengths and limitations of the

model. A first step in communicating a model's formulation is statement of the model's purpose, which clarifies the scope of the model structure and parameterization. This is highlighted in the ODD (overview, design concepts, and details) protocol [79,80], a formal framework for documenting agent-based models (ABMs). The protocol has been used in documenting several spatiotemporal ABMs of microbial communities [25,81–83]. In macrobial ecology, this protocol is often used as part of a larger modelling methodology called pattern-oriented modelling [84,85], discussed further in Section 4.5.1.

Model calibration is the process of assigning values to model parameters to best reproduce available data. Ideally, calibration is complemented by uncertainty analysis, which gauges the degree of confidence in model predictions and parameter estimates through, for example, identifiability analysis and sensitivity analysis [86,87]. The simplest approach to model calibration is to characterize each component of a system independently. This approach is suitable for simple processes, such as growth or diffusion, for which direct measurements can be made.

In contrast, it is often the case that calibration of biological models must be posed as an *inverse problem*: properties of system components cannot be measured directly and must instead be inferred from observations of overall system behavior. In such cases, model calibration involves selection of a 'goodness-of-fit' function, typically defined as a sum of squared errors (SSE)—the SSE measure aligns with a maximum likelihood measure under idealized assumptions about system and noise structure [88]. When calibrating linear models, a rich theory provides robust uncertainty analysis, such as 95% confidence intervals for parameter estimates and model predictions. When addressing nonlinear dynamic models, the theory provides less support; models that minimize the SSE can be found only through nonlinear optimization procedures (typically iterative global optimization routines [89]), and uncertainty analysis is approximate (though there are uncertainty tools designed for nonlinear systems, e.g., profile likelihoods [90]). Bayesian calibration methods, such as approximate Bayesian computing [91], offer an alternative to global optimization searches. Bayesian

methods refine uncertainty distributions for model parameter values by comparing with experimental observations.

For non-spatial models (e.g., ODEs), systematic model calibration approaches have become standard in the field of computational biology, as reviewed in [86,87]. Such ODE models are used to describe microbial community dynamics through compartmentalization. For instance, Gupta et al. [61] used a compartmental ODE model to investigate the behaviour of physically separated microbial populations; they calibrated their model parameters using a standard SSE-minimizing approach.

4.4.1 A Survey of Calibration Approaches for Spatiotemporal Models

In this section, we survey strategies recently employed for calibration of spatiotemporal models of bacterial communities against observations at (or near) the single-cell level. The corresponding data (as described in Section 4.3) are complex, and calibration of these models is challenging. Calibration strategies used in recent publications can be roughly classified into three categories (Figure 4.3): manual fitting, systematic calibration to non-spatial data, and systematic calibration against spatial summary statistics.

The simplest calibration approach is manual fitting, by which model simulations are qualitatively compared to observed data, and model parameter values are adjusted to arrive at a satisfactory alignment. This approach, often used to extend previously established model structure, is exemplified by comments such as: “We chose model parameters to qualitatively fit the experimental results...” [51], and “We adjusted the parameters of our simulations until the behaviour matched the images of real cells...” [92]. Manual fitting is a pragmatic means to arrive at qualitative agreements between model and data, which is often perfectly appropriate for modelling objectives. However, it is poorly suited for situations in which precise calibration is required, it is unlikely to provide a robust search of high-dimensional model spaces, and it is unsatisfactory in terms of reproducibility.

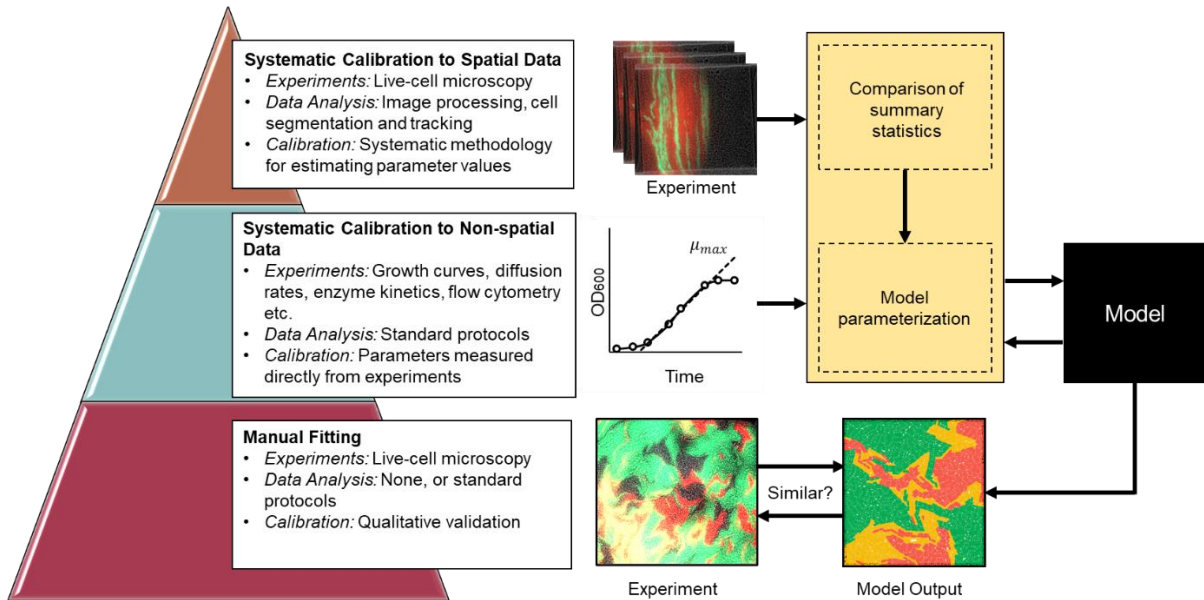


Figure 4.3. Model calibration techniques for spatiotemporal models of microbial communities. Manual fitting involves direct adjustment of parameter values to achieve qualitative agreement between model predictions and observations. Non-spatial calibration is often systematic (based on a goodness-of-fit function) but is based on experiments that do not incorporate the spatial features of the system. Spatial calibration, against spatially-distributed data, can be systematic (SSE-based), but must rely on summary statistics collected from the data.

A second common calibration tactic is to apply systematic calibration to non-spatial data compared with an aggregate model output. For example, in their study on the inhibitory role of antibiotic efflux activity from neighbouring cells, Wen et al. [59] observed interactions between two bacterial populations, one of which expressed antibiotic efflux pumps. They used an SSE-based approach to estimate growth and inhibition parameters from data obtained by suspension growth experiments. They then used those parameters in an ABM that supplemented findings from additional single-cell experiments. Another example is provided by the work of Pande et al. [93], who investigated the role of spatial segregation in cross-feeding populations. They used in-suspension growth curves obtained over a range of nutrient concentrations (via Monod growth kinetics) to fit growth parameters, which were then applied to a spatial ABM of the consortium. Such strategies rely on an assumption that the behaviours measured in suspension are representative of behaviour in the spatially structured environments under investigation.

Finally, in some instances, model developers have made full use of spatiotemporal data by systematic calibration against spatial summary statistics that capture the spatiotemporal aspects of primary interest, fitted with an SSE-based protocol. For example, Hartmann et al. [70] validated a three-dimensional cell tracking algorithm and calibrated an ABM by minimizing the error between measured and simulated summary statistics in a growing biofilm. Another example is provided by Leaman et al. [94], who collected summary statistics from spatial distributions of cells and then used a global optimization scheme to fit parameters of an ABM. Such systematic calibration approaches can be resource-intensive, often requiring detailed image processing pipelines (e.g., [70]), or numerous auxiliary experiments to fit physical or chemical parameters. For example, Leaman et al. [94] needed to measure both diffusivity of solute particles in the presence of bacteria and activation time for gene expression controlled by a quorum sensing molecule before calibrating the rest of the model's parameter values.

In the next section we present a collection of summary statistics that are suitable for validation of spatiotemporal models of bacterial community dynamics. Use of these summary statistics typically requires development of a data processing pipeline for image processing and summary statistic calculation.

4.4.2 A Catalogue of Spatiotemporal Summary Statistics for Microbial Community Dynamics

As discussed in Section 4.3, modelling projects are frequently built on spatiotemporal data that is rich and complex, resulting in a tendency to aim for qualitative agreement or calibration against non-spatial observations. Hartmann et al. [70] and Leaman et al. [94] provide examples that make more complete use of the richness of spatiotemporal data by selecting summary statistics to capture key spatial features in a quantitative manner and applying SSE-based calibration to ensure accurate representation. As we discuss below in Section 4.5.1, this strategy has been adopted for many modelling projects in the microbial ecology community, where spatiotemporal datasets of this type have been collected for decades. One of the challenges of this approach is identification of appropriate summary statistics. These should

(i) capture relevant features of the system's behaviour, (ii) be represented by model outputs, and (iii) be computationally tractable (in terms of image processing). In this section we survey summary statistics that have been used to capture spatiotemporal features of microbial dynamics (Table 4.1), along with some examples from macrobial ecology that hold promise for use in this context.

Table 4.1. Summary Statistics for Quantifying Features in Microbial Communities

Statistic	Goal of Statistic	Computational Details	Reference
Summary Statistics for Single-Species Populations			
<i>Microcolony shape</i>			
Microcolony aspect ratio	Quantify eccentricity of developing colony	Standard image processing feature, defined in 2D (or 3D, projected to 2D)	[70,95]
Dyad structure	Characterize structure of two-cell `colony` immediately before the 2 nd division	Normalized dot-product of the two cells' orientation	[95]
Biofilm base circularity	Characterize shape of the biofilm base	Unity minus aspect ratio of projection onto the horizontal plane	[70]
<i>Internal microcolony structure</i>			
Microcolony density	Quantify packedness of cells within developing colony	Standard image processing feature, defined locally or globally	[70,95–97]
Order parameter	Quantify anisotropy within developing colony	Mean of projections of orientation of neighbouring cells; defined per-cell, recorded as a colony average or as a distribution	[70,95–99]

Correlation length of scalar order parameter	Characterize ‘patchiness’: spatial scale over which orientation of neighbouring cells is aligned	Correlation of orientation as a function of distance; can be compared as a mean or a distribution	[98]
Micropatch area	Quantify ‘patchiness’; similar to correlation length of scalar order parameter	Cells are clustered into patches based on contact and relative orientation	[100]
Topological defect density	Characterize ‘patchiness’: density of topological defects (i.e. discontinuities in the order-parameter field)	Algorithm provided in [101]	[98]
Defect velocity	Characterizes the evolution of a microcolony's internal structure	The position of topological defects are tracked over time	[98,99]
Age distribution of cell poles within the developing microcolony	Characterize degree of mixing during colony development	Simple measure is distance from centre of colony to oldest cell poles. More complete measures additionally account for younger poles	[102]
<i>Other metrics</i>			
Orientation of cells at the microcolony boundary	Characterize tendency of boundary cells to align with the colony boundary	Colony boundary must be determined by a smoothing operation; cells on the boundary and the corresponding boundary orientation must be identified	[99]
Gradient of cell velocity normal to	Characterize growth inhibition	Measured by particle-image velocimetry	[96,97]

microcolony boundary	due to pressure gradients		
Cell-cell distance	Characterize cell spacing	Centroid-to-centroid distance to nearest neighbour	[70]
Vertical and radial alignment	Characterize 3D structure; identify transition from monolayer to multilayer growth	Angle formed by the z-axis and cell's major axis	[70]

Summary Statistics for Multi-species Populations

Composition

Single-strain population counts; population fractions	Captures population abundances	Cell counts for each population	[49,58,63,103,104]
Shannon species diversity index	Species biodiversity metric	Determined from cell counts for each population	[105]

Spatial configuration; nearest cell-cell adjacencies

Shannon entropy	Quantify randomness of pixel identities in an image	Total population of each species, heterospecific neighbouring cell counts	[35,105]
Intensity correlation quotient	Characterize colocalization or exclusion of pairs species in space	Determined by sum of pixel intensities for each fluorescence channel	[35]
Contagion index	Quantify dispersion and intermixing of different populations; deviation from maximum entropy state	Total population of each species, heterospecific neighbouring cell counts	[106]
Probability matrix for adjacent species identities	Global quantification for likelihood of nearest	Identify neighbour to cell centroid	[73]

	interspecific adjacencies		
Neighbour index	Characterize interspecific adjacencies relative to the initial adjacencies	Count physical contacts between pairs of cells of different phenotypes	[104]
<i>Spatial configuration; cell-cell adjacencies within a neighbourhood</i>			
Proportion of conspecific neighbours	Quantify interspecies mixing from a probability distribution; β -diversity metric	Probability that a cell is located a defined distance away from other members of its own species	[107,108]
Structure factor	Quantifies characteristic length scales of spatial patterning; used to characterize transition from well-mixed to structured populations	Normalized spatial Fourier transform of image data	[109]
Segregation index	Normalized metric to quantify population segregation	Cell neighbourhood interaction distance, heterospecific neighbouring cell counts	[110–113]
<i>Spatial configuration: shape</i>			
Colony edge roughness	Increased in communities with antagonistic interactions	Standard deviation of microcolony radius	[103]
Fractal dimension	“Jaggedness” of species boundaries	Distance of each pixel to nearest border of two different populations; algorithm provided in [114]	[115,116]
<i>Spatial configuration; sectors</i>			

Intermixing index	Estimate spatial mixing between multiple species	Average number of species transitions along a straight line or arc	[117,118]
Single-strain sector size	Indirect measurement of spatial mixing between multiple species	Length scale of single-strain patches	[104,119]
<i>Temporal</i>			
Order parameter; phase transitions	Quantify synchrony in gene expression between populations	Spatiotemporal data is processed into kymographs; algorithm introduced in [120]	[53]

Monolayer growth is the simplest set-up for observing single-cell characteristics of microbial population dynamics. In this setting, single-cell features such as elongation rate and division length threshold can be measured directly. The simplest population to study is an isolated microcolony descended from a single cell. Several groups have proposed summary statistics to capture development of such microcolonies. Volfson et al. [96] were one of the first to compare simulations of an ABM to time-lapse images of developing microcolonies within microfluidic devices. They calculated microcolony density, a cell velocity gradient, and an order parameter quantifying the global anisotropy in cell orientation. These summary statistics have been used to calibrate parameters governing physical interactions between rod-shaped bacteria in more recent ABM projects, e.g., [97]. Doumic et al. [95] used similar metrics in their model of microcolony growth that incorporates unequal mass distribution upon cell division. They also considered the microcolony aspect ratio and the relative orientation of the two daughter cells just prior to the second division (called ‘ d_2 ’ in Doumic et al., and dyad structure in Table 4.1). Doumic et al. highlight additional summary statistics for microcolony development: orientation of cells at the colony boundary (this is referred to as ‘active anchoring’ in [99]) and relative position with respect to age of individual cell poles within the colony [102].

Monoculture microcolony development has also been characterized using summary statistics from liquid crystal theory. These measures quantify the degree of physical alignment between neighbouring cells. The order parameter used by Volfson et al. [96] (mentioned above) is one such example. van Holthe tot Echten et al. [98] took the mean of this measure over the entire colony to compare the evolution of real microcolonies to ABM simulations. Dell’Arciprete et al. [99] use this same statistic, which they call the global order parameter, to summarize overall microcolony orientational disorder. Orientational order can also be quantified with discrete measurements. You et al. [100] investigate the inner structure of microcolonies by segmenting them into patches of similarly oriented cells and comparing the distribution of patch areas between ABM simulations and experiments. van Holthe tot Echten et al. [98] also use correlation length of the scalar order parameter (a measure of patch size) and topological defect density as additional measures of microcolony structure. (Topological defects are discontinuities in the orientation field that arise at boundaries between patches of similarly-oriented cells. These are locations that lack a representative trend in orientation.)

Studies of monolayer growth provide valuable insights into microbial activity, but they represent an idealized version of microcolony formation. In contrast, Hartmann et al. [70] present a comprehensive study of biofilm formation in three dimensions. They present novel imaging and image-processing tools that allow single-cell level tracking of a *V. cholera* biofilm from a single progenitor to about 10,000 cells. To quantify growth of this population, they make use of a collection of spatial summary statistics: vertical and radial alignment, local order, cell-to-cell distance, density, and aspect ratio of overall population and biofilm base.

In multi-species communities, population counts of each species are a simple, key summary statistic that encapsules population dynamics (Figure 4.4a). Microbial interactions that have been summarized using population fractions include secretion of nutrients and toxins [103], cell lysis by T6SS [63], and competition for space in microfluidic traps [49]. While population fractions are often measured globally, localized measures are also used. For example, Bottery et al. [104] measured population fractions as a function of a microcolony radius, while Dal Co et al. [58] measured population fractions within a given radius for each

cell to quantify the length scale of interactions mediated by secretion of diffusible molecules. In microbial communities consisting of a large number of species, an index summarizing biodiversity, such as the Shannon diversity index [105], may be more informative than specific population fractions.

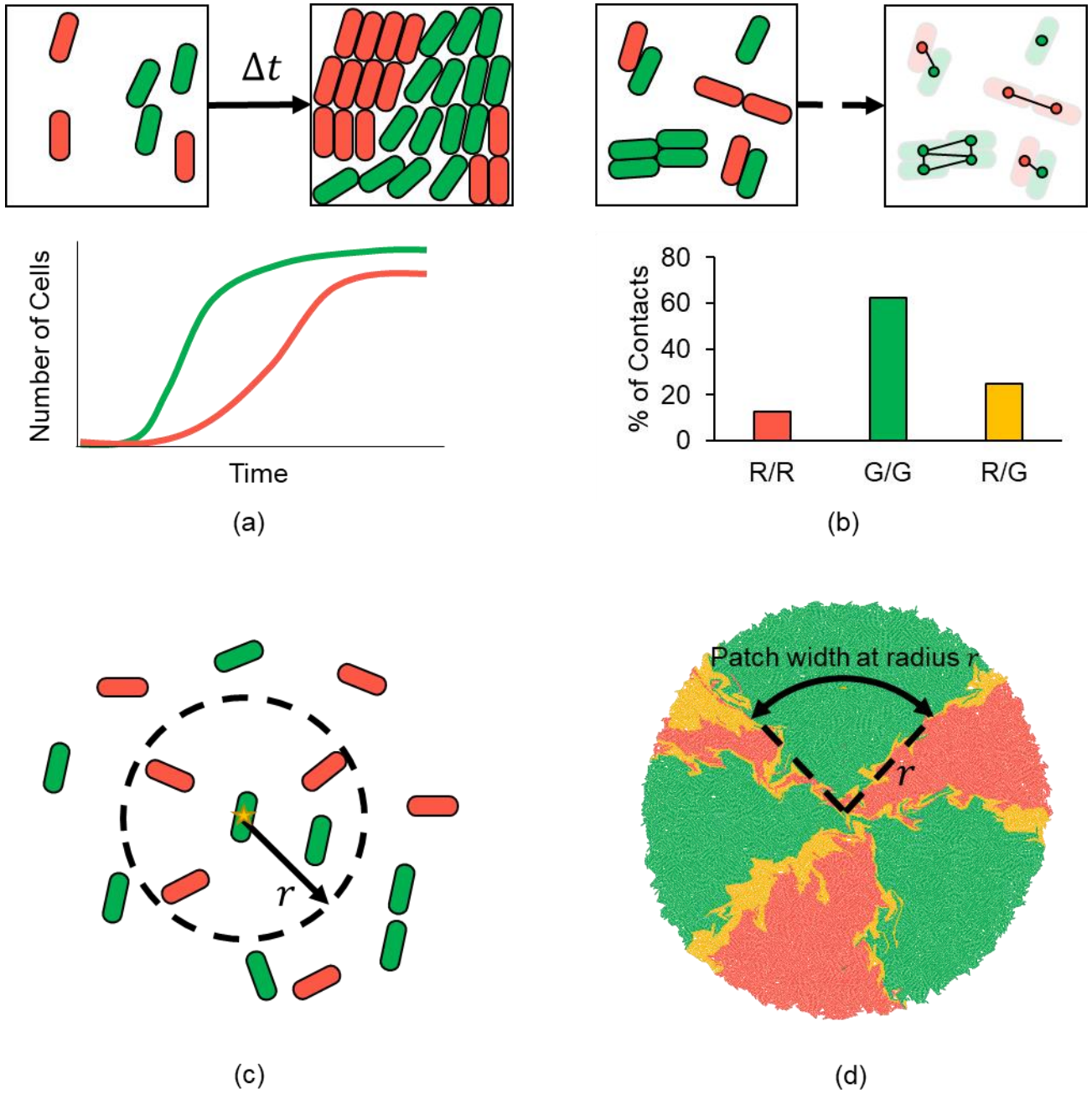


Figure 4.4. Calculating spatiotemporal summary statistics for microbial communities. (a) Population counts over time capture the overall dynamics in a multi-species community. (b) The frequency of adjacent species in physical contact, determined by a contact network, provides a measure of intermixing between different species. (c) Summary statistics can be calculated from data averaged within a particular cell's neighbourhood. (d) Single-species patch metrics, such as patch width and number of sectors, are useful for quantifying spatial patterns on a larger colony scale.

Most summary statistics that describe spatial patterns in multi-species communities quantify the degree of inter-species mixing or (conversely) of mono-species patchiness. These measures are commonly used in the field of landscape ecology [121]. Metrics from landscape ecology traditionally rely on counting adjacencies between image pixels, each of which is assigned a value corresponding to its dominant occupant. These metrics can be applied in the same manner for low-magnification microscopy images of microbial colonies. Shannon entropy is a canonical metric for species mixing [105]; it quantifies the overall disorder between any number of populations by counting like- and non-alike pixel adjacencies. Kong et al. [35] used this measure to assess the extent of red-green pixel colocalization in two-strain microbial communities from microscopy images taken at $7\times$ magnification. Li and Reynolds [106] developed a contagion index [122] that quantifies the deviation from the maximum entropy state using the same type of pixel adjacency counts. This contagion index is used widely in landscape ecology because it captures both aggregation of single populations and intermixing of different populations.

Landscape ecology metrics could be extended to higher-magnification single-cell data by generating a physical contact network and accounting for non-rectangular adjacency structure (Figure 4.4b). Alternatively, single-cell images can be smoothed until continuous single-species patches are formed [104,119]. Other metrics defining patch shape, aggregation, and species/strain diversity (discussed below) could also translate to the single-cell level. Mony et al. [123] discuss applications of other higher level principles from landscape ecology to analysis of microbial community assembly and structure.

While the contagion index has not yet been applied to microbial studies, related measures have been used. For example, Bottery et al. [104] counted physical contacts in pairs

of cells of differing strain/species (Figure 4.4b). They normalized these counts to initial neighbour counts, arriving at a metric they called the neighbour index. An alternative intermixing measure that does not require counting all physical contacts between cells is a probability matrix for adjacent species identities, computed by identifying the species/strain of a cell's nearest neighbour. Glass and Riedel-Kruse [73] used this type of measurement to quantify effects of surface nanobodies and antigens on cell-cell adhesion.

Summary statistics that describe proportions of species within some defined neighbourhood (Figure 4.4c) are also used to describe intermixing of microbial populations. The segregation index [110–113] measures the degree to which cells within a given neighbourhood radius are related to one another (by genotype or phenotype). In this case, the radius is defined as the distance over which interactions mediated by small molecules are expected to equally influence all cells within the neighbourhood [113]. The segregation index has been applied to simulated data in numerous microbial ABM studies but has yet to see use in the context of single-cell microscopy data. Generalizing this measure, the proportion of conspecific neighbours is defined as the probability that two randomly selected individuals separated by some defined distance will belong to the same population [107]. Computing this metric over a large sample of individual pairs provides the proportion of conspecific neighbours as a function of distance. An alternative way to define a neighbourhood is by a characteristic length scale. McNally et al. [109] used a static structure factor to identify transitions from well-mixed to segregated states in antagonistic two-strain communities. This metric was computed using Fourier transforms of binarized pixel intensities to assess spatial (patch size) frequencies of each strain within a characteristic length scale.

Other metrics for defining intermixing on larger scales use the number of single-species patches as a measure of interspecific mixing. For example, the intermixing index is determined by the average number of single-species patch transitions along a line or arc. This metric has been used as a measure of species colocalization in low-magnification images of biofilms and microbial colonies [117,118]. Blanchard and Lu [103] and Bottery et al. [104] used the number

of single-strain sectors of a circular colony to characterize spatial patterns in two-strain communities growing on a surface with open boundary conditions (Figure 4.4d).

Some spatial patterns may be observable from the physical shape of single-species sectors or colony boundaries (Figure 4.4d). Kan et al. [115] and Rudge et al. [116] measured the fractal dimension of species patch boundaries, which quantifies jaggedness. Blanchard and Lu [103] noted that the roughness of a growing colony's edge increases when there are antagonistic interactions between different strains. Amor et al. [119] and Bottery et al. [104] used sector widths as an indirect measurement of spatial mixing, because larger widths imply less mixing. The perimeter-to-area ratio of single-species sectors could also be appropriate as a summary statistic for shape [121], although it has not been used yet in microbial studies.

The summary statistics described above are applicable to end-point measurements. Of course, these can be measured through times series, but alternative measures rely explicitly on time series, e.g., through windowed averages and autocorrelation [124]. Periodicity can also be used as a temporal metric, quantified by, e.g., a periodic order parameter, as demonstrated by Kim et al. [53], who summarized spatiotemporal synchronization of gene expression in a two-strain community. Time derivatives of summary statistics can also be assessed. For example, Dell'Arciprete et al. [99] and Van Holthe tot Echten et al. [98], discussed above, both use the velocity of topological defects to characterize microcolony dynamics.

4.5 Outlooks

There is no doubt that spatiotemporal models of microbial communities will continue to grow in complexity (and corresponding computational requirements) as researchers continue making advances in synthetic ecology, microbiome engineering, and in characterizing natural systems. In this section, we survey some outlooks for standardizing and streamlining model development and validation.

4.5.1 Pattern-Oriented Modelling as a Guideline for Standardizing Microbiological Models

It can be challenging to describe agent/individual-based models efficiently, but complete descriptions are crucial; incomplete reporting leads to difficulties with subsequent implementation and replication, as demonstrated by Donkin et al. [125] and discussed in [124,126]. Furthermore, systematic model documentation can improve model quality by enforcing critical thinking about the model’s objective, formulation, implementation, and validation. In surveying modelling practices for microbial communities, we found that modelling and documentation practices vary considerably, especially regarding model calibration. A systematic framework for model development and testing, referred to as pattern-oriented modelling (POM) [84,85], sees frequent use in macrobial ecology and has been used occasionally in microbial settings as well [127–130].

POM addresses “the multi-criteria design, selection and calibration of models of complex systems” [85]. The framework formalizes all stages of the modelling pipeline, from model formulation, to testing, to calibration and validation. The ‘patterns’ in POM are any quantifiable features of model simulations; we referred to these as summary statistics in Section 4.4. These measures are most useful when they span ecological scales: individual, population, community, ecosystem. As highlighted by the POM framework, summary statistics facilitate validation by reducing system dimensionality [131]. Moreover, they can guide model formulation by focusing attention on the aspects of simulations that will be quantitatively captured.

POM’s model validation strategy is standard [124,131]: begin with qualitative comparison of model predictions with experimental data, then sample the parameter space to determine the sensitivity of summary statistics to parameter values (typically done in a one-at-a-time fashion, given computational costs) (e.g., [108]). Acceptable parameter fits are then determined based on systematic minimization of SSE quality-of-fit measures using a weighted average of the summary statistics, as in, e.g., [132–134]. Documentation of all model

formulations and parameter sets tested can provide insights into model behaviour and can potentially reveal underlying mechanisms of emergent community properties.

4.5.2 Feature Identification through Topological Data Analysis

The selection of appropriate ‘patterns’ is a subjective task, as acknowledged by the architects of POM [85]. Moreover, it is not always clear how best to quantify these patterns as summary statistics once they have been identified. Some features are easy to represent numerically (e.g., average population density), but many relevant patterns are qualitative, or manifest as complex spatiotemporal configurations. In some cases, existing theory can offer tools to quantify these features, such as order parameters from liquid crystals, or Fourier coefficients to identify feature scales (both described in Section 4.4). In the absence of such tools, many researchers rely on visual inspection, which introduces subjectivity into the calibration pipeline. A generic approach to pattern identification is provided by the recently developed tools of topological data analysis (TDA).

TDA provides tools to quantify topological (i.e., qualitative) features within datasets. It can be applied either to discrete datasets, like those from ABMs, or to continuous data, like those from PDE models. TDA encompasses a wide variety of tools and techniques. Here we focus on the most popular: persistence homology. (For a general overview of the field, see [135]; a broad discussion of applications to biology is presented in [136].) Persistence homology can be thought of as a non-linear analogue of a more familiar technique: principle component analysis (PCA) [137]. PCA is used to identify the variational structure within datasets: if there are correlations within the data, the points will tend to cluster around certain linear subspaces (lines, planes, etc.) and the data will exhibit less variation in the directions perpendicular to these subspaces. One application of PCA is dimensionality reduction. Data can be projected onto these linear subspaces, thereby reducing the dimensionality of the dataset with minimal information loss.

Whereas PCA identifies linear structure within a dataset, persistence homology identifies arbitrary nonlinear structure. A basic persistence homology workflow for discrete

data can be described as follows (Figure 4.5). First, data is represented as a collection of points in some (typically) high-dimensional space that characterizes features of interest. If, for example, TDA were to be used to characterize the results of an agent-based simulation, each point might represent an individual agent, with the coordinates corresponding to features of that agent: e.g., species, position, length, orientation. To proceed, a length scale L is chosen, a ball of radius L is constructed around each point, and the topological features of the shape thus produced are determined, e.g., connected components and loops. This analysis is repeated over a wide range of length scales (Figure 4.5a-c). At small scales, it will produce only a cloud of disconnected points. As the length scale, L , increases, neighbouring balls intersect, forming larger and larger structures until, finally, they merge into one fully connected component. As L varies, the length scales at which various topological features occur (i.e., over which they persist) is recorded. Each topological feature is thus associated to a pair of numbers: the smallest and largest scales at which the feature exists. These can be represented by a persistence ‘barcode’ in which bars are plotted against length scale. Each bar corresponds to topological feature; the bars represent the length scales over which the features occur. (In Figure 5d, the teal bars end at scales at which connected components merge.) As an alternative visualization, the pairs of length scales associated to each topological feature can be plotted in a persistence diagram (Figure 5e; any features near the diagonal occur over only a short range of length scales and may be dismissed as spurious). A similar analysis can be applied to continuous data (e.g., from a PDE model) by characterizing the topology of the sublevel sets of a continuous function, such as population density [135].

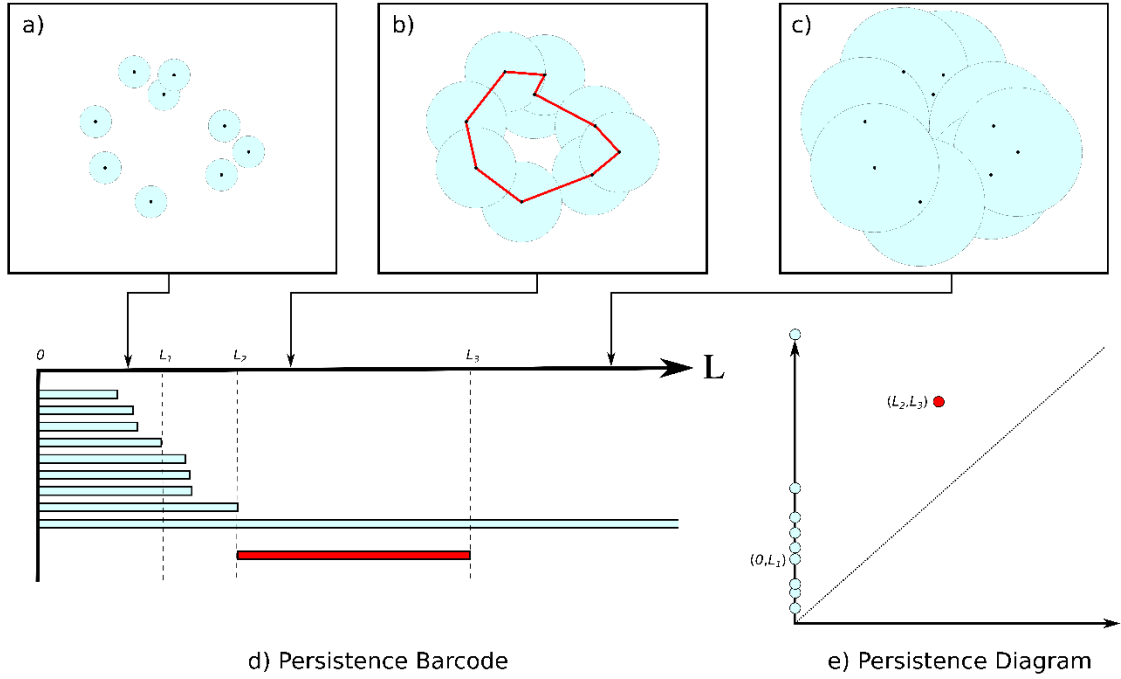


Figure 4.5. Persistence homology. Panels (a-c) illustrate the topology of a dataset changing as the length scale, L is varied. (a) For small values of L , the balls (disks) are mostly disconnected; only two of the nine intersect. (b) At an intermediate scale, all nine balls intersect, forming a single connected component, giving rise to a loop. (c) At larger scales, there is a single connected component and no loop. (d) The progression illustrated in (a-c) is documented in the persistence barcode; the blue bars correspond to separate connected components, the ends of which corresponds to intersection (merge) events, e.g., at $L = L_1$. The red bar corresponds to the loop which forms at $L = L_2$ and which becomes filled in at $L = L_3$. (e) The same information can be represented in persistence diagram in which the (x,y) coordinates of points correspond to the right and left ends, respectively, of each bar in the barcode.

Much like PCA, persistence homology provides a natural way to reduce the dimensionality of the dataset. Features that persist over a narrow range of length scales are likely spurious and can be discarded. The features that persist over a wide range of length scales more likely correspond to meaningful structure in the dataset: connected components might correspond to discrete clusters, loops to periodicity. Persistence diagrams can be compared to one another using a variety of metrics, allowing them to be used directly for model calibration. They can also serve as a starting point for development of custom summary statistics.

This type of analysis was applied by Topaz et al. [138] to study ABMs demonstrating swarming behaviour. In that case, the agents each have a position and a velocity, so clusters of points in position-velocity space correspond to swarms—closely grouped agents exhibiting collective motion. Previously, researchers had developed case-specific order parameters to quantify such behaviour [139–141]. By applying TDA, Topaz et al. [138] were able to detect features that the previous *ad hoc* metrics failed to quantify. Similar approaches could be applied to characterize the dynamics of microbial communities.

4.5.3 Machine Learning Algorithms to Accelerate Model Calibration

Conducting global parameter sweeps over high-dimensional parameter spaces is often infeasible due to computational limitations. For spatiotemporal models, this problem occurs due to long simulation runtimes and is exacerbated further for stochastic models, where large simulation ensembles may be required. One approach to address this challenge is to create a simplified input-output representation of the mathematical model of interest. Such an abstraction is known as a surrogate model (also commonly referred to as a metamodel, or emulator). A surrogate model is constructed by fitting a statistical model or machine learning model to training data generated from the mathematical model of interest. Surrogate models can run several orders of magnitude faster than ABM or PDE models that they are fit to, which has motivated their application in calibration of spatiotemporal models of microbial communities [142–144]. A surrogate model enables faster exploration of the parameter space for both simulation and uncertainty analysis, albeit at the cost of relying on emulator approximation of model behavior. Although this approach is relatively new to microbial community modelling, surrogate models have been used extensively in other fields, including engineering design [145], climate simulation [146], health economics and public health [147], and ecology [148].

Surrogate models can provide significantly more efficient simulation engines when compared to the original model formulations. However, when comparing computational costs, the time required to obtain the training data for the surrogate model must be considered. The time required for training depends on several factors, such as parameter space complexity,

sampling methods for parameter values, and computational cost of the original model. For example, surrogate models of an ABM for biofilm growth required approximately 1000 hours of serial computing time (Table 4.2) [143,144]. In contrast, a surrogate model of a PDE model describing spatiotemporal dynamics of pattern-forming bacteria required approximately 100000 hours of serial computing time [142]. These training times can be reduced through parallelization. For example, the actual time required to generate the training data for the works listed in Table 4.2 would range approximately from 1-7 days if 64 simulations were constantly run in parallel.

Table 4.2. Requirements for Generating Training Data for Microbial Community Models

Model Type	Parameters	Simulations Required	Time per Simulation (h)	Serial Computing Time (h)	Est. Parallel Computing Time on 64 CPU cores (h)	Reference
PDE	231	100000	0.0972	9720	152	[142]
ABM	32	300	5-6	1500-1800	23-28	[143]
ABM	7	100	6-8	600-800	9-13	[144]

The use of surrogates for calibrating microbial community models requires familiarity with data sampling methods, supervised machine learning algorithms, and ‘big data’ processing tools. Such projects may demand collaboration between data scientists and modellers who want to access these tools. Software packages, such as SUMO [149], SMT [150], and *spartan* [151], are available to facilitate the process of generating surrogate models. A brief discussion of surrogate model selection for ABM applications is provided in [152].

Surrogate models are not the only use for machine learning in this area. Lee et al. [153] reversed the standard modelling pipeline (of training a model on experimental data) by training a neural network to ABM data, and then using that network to infer microbial interactions from microscopy data. They demonstrated that the magnitude and direction of interspecific interactions could be quantified from steady-state spatial distributions of two interacting bacterial populations. Their work provides a new perspective on the use of mathematical

models and machine learning to supplement experiments on microbial communities. Further advances in modelling and computer science may reveal previously unexplored features of the rich experimental data associated with single-cell observations.

4.6 Conclusions

Potential applications of microbial communities in biotechnology, health, agriculture, and energy have motivated efforts to design and manipulate both synthetic and natural communities in a predictable fashion. Predictive tools such as ABMs and PDE models will be an essential part of the microbiome engineering toolbox. Systematically calibrating microbial community models to single-cell resolution data is challenging due to the high-dimensionality of the data, the intensive image processing requirements, and the specific data processing algorithms required to generate summary statistics. The study of Hartmann et al. [70] exemplifies how single-cell data collection and application of systematic calibration techniques can be used to predict community-level properties from single-cell behaviours.

We have compiled a collection of summary statistics relevant to microbial communities to facilitate quantitative comparison between experimental data and simulation outputs. Systematic calibration with the aid of these summary statistics can increase confidence in model predictions and overall model utility. Moreover, such calibration can improve the reusability of sub-models and specific parameter values by allowing developers to confidently use or build upon these calibrated models. This modular approach is already used in synthetic circuit design workflows (e.g., [154]).

Work in microbial ecology demonstrates how adopting standard model documentation procedures (e.g., the ODD protocol and POM framework) can result in more reproducible and therefore more useful models. Adoption of a similar standard in microbial ecology could yield similar boons in reproducibility. Topological data analysis (TDA) and machine learning algorithms hold potential for facilitating systematic selection of summary statistics and efficient exploration of high-dimensional parameter spaces. As experimental methods and computational techniques continue to improve, it is expected that models will play a prominent

role in rationally manipulating microbial communities in complex environments such as bioreactors, guts, soils, and wastewater treatment plants.

4.7 References

1. Xu J. Invited review: Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol.* 2006;15: 1713–1731. doi:10.1111/j.1365-294X.2006.02882.x
2. Rousk J, Bengtson P. Microbial regulation of global biogeochemical cycles. *Front Microbiol.* 2014;5: 103. doi:10.3389/fmicb.2014.00103
3. Newman DK, Banfield JF. Geomicrobiology: How Molecular-Scale Interactions Underpin Biogeochemical Systems. *Science.* 2002;296: 1071–1077. doi:10.1126/science.1010716
4. Wang B, Yao M, Lv L, Ling Z, Li L. The Human Microbiota in Health and Disease. *Engineering.* 2017;3: 71–82. doi:10.1016/J.ENG.2017.01.008
5. Sabra W, Dietz D, Tjahjajari D, Zeng A-P. Biosystems analysis and engineering of microbial consortia for industrial biotechnology. *Eng Life Sci.* 2010;10: 407–421. doi:10.1002/elsc.201000111
6. Qureshi N, Annous BA, Ezeji TC, Karcher P, Maddox IS. Biofilm reactors for industrial bioconversion processes: employing potential of enhanced reaction rates. *Microb Cell Fact.* 2005;4: 24. doi:10.1186/1475-2859-4-24
7. Borody TJ, Khoruts A. Fecal microbiota transplantation and emerging applications. *Nat Rev Gastroenterol Hepatol.* 2012;9: 88–96. doi:10.1038/nrgastro.2011.244
8. Lee ED, Aurand ER, Friedman DC. Engineering Microbiomes—Looking Ahead. *ACS Synth Biol.* 2020;9: 3181–3183. doi:10.1021/acssynbio.0c00558
9. Werner GDA, Strassmann JE, Ivens ABF, Engelmoer DJP, Verbruggen E, Queller DC, et al. Evolution of microbial markets. *Proc Natl Acad Sci.* 2014;111: 1237–1244. doi:10.1073/pnas.1315980111
10. Hsu RH, Clark RL, Tan JW, Ahn JC, Gupta S, Romero PA, et al. Microbial Interaction Network Inference in Microfluidic Droplets. *Cell Syst.* 2019;9: 229–242.e4. doi:10.1016/j.cels.2019.06.008
11. Strickland MS, Lauber C, Fierer N, Bradford MA. Testing the functional significance of microbial community composition. *Ecology.* 2009;90: 441–451. doi:10.1890/08-0296.1
12. Galand PE, Pereira O, Hochart C, Auguet JC, Debross D. A strong link between marine microbial community composition and function challenges the idea of functional redundancy. *ISME J.* 2018;12: 2470–2478. doi:10.1038/s41396-018-0158-1
13. Rivett DW, Bell T. Abundance determines the functional role of bacterial phylotypes in complex communities. *Nat Microbiol.* 2018;3: 767–772. doi:10.1038/s41564-018-0180-0
14. Yanni D, Márquez-Zacarías P, Yunker PJ, Ratcliff WC. Drivers of Spatial Structure in Social Microbial Communities. *Curr Biol.* 2019;29: R545–R550. doi:10.1016/j.cub.2019.03.068
15. Kim W, Racimo F, Schluter J, Levy SB, Foster KR. Importance of positioning for microbial evolution. *Proc Natl Acad Sci.* 2014;111: E1639–E1647. doi:10.1073/pnas.1323632111

16. Gorter FA, Manhart M, Ackermann M. Understanding the evolution of interspecies interactions in microbial communities. *Philos Trans R Soc B Biol Sci.* 2020;375: 20190256. doi:10.1098/rstb.2019.0256
17. Pagaling E, Strathdee F, Spears BM, Cates ME, Allen RJ, Free A. Community history affects the predictability of microbial ecosystem development. *ISME J.* 2014;8: 19–30. doi:10.1038/ismej.2013.150
18. Fukami T. Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. *Annu Rev Ecol Evol Syst.* 2015;46: 1–23. doi:10.1146/annurev-ecolsys-110411-160340
19. Lindemann SR, Bernstein HC, Song H-S, Fredrickson JK, Fields MW, Shou W, et al. Engineering microbial consortia for controllable outputs. *ISME J.* 2016;10: 2077–2084. doi:10.1038/ismej.2016.26
20. Succurro A, Ebenhöf O. Review and perspective on mathematical modeling of microbial ecosystems. *Biochem Soc Trans.* 2018;46: 403–412. doi:10.1042/BST20170265
21. Ravikrishnan A. Systems-level modelling of microbial communities: theory and practice. Raman K, editor. Boca Raton: CRC Press, Taylor & Francis Group; 2019.
22. Cao X, Hamilton JJ, Venturelli OS. Understanding and Engineering Distributed Biochemical Pathways in Microbial Communities. *Biochemistry.* 2019;58: 94–107. doi:10.1021/acs.biochem.8b01006
23. Naylor J, Fellermann H, Krasnogor N. Easybiotics: a GUI for 3D physical modelling of multi-species bacterial populations. *Bioinformatics.* 2019;35: 3859–3860. doi:10.1093/bioinformatics/btz131
24. Dukovski I, Bajić D, Chacón JM, Quintin M, Vila JCC, Sulheim S, et al. A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS). *Nat Protoc.* 2021;16: 5030–5082. doi:10.1038/s41596-021-00593-3
25. Bogdanowski A, Banitz T, Muhsal LK, Kost C, Frank K. McComedy: A user-friendly tool for next-generation individual-based modeling of microbial consumer-resource systems. *PLOS Comput Biol.* 2022;18: e1009777. doi:10.1371/journal.pcbi.1009777
26. Hellweger FL, Clegg RJ, Clark JR, Plugge CM, Kreft J-U. Advancing microbial sciences by individual-based modelling. *Nat Rev Microbiol.* 2016;14: 461–471. doi:10.1038/nrmicro.2016.62
27. Kreft J-U, Plugge CM, Grimm V, Prats C, Leveau JHJ, Banitz T, et al. Mighty small: Observing and modeling individual microbes becomes big science. *Proc Natl Acad Sci.* 2013;110: 18027–18028. doi:10.1073/pnas.1317472110
28. Grimm V, Revilla E, Berger U, Jeltsch F, Mooij WM, Railsback SF, et al. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science.* 2005;310: 987–991. doi:10.1126/science.1116681
29. Potvin-Trottier L, Luro S, Paulsson J. Microfluidics and single-cell microscopy to study stochastic processes in bacteria. *Curr Opin Microbiol.* 2018;43: 186–192. doi:10.1016/j.mib.2017.12.004
30. Wessel AK, Hmelo L, Parsek MR, Whiteley M. Going local: technologies for exploring bacterial microenvironments. *Nat Rev Microbiol.* 2013;11: 337–348. doi:10.1038/nrmicro3010

31. Bengtsson-Palme J. Microbial model communities: To understand complexity, harness the power of simplicity. *Comput Struct Biotechnol J*. 2020;18: 3987–4001. doi:10.1016/j.csbj.2020.11.043
32. Becton Dickinson and Company. FlowJo™ Software Version 10. 2021. Available: <https://docs.flowjo.com/flowjo/>
33. Zuba-Surma EK, Kucia M, Abdel-Latif A, Lillard JW, Ratajczak MZ. The ImageStream System: a key step to a new era in imaging. *Folia Histochem Cytobiol*. 2007;45: 279–290.
34. Young JW, Locke JCW, Altinok A, Rosenfeld N, Bacarian T, Swain PS, et al. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc*. 2012;7: 80–88. doi:10.1038/nprot.2011.432
35. Kong W, Meldgin DR, Collins JJ, Lu T. Designing microbial consortia with defined social interactions. *Nat Chem Biol*. 2018;14: 821–829. doi:10.1038/s41589-018-0091-7
36. Malwade A, Nguyen A, Sadat-Mousavi P, Ingalls BP. Predictive Modeling of a Batch Filter Mating Process. *Frontiers in Microbiology*. 2017. p. 461. doi:10.3389/fmicb.2017.00461
37. del Campo I, Ruiz R, Cuevas A, Revilla C, Vielva L, de la Cruz F. Determination of conjugation rates on solid surfaces. *Plasmid*. 2012;67: 174–182. doi:10.1016/j.plasmid.2012.01.008
38. Sanchez A, Gore J. Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations. *PLOS Biol*. 2013;11: e1001547. doi:10.1371/journal.pbio.1001547
39. Granato ET, Foster KR. The Evolution of Mass Cell Suicide in Bacterial Warfare. *Curr Biol*. 2020;30: 2836–2843.e3. doi:10.1016/j.cub.2020.05.007
40. Berk V, Fong JCN, Dempsey GT, Develioglu ON, Zhuang X, Liphardt J, et al. Molecular Architecture and Assembly Principles of *Vibrio cholerae* Biofilms. *Science*. 2012;337: 236–239. doi:10.1126/science.1222981
41. Bridier A, Briandet R. Contribution of Confocal Laser Scanning Microscopy in Deciphering Biofilm Tridimensional Structure and Reactivity BT. In: Donelli G, editor. *Microbial Biofilms: Methods and Protocols*. New York, NY: Springer New York; 2014. pp. 255–266. doi:10.1007/978-1-4939-0467-9_18
42. Joyce G, Robertson BD, Williams KJ. A modified agar pad method for mycobacterial live-cell imaging. *BMC Res Notes*. 2011;4: 73. doi:10.1186/1756-0500-4-73
43. Wang X, Koehler SA, Wilking JN, Sinha NN, Cabeen MT, Srinivasan S, et al. Probing phenotypic growth in expanding *Bacillus subtilis* biofilms. *Appl Microbiol Biotechnol*. 2016;100: 4607–4615. doi:10.1007/s00253-016-7461-4
44. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, et al. Robust Growth of *Escherichia coli*. *Curr Biol*. 2010;20: 1099–1103. doi:10.1016/j.cub.2010.04.045
45. Ulman V, Maška M, Magnusson KEG, Ronneberger O, Haubold C, Harder N, et al. An objective comparison of cell-tracking algorithms. *Nat Methods*. 2017;14: 1141–1152. doi:10.1038/nmeth.4473
46. Jeckel H, Drescher K. Advances and opportunities in image analysis of bacterial cells and communities. *FEMS Microbiol Rev*. 2021;45. doi:10.1093/femsre/fuaa062

47. Weitz M, Mückl A, Kapsner K, Berg R, Meyer A, Simmel FC. Communication and Computation by Bacteria Compartmentalized within Microemulsion Droplets. *J Am Chem Soc.* 2014;136: 72–75. doi:10.1021/ja411132w
48. Chen Y, Kim JK, Hirning AJ, Josić K, Bennett MR. Emergent genetic oscillations in a synthetic microbial consortium. *Science.* 2015;349: 986–989. doi:10.1126/science.aaa3794
49. Alnahhas RN, Winkle JJ, Hirning AJ, Karamched B, Ott W, Josić K, et al. Spatiotemporal Dynamics of Synthetic Microbial Consortia in Microfluidic Devices. *ACS Synth Biol.* 2019;8: 2051–2058. doi:10.1021/acssynbio.9b00146
50. Alnahhas RN, Sadeghpour M, Chen Y, Frey AA, Ott W, Josić K, et al. Majority sensing in synthetic microbial consortia. *Nat Commun.* 2020;11: 3659. doi:10.1038/s41467-020-17475-z
51. Scott SR, Din MO, Bittihn P, Xiong L, Tsimring LS, Hasty J. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nat Microbiol.* 2017;2: 17083. doi:10.1038/nmicrobiol.2017.83
52. van Vliet S, Hol FJH, Weenink T, Galajda P, Keymer JE. The effects of chemical interactions and culture history on the colonization of structured habitats by competing bacterial populations. *BMC Microbiol.* 2014;14: 116. doi:10.1186/1471-2180-14-116
53. Kim JK, Chen Y, Hirning AJ, Alnahhas RN, Josić K, Bennett MR. Long-range temporal coordination of gene expression in synthetic microbial consortia. *Nat Chem Biol.* 2019;15: 1102–1109. doi:10.1038/s41589-019-0372-9
54. Hol FJH, Voges MJ, Dekker C, Keymer JE. Nutrient-responsive regulation determines biodiversity in a colicin-mediated bacterial community. *BMC Biol.* 2014;12: 68. doi:10.1186/s12915-014-0068-2
55. Massalha H, Korenblum E, Malitsky S, Shapiro OH, Aharoni A. Live imaging of root–bacteria interactions in a microfluidics setup. *Proc Natl Acad Sci.* 2017;114: 4549 LP – 4554. doi:10.1073/pnas.1618584114
56. Chuang JS, Frentz Z, Leibler S. Homeorhesis and ecological succession quantified in synthetic microbial ecosystems. *Proc Natl Acad Sci.* 2019;116: 14852–14861. doi:10.1073/pnas.1901055116
57. Dal Co A, van Vliet S, Ackermann M. Emergent microscale gradients give rise to metabolic cross-feeding and antibiotic tolerance in clonal bacterial populations. *Philos Trans R Soc B Biol Sci.* 2019;374: 20190080. doi:10.1098/rstb.2019.0080
58. Dal Co A, van Vliet S, Kiviet DJ, Schlegel S, Ackermann M. Short-range interactions govern the dynamics and functions of microbial communities. *Nat Ecol Evol.* 2020;4: 366–375. doi:10.1038/s41559-019-1080-2
59. Wen X, Langevin AM, Dunlop MJ. Antibiotic export by efflux pumps affects growth of neighboring bacteria. *Sci Rep.* 2018;8: 15120. doi:10.1038/s41598-018-33275-4
60. Moffitt JR, Lee JB, Cluzel P. The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab Chip.* 2012;12: 1487–1494. doi:10.1039/C2LC00009A

61. Gupta S, Ross TD, Gomez MM, Grant JL, Romero PA, Venturelli OS. Investigating the dynamics of microbial consortia in spatially structured environments. *Nat Commun.* 2020;11: 2418. doi:10.1038/s41467-020-16200-0
62. LeRoux M, De Leon JA, Kuwada NJ, Russell AB, Pinto-Santini D, Hood RD, et al. Quantitative single-cell characterization of bacterial interactions reveals type VI secretion is a double-edged sword. *Proc Natl Acad Sci.* 2012;109: 19804–19809. doi:10.1073/pnas.1213963109
63. Smith WPJ, Vettiger A, Winter J, Ryser T, Comstock LE, Basler M, et al. The evolution of the type VI secretion system as a disintegration weapon. *PLOS Biol.* 2020;18: e3000720. doi:10.1371/journal.pbio.3000720
64. Steinbach G, Crisan C, Ng SL, Hammer BK, Yunker PJ. Accumulation of dead cells from contact killing facilitates coexistence in bacterial biofilms. *J R Soc Interface.* 2020;17: 20200486. doi:10.1098/rsif.2020.0486
65. Seoane J, Yankelevich T, Dechesne A, Merkey B, Sternberg C, Smets BF. An individual-based approach to explain plasmid invasion in bacterial populations. *FEMS Microbiol Ecol.* 2011;75: 17–27. doi:10.1111/j.1574-6941.2010.00994.x
66. Cooper RM, Tsimring L, Hasty J. Inter-species population dynamics enhance microbial horizontal gene transfer and spread of antibiotic resistance. Cooper B, editor. *Elife.* 2017;6: e25950. doi:10.7554/eLife.25950
67. Borgeaud S, Metzger LC, Scignari T, Blokesch M. The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. *Science.* 2015;347: 63–67. doi:10.1126/science.1260064
68. Hol FJH, Whitesides GM, Dekker C. Bacteria-in-paper, a versatile platform to study bacterial ecology. *Ecol Lett.* 2019;22: 1316–1323. doi:10.1111/ele.13274
69. Hengoju S, Tovar M, Man DKW, Buchheim S, Rosenbaum MA. Droplet Microfluidics for Microbial Biotechnology. *Advances in Biochemical Engineering/Biotechnology.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2020. pp. 1–29. doi:10.1007/10_2020_140
70. Hartmann R, Singh PK, Pearce P, Mok R, Song B, Díaz-Pascual F, et al. Emergence of three-dimensional order and structure in growing biofilms. *Nat Phys.* 2019;15: 251–256. doi:10.1038/s41567-018-0356-9
71. Nijjer J, Li C, Zhang Q, Lu H, Zhang S, Yan J. Mechanical forces drive a reorientation cascade leading to biofilm self-patterning. *Nat Commun.* 2021;12: 6632. doi:10.1038/s41467-021-26869-6
72. Valm AM, Welch JLM, Rieken CW, Hasegawa Y, Sogin ML, Oldenbourg R, et al. Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc Natl Acad Sci.* 2011;108: 4152–4157. doi:10.1073/pnas.1101134108
73. Glass DS, Riedel-Kruse IH. A Synthetic Bacterial Cell-Cell Adhesion Toolbox for Programming Multicellular Morphologies and Patterns. *Cell.* 2018;174: 649–658.e16. doi:10.1016/j.cell.2018.06.041

74. Espeso DR, Martínez-García E, de Lorenzo V. Quantitative assessment of morphological traits of planktonic bacterial aggregates. *Water Res.* 2021;188: 116468. doi:10.1016/j.watres.2020.116468
75. van Gestel J, Bareia T, Tenennbaum B, Dal Co A, Guler P, Aframian N, et al. Short-range quorum sensing controls horizontal gene transfer at micron scale in bacterial communities. *Nat Commun.* 2021;12: 2324. doi:10.1038/s41467-021-22649-4
76. Smith WPJ, Davit Y, Osborne JM, Kim W, Foster KR, Pitt-Francis JM. Cell morphology drives spatial patterning in microbial communities. *Proc Natl Acad Sci.* 2017;114: E280–E286. doi:10.1073/pnas.1613007114
77. Estrela S, Brown SP. Community interactions and spatial structure shape selection on antibiotic resistant lineages. *PLOS Comput Biol.* 2018;14: e1006179. doi:10.1371/journal.pcbi.1006179
78. Sousa JAM de, Rocha EPC. Environmental structure drives resistance to phages and antibiotics during phage therapy and to invading lysogens during colonisation. *Sci Rep.* 2019;9: 3149. doi:10.1038/s41598-019-39773-3
79. Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, et al. A standard protocol for describing individual-based and agent-based models. *Ecol Modell.* 2006;198: 115–126. doi:10.1016/j.ecolmodel.2006.04.023
80. Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: A review and first update. *Ecol Modell.* 2010;221: 2760–2768. doi:10.1016/j.ecolmodel.2010.08.019
81. Li B, Taniguchi D, Gedara JP, Gogulancea V, Gonzalez-Cabaleiro R, Chen J, et al. NUFEB: A massively parallel simulator for individual-based modelling of microbial communities. *PLOS Comput Biol.* 2019;15: e1007125. doi:10.1371/journal.pcbi.1007125
82. Jayathilake PG, Gupta P, Li B, Madsen C, Oyebamiji O, González-Cabaleiro R, et al. A mechanistic Individual-based Model of microbial communities. *PLoS One.* 2017;12: e0181965. doi:10.1371/journal.pone.0181965
83. Prats C, Ferrer J, Gras A, Ginovart M. Individual-based modelling and simulation of microbial processes: yeast fermentation and multi-species composting. *Math Comput Model Dyn Syst.* 2010;16: 489–510. doi:10.1080/13873954.2010.481809
84. Grimm V, Frank K, Jeltsch F, Brandl R, Uchmański J, Wissel C. Pattern-oriented modelling in population ecology. *Sci Total Environ.* 1996;183: 151–166. doi:10.1016/0048-9697(95)04966-5
85. Grimm V, Railsback SF. Pattern-oriented modelling: a ‘multi-scope’ for predictive systems ecology. *Philos Trans R Soc B Biol Sci.* 2012;367: 298–310. doi:10.1098/rstb.2011.0180
86. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models. *FEBS J.* 2009;276: 886–902. doi:10.1111/j.1742-4658.2008.06844.x
87. Villaverde AF, Pathirana D, Fröhlich F, Hasenauer J, Banga JR. A protocol for dynamic model calibration. *arXiv Prepr arXiv210512008.* 2021.

88. Myung IJ. Tutorial on maximum likelihood estimation. *J Math Psychol.* 2003;47: 90–100. doi:10.1016/S0022-2496(02)00028-7
89. Törn A, Zilinskas A. Global optimization. *Lect notes Comput Sci.* 1989;350.
90. Kreutz C, Raue A, Kaschek D, Timmer J. Profile likelihood in systems biology. *FEBS J.* 2013;280: 2564–2571. doi:10.1111/febs.12276
91. Beaumont MA. Approximate Bayesian Computation. *Annu Rev Stat Its Appl.* 2019;6: 379–403. doi:10.1146/annurev-statistics-030718-105212
92. Goñi-Moreno A, Amos M. DiSCUS: A Simulation Platform for Conjugation Computing. In: Calude CS, Dinneen MJ, editors. *Unconventional Computation and Natural Computation.* Cham: Springer International Publishing; 2015. pp. 181–191.
93. Pande S, Kaftan F, Lang S, Svatoš A, Germerodt S, Kost C. Privatization of cooperative benefits stabilizes mutualistic cross-feeding interactions in spatially structured environments. *ISME J.* 2016;10: 1413–1423. doi:10.1038/ismej.2015.212
94. Leaman EJ, Geuther BQ, Behkam B. Quantitative Investigation of the Role of Intra-/Intercellular Dynamics in Bacterial Quorum Sensing. *ACS Synth Biol.* 2018;7: 1030–1042. doi:10.1021/acssynbio.7b00406
95. Doumic M, Hecht S, Peurichard D. A purely mechanical model with asymmetric features for early morphogenesis of rod-shaped bacteria micro-colony. *Math Biosci Eng.* 2020;17: 6873–6908. doi:10.3934/mbe.2020356
96. Volfson D, Cookson S, Hasty J, Tsimring LS. Biomechanical ordering of dense cell populations. *Proc Natl Acad Sci.* 2008;105: 15346–15351. doi:10.1073/pnas.0706805105
97. Goñi-Moreno A, Amos M, de la Cruz F. Multicellular Computing Using Conjugation for Wiring. *PLoS One.* 2013;8: e65986. doi:10.1371/journal.pone.0065986
98. van Holthe tot Echten D, Nordemann G, Wehrens M, Tans S, Idema T. Defect dynamics in growing bacterial colonies. *arXiv Prepr.* 2020.
99. Dell’Arciprete D, Blow ML, Brown AT, Farrell FDC, Lintuvuori JS, McVey AF, et al. A growing bacterial colony in two dimensions as an active nematic. *Nat Commun.* 2018;9: 4190. doi:10.1038/s41467-018-06370-3
100. You Z, Pearce DJG, Sengupta A, Giomi L. Geometry and Mechanics of Microdomains in Growing Bacterial Colonies. *Phys Rev X.* 2018;8: 31065. doi:10.1103/PhysRevX.8.031065
101. Zapotocky M, Goldbart PM, Goldenfeld N. Kinetics of phase ordering in uniaxial and biaxial nematic films. *Phys Rev E.* 1995;51: 1216–1235. doi:10.1103/PhysRevE.51.1216
102. Duvernoy M-C, Mora T, Ardré M, Croquette V, Bensimon D, Quilliet C, et al. Asymmetric adhesion of rod-shaped bacteria controls microcolony morphogenesis. *Nat Commun.* 2018;9: 1120. doi:10.1038/s41467-018-03446-y
103. Blanchard AE, Lu T. Bacterial social interactions drive the emergence of differential spatial colony structures. *BMC Syst Biol.* 2015;9: 59. doi:10.1186/s12918-015-0188-5
104. Bottery MJ, Passaris I, Dytham C, Wood AJ, van der Woude MW. Spatial Organization of Expanding Bacterial Colonies Is Affected by Contact-Dependent Growth Inhibition. *Curr Biol.* 2019;29: 3622-3634.e5. doi:10.1016/j.cub.2019.08.074

105. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* 1948;27: 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x
106. Li H, Reynolds JF. A new contagion index to quantify spatial patterns of landscapes. *Landsc Ecol.* 1993;8: 155–162. doi:10.1007/BF00125347
107. Condit R, Pitman N, Leigh EG, Chave J, Terborgh J, Foster RB, et al. Beta-Diversity in Tropical Forest Trees. *Science.* 2002;295: 666–669. doi:10.1126/science.1066854
108. May F, Huth A, Wiegand T. Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proc R Soc B Biol Sci.* 2015;282: 20141657. doi:10.1098/rspb.2014.1657
109. McNally L, Bernardy E, Thomas J, Kalziqi A, Pentz J, Brown SP, et al. Killing by Type VI secretion drives genetic phase separation and correlates with increased cooperation. *Nat Commun.* 2017;8: 14371. doi:10.1038/ncomms14371
110. Nadell CD, Foster KR, Xavier JB. Emergence of Spatial Structure in Cell Groups and the Evolution of Cooperation. *PLOS Comput Biol.* 2010;6: e1000716. doi:10.1371/journal.pcbi.1000716
111. Estrela S, Brown SP. Metabolic and Demographic Feedbacks Shape the Emergent Spatial Structure and Function of Microbial Communities. *PLOS Comput Biol.* 2013;9: e1003398. doi:10.1371/journal.pcbi.1003398
112. Mitri S, Xavier JB, Foster KR. Social evolution in multispecies biofilms. *Proc Natl Acad Sci.* 2011;108: 10839 LP – 10846. doi:10.1073/pnas.1100292108
113. Schluter J, Schoech AP, Foster KR, Mitri S. The Evolution of Quorum Sensing as a Mechanism to Infer Kinship. *PLOS Comput Biol.* 2016;12: e1004848. doi:10.1371/journal.pcbi.1004848
114. Bérubé D, Jébrak M. High precision boundary fractal analysis for shape characterization. *Comput Geosci.* 1999;25: 1059–1071. doi:10.1016/S0098-3004(99)00067-9
115. Kan A, Del Valle I, Rudge T, Federici F, Haseloff J. Intercellular adhesion promotes clonal mixing in growing bacterial populations. *J R Soc Interface.* 2018;15: 20180406. doi:10.1098/rsif.2018.0406
116. Rudge TJ, Federici F, Steiner PJ, Kan A, Haseloff J. Cell Polarity-Driven Instability Generates Self-Organized, Fractal Patterning of Cell Layers. *ACS Synth Biol.* 2013;2: 705–714. doi:10.1021/sb400030p
117. Momeni B, Brileya KA, Fields MW, Shou W. Strong inter-population cooperation leads to partner intermixing in microbial communities. *Elife.* 2013;2: e00230.
118. Goldschmidt F, Regoes RR, Johnson DR. Successive range expansion promotes diversity and accelerates evolution in spatially structured microbial populations. *ISME J.* 2017;11: 2112–2123. doi:10.1038/ismej.2017.76
119. Amor DR, Montañez R, Duran-Nebreda S, Solé R. Spatial dynamics of synthetic microbial mutualists and their parasites. *PLOS Comput Biol.* 2017;13: e1005689. doi:10.1371/journal.pcbi.1005689
120. Shinomoto S, Kuramoto Y. Phase Transitions in Active Rotator Systems. *Prog Theor Phys.* 1986;75: 1105–1110. doi:10.1143/PTP.75.1105

121. Farina A, editor. *Methods in landscape ecology. Principles and Methods in Landscape Ecology: Towards a Science of Landscape*. Dordrecht: Springer Netherlands; 2006. pp. 313–391. doi:10.1007/978-1-4020-5535-5_8
122. Riitters KH, O'Neill R V, Wickham JD, Jones KB. A note on contagion indices for landscape analysis. *Landsc Ecol*. 1996;11: 197–202. doi:10.1007/BF02071810
123. Mony C, Vandenkoornhuyse P, Bohannan BJM, Peay K, Leibold MA. A Landscape of Opportunities for Microbial Ecology Research. *Front Microbiol*. 2020;11: 2964. doi:10.3389/fmicb.2020.561427
124. Railsback SF, Grimm V. *Agent-Based and Individual-Based Modeling: A Practical Introduction*. 2nd ed. Princeton University Press; 2019.
125. Donkin E, Dennis P, Ustalakov A, Warren J, Clare A. Replicating complex agent based models, a formidable task. *Environ Model Softw*. 2017;92: 142–151. doi:10.1016/j.envsoft.2017.01.020
126. Schmolke A, Thorbek P, DeAngelis DL, Grimm V. Ecological models supporting environmental decision making: a strategy for the future. *Trends Ecol Evol*. 2010;25: 479–486. doi:10.1016/j.tree.2010.05.001
127. Banitz T, Johst K, Wick LY, Fetzner I, Harms H, Frank K. The Relevance of Conditional Dispersal for Bacterial Colony Growth and Biodegradation. *Microb Ecol*. 2012;63: 339–347. doi:10.1007/s00248-011-9927-3
128. Werisch M, Berger U, Berendonk TU. Conjugative plasmids enable the maintenance of low cost non-transmissible plasmids. *Plasmid*. 2017;91: 96–104. doi:10.1016/j.plasmid.2017.04.004
129. Koleva KZ, Hellweger FL. From protein damage to cell aging to population fitness in *E. coli*: Insights from a multi-level agent-based model. *Ecol Modell*. 2015;301: 62–71. doi:10.1016/j.ecolmodel.2015.01.024
130. Hellweger FL, Fredrick ND, McCarthy MJ, Gardner WS, Wilhelm SW, Paerl HW. Dynamic, mechanistic, molecular-level modelling of cyanobacteria: *Anabaena* and nitrogen interaction. *Environ Microbiol*. 2016;18: 2721–2731. doi:10.1111/1462-2920.13299
131. Gallagher CA, Chudzinska M, Larsen-Gray A, Pollock CJ, Sells SN, White PJC, et al. From theory to practice in pattern-oriented modelling: identifying and using empirical patterns in predictive models. *Biol Rev*. 2021;96: 1868–1888. doi:10.1111/brv.12729
132. Stenglein JL, Gilbert JH, Wydeven AP, Van Deelen TR. An individual-based model for southern Lake Superior wolves: A tool to explore the effect of human-caused mortality on a landscape of risk. *Ecol Modell*. 2015;302: 13–24. doi:10.1016/j.ecolmodel.2015.01.022
133. Fordham DA, Haythorne S, Brown SC, Buettel JC, Brook BW. poems: R package for simulating species' range dynamics using pattern-oriented validation. *Methods Ecol Evol*. 2021;12: 2364–2371. doi:10.1111/2041-210X.13720
134. van der Vaart E, Beaumont MA, Johnston ASA, Sibly RM. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecol Modell*. 2015;312: 182–190. doi:10.1016/j.ecolmodel.2015.05.020

135. Chazal F, Michel B. An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. *Front Artif Intell.* 2021;4: 108. doi:10.3389/frai.2021.667963
136. Amézquita EJ, Quigley MY, Ophelders T, Munch E, Chitwood DH. The shape of things to come: Topological data analysis and biology, from molecules to organisms. *Dev Dyn.* 2020;249: 816–833. doi:10.1002/dvdy.175
137. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemom Intell Lab Syst.* 1987;2: 37–52. doi:10.1016/0169-7439(87)80084-9
138. Topaz CM, Ziegelmeier L, Halverson T. Topological Data Analysis of Biological Aggregation Models. *PLoS One.* 2015;10: e0126383. doi:10.1371/journal.pone.0126383
139. Couzin ID, Krause J, James R, Ruxton GD, Franks NR. Collective Memory and Spatial Sorting in Animal Groups. *J Theor Biol.* 2002;218: 1–11. doi:10.1006/jtbi.2002.3065
140. Chuang Y, D’Orsogna MR, Marthaler D, Bertozzi AL, Chayes LS. State transitions and the continuum limit for a 2D interacting, self-propelled particle system. *Phys D Nonlinear Phenom.* 2007;232: 33–47. doi:10.1016/j.physd.2007.05.007
141. Huepe C, Aldana M. New tools for characterizing swarming systems: A comparison of minimal models. *Phys A Stat Mech its Appl.* 2008;387: 2809–2822. doi:10.1016/j.physa.2008.01.081
142. Wang S, Fan K, Luo N, Cao Y, Wu F, Zhang C, et al. Massive computational acceleration by using neural networks to emulate mechanism-based biological models. *Nat Commun.* 2019;10: 4354. doi:10.1038/s41467-019-12342-y
143. Oyebamiji OK, Wilkinson DJ, Jayathilake PG, Curtis TP, Rushton SP, Li B, et al. Gaussian process emulation of an individual-based model simulation of microbial communities. *J Comput Sci.* 2017;22: 69–84. doi:10.1016/j.jocs.2017.08.006
144. Oyebamiji OK, Wilkinson DJ, Li B, Jayathilake PG, Zuliani P, Curtis TP. Bayesian emulation and calibration of an individual-based model of microbial communities. *J Comput Sci.* 2019;30: 194–208. doi:10.1016/j.jocs.2018.12.007
145. Wang GG, Shan S. Review of Metamodeling Techniques in Support of Engineering Design Optimization. *J Mech Des.* 2006;129: 370–380. doi:10.1115/1.2429697
146. Salter JM, Williamson D. A comparison of statistical emulation methodologies for multi-wave calibration of environmental models. *Environmetrics.* 2016;27: 507–523. doi:10.1002/env.2405
147. Degeling K, IJzerman MJ, Lavieri MS, Strong M, Koffijberg H. Introduction to metamodeling for reducing computational burden of advanced analyses with health economic models: a structured overview of metamodeling methods in a 6-step application process. *Med Decis Mak.* 2020;40: 348–363. doi:10.1177/0272989X20912233
148. Marie G, Simioni G. Extending the use of ecological models without sacrificing details: a generic and parsimonious meta-modelling approach. *Methods Ecol Evol.* 2014;5: 934–943. doi:10.1111/2041-210X.12250
149. Gorissen D, Couckuyt I, Demeester P, Dhaene T, Crombecq K. A surrogate modeling and adaptive sampling toolbox for computer based design. *J Mach Learn Res.* 2010;11: 2051–2055.

150. Bouhlel MA, Hwang JT, Bartoli N, Lafage R, Morlier J, Martins JRRA. A Python surrogate modeling framework with derivatives. *Adv Eng Softw.* 2019;135: 102662. doi:10.1016/j.advengsoft.2019.03.005
151. Alden K, Cosgrove J, Coles M, Timmis J. Using Emulation to Engineer and Understand Simulations of Biological Systems. *IEEE/ACM Trans Comput Biol Bioinforma.* 2020;17: 302–315. doi:10.1109/TCBB.2018.2843339
152. Pietzsch B, Fiedler S, Mertens KG, Richter M, Scherer C, Widyastuti K, et al. Metamodels for Evaluating, Calibrating and Applying Agent-Based Models: A Review. *J Artif Soc Soc Simul.* 2020;23: 9. doi:10.18564/jasss.4274
153. Lee J-Y, Sadler NC, Egbert RG, Anderton CR, Hofmockel KS, Jansson JK, et al. Deep learning predicts microbial interactions from self-organized spatiotemporal patterns. *Comput Struct Biotechnol J.* 2020;18: 1259–1269. doi:10.1016/j.csbj.2020.05.023
154. Watanabe L, Nguyen T, Zhang M, Zundel Z, Zhang Z, Madsen C, et al. iBioSim 3: A Tool for Model-Based Genetic Circuit Design. *ACS Synth Biol.* 2019;8: 1560–1563. doi:10.1021/acssynbio.8b00078

Chapter 5

A pipeline for calibrating agent-based models of microbial populations: From image collection to model parameterization

The work included in this chapter is published: A. Ahmadi, A. Yip, J. Chalaturnyk, M.G. Aucoin, B.P. Ingalls. A pipeline for calibrating agent-based models of microbial populations: From image collection to model parameterization. *IFAC-PapersOnLine*, 58 (2024), 1-6.

As highlighted in Chapter 4, there are few reports describing the use of methodical, quantitative approaches for calibrating agent-based models to experimental data, especially data involving multiple spatial dimensions. The purpose of this chapter is to develop a *pattern-oriented modelling*-inspired methodology (or pipeline) to calibrate agent-based models of microbial communities to microscopy images of bacteria at single-cell resolution. To demonstrate our methodology in a relatively simple scenario, we sought to capture spatio-temporal patterns that occur in isogenic microbial populations growing on a two-dimensional surface. We applied our methodology to calibrate parameter values governing biophysical interactions in the agent-based model CellModeller.

In Chapter 6, I apply parameter values inferred from this work in simulations of bacteria growing in microfluidic environments, and I apply the same pipeline to calibrate a bacterial conjugation model in CellModeller.

5.1 Abstract

We present a framework for calibration of agent-based models of bacilliform (rod-shaped) bacterial populations based on time-lapse microscopy at single-cell resolution. Our approach draws on pattern-oriented modelling for feature selection, followed by sensitivity analysis and Bayesian inference to arrive at a model calibration that aligns with experimental observations.

We illustrate this pipeline by calibrating a model of microcolony formation against observations of monolayer *E. coli* population growth.

5.2 Introduction

Microbiomes play essential roles in health and medicine, environmental remediation, and agriculture. Ongoing research is aimed at understanding microbial communities and harnessing their biotechnological potential (Lee et al., 2020). The spatial complexity of environments such as the mammalian gut or soil pose challenges for studying the behaviour of microbial communities in their natural habitats. Spatio-temporal mathematical models can be used to support such studies by characterizing spatial heterogeneity. In particular, spatially explicit agent-based models (ABMs) can be used to simulate complex inter-cellular processes. ABMs can also incorporate molecular events, such as gene regulation and metabolic reactions in single cells, adding to their flexibility in describing behaviours in microbial communities. Emergent features at the population- and community-scale can emerge from such processes, reflecting the behaviour of biological systems. Such models have been previously used to reproduce, for instance, coordinated oscillations in gene expression (Matyjaszkiewicz et al., 2017), coordinated cell lysis through quorum-sensing (Scott et al., 2017), and spatial patterns arising from antagonistic interactions between bacterial populations (Smith et al., 2020).

Previously, we surveyed methods for calibrating spatio-temporal models of microbial communities (Yip et al., 2022); we found few reports describing the use of methodical, quantitative approaches for calibrating ABMs to experimental data. Instead, ABMs are frequently developed to capture general qualitative trends, and so are calibrated informally. In contrast, well-calibrated computational models not only capture trends, but can be used to test quantitative hypotheses and can facilitate rapid prototyping and testing of experiments over broad design spaces.

ABMs often incorporate high-level abstractions of agent behaviours that cannot be directly observed and are typically implemented as stochastic processes; this makes their calibration challenging. Within the ecology community, a standard for calibrating ABMs has

emerged which involves a detailed documentation framework (Grimm et al., 2006) accompanied by model calibration and validation through *pattern-oriented modelling* (POM) (Grimm et al., 2005). The POM framework supports quantitative and reproducible calibration of models and assessment of the uncertainty associated with model predictions.

Here, we present a POM-inspired pipeline for calibrating ABMs of bacterial populations to single-cell resolution time-lapse microscopy data. To demonstrate the pipeline's use, we calibrated the biophysical parameters of an ABM against time-lapse observations of *E. coli* forming microcolonies (Figure 3.2). Following the POM framework, we methodically selected patterns against which to calibrate, guided by sensitivity analysis, and then used an iterative Bayesian approach for parameter inference. We demonstrate the pipeline with a relatively simple calibration task in which model biophysics depends on just two parameter values; this scenario allows for transparent illustration of each step of the process. In ongoing work we are applying the pipeline to more complex settings, addressing multiple microbial populations involved in a range of ecological interactions.

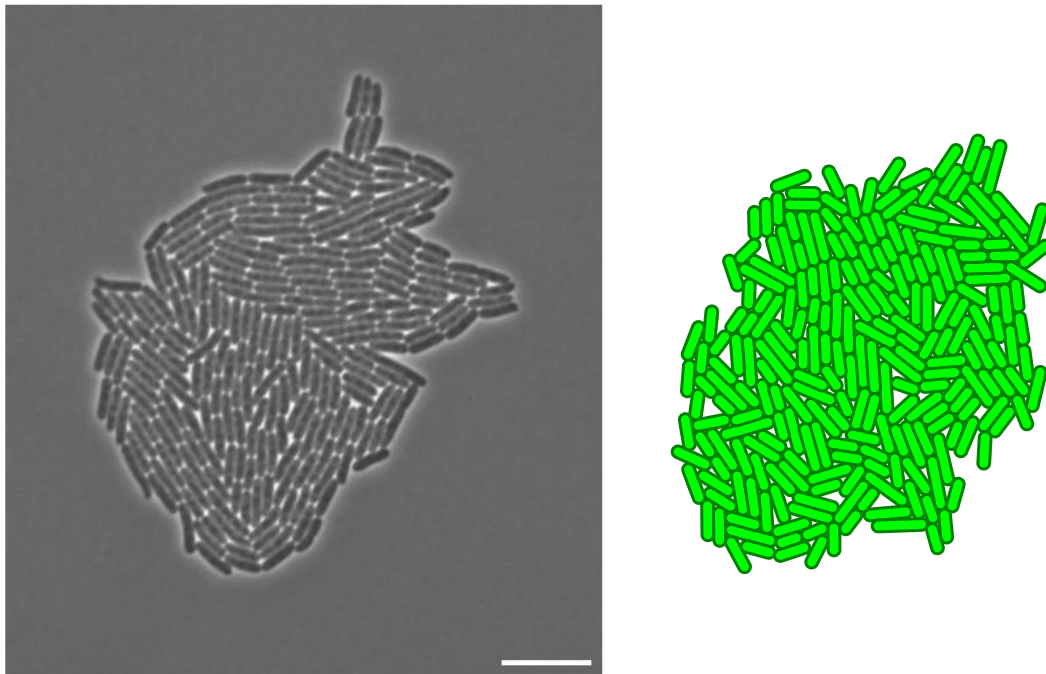


Figure 5.1. Monolayer microcolony of *E. coli* (left) and corresponding agent-based model simulation (right). Scale bar: 10 μm .

5.3 Methods

5.3.1 Microscopy

An overnight culture of *E. coli* MG1655 was diluted 1:100 in 3 mL of LB Miller medium. Cells were grown at 37 °C with shaking until the OD₆₀₀ reached 0.4-0.6. Samples were then prepared for imaging as described in cite (Young et al., 2012). Briefly, cells were diluted to an OD₆₀₀ of 0.01-0.05. Then, 2.25 μL of culture was spotted onto LB pads made with 1.5% low-melt agarose. After 5 minutes, the agarose pads were inverted and sealed in a 50 mm No. 1.5 coverslip dish (Mattek P50G-1.5-30-F). The cells were grown at 37 °C in a GeneFrame incubation chamber (Pecon Incubator P S Compact) secured to the stage of a Zeiss Axio Observer inverted microscope equipped with an Axiocam 506 mono camera (Zeiss). With a 63x oil objective, phase contrast images were taken every 3 minutes at 6-8 different locations using an automated stage controlled by ZEN Blue 2.6 software until cells were no longer in monolayer (from 3.6 to 5 h). Intensity and exposure time were set to 50% and 150 ms, respectively. A total of 24 microcolonies, reaching end-point populations between 140-225 cells, were observed across four biological replicate experiments.

5.3.2 Cell tracking and measurements

Time-lapse images were processed using ImageJ's 'unsharp mask' filter (Schneider et al., 2012). A custom Python script was employed for registration-based jitter correction ([GitHub](#)). Cells were segmented with Omnipose (Cutler et al., 2022); tracking and feature assignment were carried out with CellProfiler (Stirling et al., 2021), following the pipeline outlined in Ahmadi et al. (2024). A custom post-processing tracking correction tool ([Github](#)) was applied to ensure accurate assessment of cell propagation and division events.

Cell morphology and orientation were determined in CellProfiler by fitting an ellipse to each cell object: object length is assigned as the major axis length; orientation is defined by the angle between this major axis and the x-axis of the frame. The cell centroid is assigned as

the centre of the ellipse. Elongation rates were assigned as the slope of the best-fit line to the logarithm of cell length over time; cell radii were likewise assigned as the average radius (minor axis length of fitted ellipse) from birth to division. Division lengths were recorded as the cell length in the frame before division.

5.3.3 Agent-based modelling

Agent-based model simulations were performed with CellModeller (Rudge et al., 2012), in which cells are represented as rigid capsules that elongate exponentially at a rate proportional to their length. Cell movement is determined using an energy minimization algorithm to achieve quasi-stationary mechanical equilibrium, whereby growth-induced cell overlaps are minimized and resolved via an elastic spring model. For further details on the biophysical model derivation and implementation, we refer the reader to the technical documentation on CellModeller (Rudge et al., 2012) and the ODD (Overview, Design, and Details) protocol in Appendix E.

We focused on two parameters that characterize the biophysics engine. The parameter α penalizes displacement of cells in the overlap resolution step, characterizing whether the cells behave as rigid objects (α small) or are flexible enough to allow unresolved overlap (α large). The parameter γ characterizes growth restriction, i.e. the growth penalty induced by neighbouring cells (and any other obstructions). Large values of γ allow cell growth to push through such restrictions. In contrast, when γ is low, growth is limited by the presence of barriers. The default CellModeller settings impose $\alpha = 1/\gamma$, reflecting the relationship between allowable cell overlap ($\propto \alpha$) and cell rigidity ($\propto \gamma$). Here, we separately calibrated α and γ to allow a wider range of behaviours.

Parameters for cell radius (r), maximum growth rate (μ_{max}), and division length threshold (L_{div}) were obtained by fitting probability distributions to experimental measurements ($n = 16385$ cell growth profiles) using SciPy (Virtanen et al., 2020). Growth rates below 0 h^{-1} (indicative of shrinking) and above 2.1 h^{-1} (equivalent to a doubling time of 20 minutes) were removed as outliers. The division length threshold (assigned to newborn

daughter cells) was sampled from a log-normal distribution fitted to experimental measurements, with a minimum sampling threshold of $2\text{ }\mu\text{m}$ to avoid producing cells shorter than $1\text{ }\mu\text{m}$ (such small cells were not observed in the experimental dataset). Maximum growth rates were assigned in a similar manner, sampling from a normal distribution fitted to experimental measurements. All cells were assigned the same radius, set as the experimental mean (consistent with the low variability in observed radii (s.d. 0.07, $n = 16385$)). Model variables and parameters are summarized in Table 5.1.

Table 5.1. CellModeller settings.

Name	Value	Units
r mean	0.67	μm
L_{div} scale parameter	4.80	μm
L_{div} shape parameter	0.42	μm
L_{div} location parameter	-0.32	μm
Division axis noise	$U(-0.001, 0.001)$	Rad
μ_{max} mean	1.15	h^{-1}
μ_{max} s.d.	0.31	h^{-1}
Time step	0.025	h

5.3.4 Summary statistics

The procedures described below were used to assess features for both experimental observations and for simulated microcolony behaviour.

Aspect Ratio

A minimal ellipse was fit to each microcolony. The fitted ellipse contained the centroid of each cell and was determined by the *Löwner-John* approach ([Github](#)), via Welzl's algorithm (Welzl, 2005). The microcolony aspect ratio was defined as the ratio of major axis length to minor axis length.

Convexity

Convexity was defined as the ratio of the perimeter of the convex hull of the microcolony to the perimeter of the microcolony. The boundary of the microcolony was defined using the [AlphaShape](#) package, with the boundary constructed in reference to the end-points of all cells. For each microcolony, the alpha-shape parameter was tuned to arrive at the minimum boundary for which the interior was a connected set. The length of the resulting perimeter was assessed using the Shoelace formula (Meister, 1769). The microcolony convex hull was determined directly from the collection of cell end-points with the Quickhull algorithm (Barber et al., 1996) implemented in SciPy.

Order Parameter

Isotropy among the cells within each microcolony was quantified by a global order parameter (Dell'Arciprete et al., 2018),

$$S = \frac{1}{N} \sum_i^N \cos(2\theta_i) \quad (5.1)$$

where N is the number of cells in the microcolony, and θ_i is the angle between the orientation of cell i and the average orientation of all cells in the microcolony.

Density

Microcolony density was determined from processed data (i.e. cell feature sets). Reconstructed binary pixel location data was used to determine a total area by sketching each microcolony's boundary through contour generation with OpenCV (Bradski et al., 2000) and counting the pixels in the interior using NumPy (Harris et al., 2020). The occupied fraction was taken as microcolony density.

Agreement with exponential growth

Microcolony growth trajectories (sampled every 0.5 h) were fit to an exponential growth profile $\ln N(t) = \ln N_0 + \mu t$, where N is cell number. Agreement was quantified as the resulting R^2 value.

Normalized growth rate

Microcolony growth trajectories (sampled every 0.5 h) were fit to an exponential growth profile. Fitted growth rates were normalized by dividing by the maximum achievable growth rate in the experiments given as $\ln(N_{max})/t_{min}$, where N_{max} is the largest number of cells in all samples (225 cells), and t_{min} is the total time elapsed in the sample with the shortest duration (3.6 h).

5.3.5 Parameter inference

Model parameter values were inferred with the Approximate Bayesian Computation Scheduled Monte Carlo (ABC-SMC) method, implemented in the pyABC package (Klinger et al., 2018), using the algorithmic parameters in Table 5.2. The error tolerance (ε) was defined as the median of distances of the previous population. The initial ε was chosen by obtaining 40 particles by sampling logarithmically from the prior distributions.

Table 5.2. ABC-SMC settings.

Setting	Value
Replicates per parameter set	10
Prior distribution (α)	$U(0.01, 100)$
Prior distribution (γ)	$U(0.1, 1000)$
Accepted simulations per round	40

The statistical distance between experimental and simulated feature distributions was measured using Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951),

$$D_{KL}(P||Q) = \sum \ln \left(\frac{P(x)}{Q(x)} \right) \quad (5.2)$$

where $P(x)$ represents the experimental observed probability density, and $Q(x)$ is the probability density obtained from simulation. D_{KL} was calculated using the method of Perez-Cruz (2008). Ten replicates for the same parameter values were used for generating $Q(x)$. Euclidean distance was used to quantify the total error of all summary statistics, which were equally weighted.

5.4 Results and Discussion

We applied an initial sensitivity analysis to a range of microcolony features drawn from the feature library in Yip et al. (2022) to determine their response to variation in the values of the biophysical parameters α and γ (data not shown). This led us to a collection of six promising summary statistics describing microcolony shape, interior structure, and population growth: aspect ratio, convexity, order parameter, density, agreement with exponential growth from initial state, and population growth rate (details in Section 5.3). We characterized the end-state of each microcolony time-lapse in terms of these six measures (Figure 5.2).

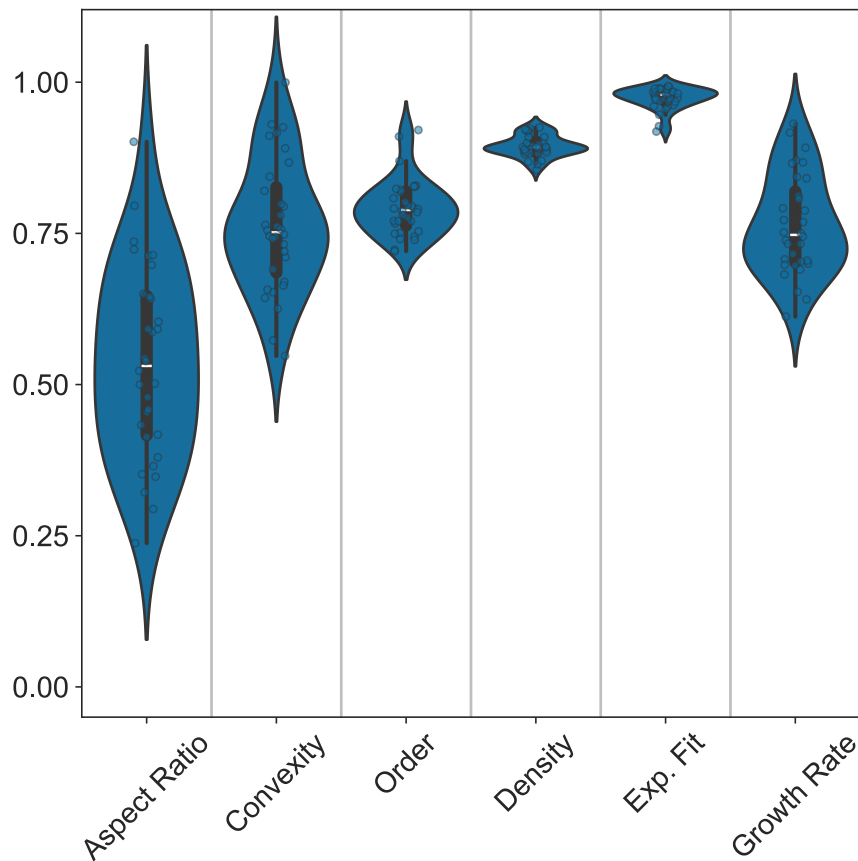


Figure 5.2. Distribution of summary statistics for experimental observations of 24 microcolonies with end-point populations of 140-225 cells. Lower box boundary: 25th percentile; upper box boundary: 75th percentile; line inside box: median; whiskers: 1.5 $\times$ interquartile range.

Microcolony aspect ratio had the highest variance of all summary statistics. From our anecdotal observation, low microcolony aspect ratios occurred when filamentous (overlong) cells appeared in the early stages of microcolony formation. *E. coli* forms filamentous cells in response to environmental stress and DNA damage, which could have occurred during transfer from the liquid culture onto the agarose pad. The division dynamics of these filamentous cells differs substantially from healthy cells (Wehrens et al., 2018), and so the occasional presence of such cells poses a challenge to predictions of colony structure.

Deviations from convexity occurred when branches of cell growth occurred close to the end-point observation. Gaps produced by such branches are invariably filled in by later growth. The transient nature of these patterns could be further explored by a temporal measure of convexity, though we did not investigate that here.

The inner structure of microcolonies tended to be orderly, resulting in high measure of cross-colony isotropy. This finding is consistent with our finding of high density: the microcolonies were tightly packed, which is in agreement with previous observations that cells tend to grow in an optimal, space-filling manner (Volfson et al., 2008).

The population growth rate showed some variability, while aligning closely with exponential growth predictions. This suggests that, at the observed microcolony sizes, the mechanical resistance faced by cells at the interior of the microcolony has little influence on their elongation rate, which is in agreement with previous analyses of growth rates in bacterial microcolonies (Dell’Arciprete et al., 2018).

We next carried out a global multivariate sensitivity analysis to gain a more accurate picture of how informative these six features would be towards model calibration. Parameter sets were chosen by sweeping logarithmically through predetermined ranges of parameter values (Table 5.2). For each parameter set, an ensemble of 50 simulations was produced starting from a single cell, and run until the population reached 200 cells. The results, shown in Figure 5.3, indicate that these summary statistics carry useful information for discriminating the values of α and γ .

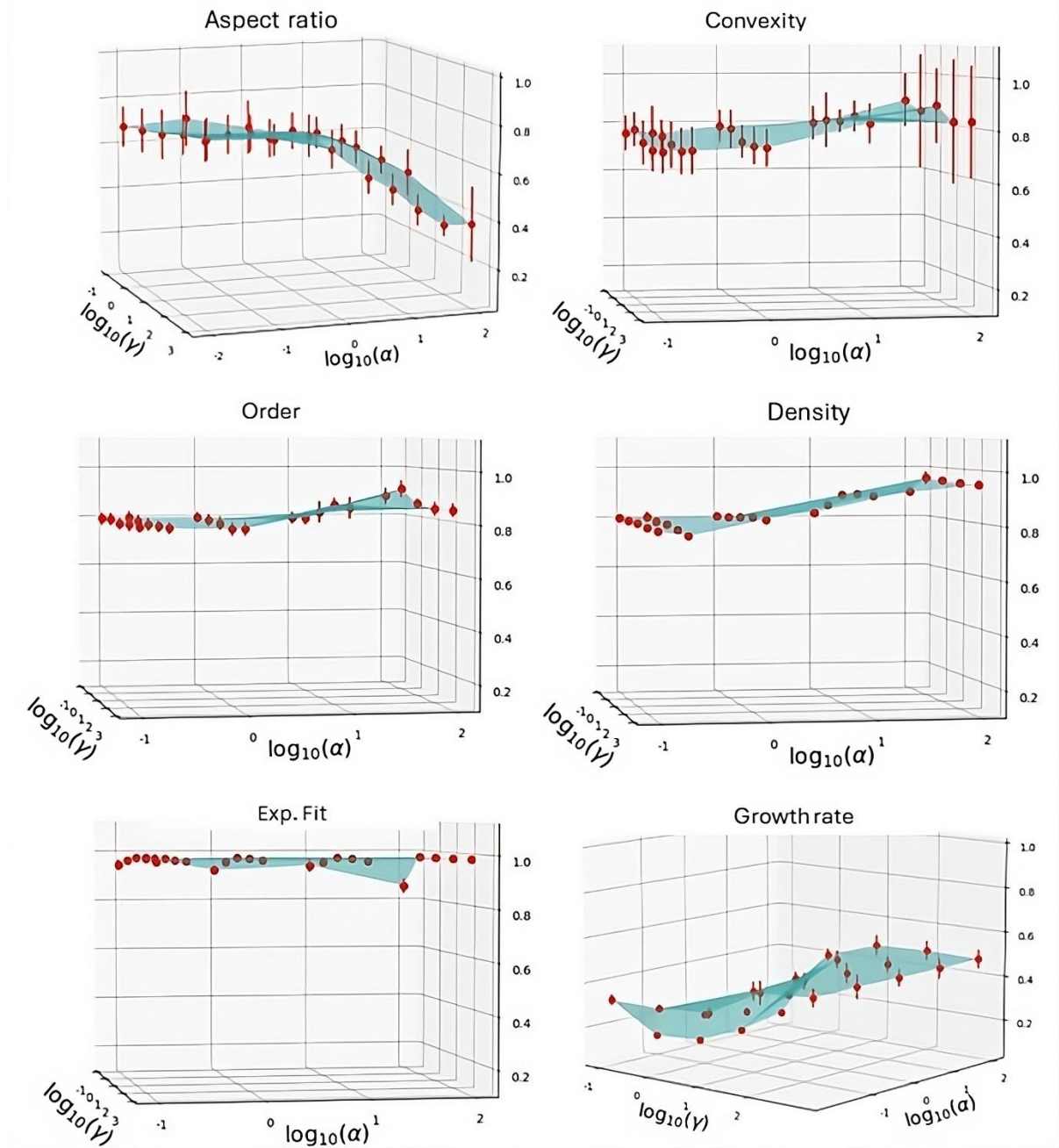


Figure 5.3. Global sensitivity analysis across γ and α values. Red bars represent the standard deviation of each feature value determined from an ensemble of 50 simulations for each parameter pair.

In particular, sensitivity analyses show that microcolony density, convexity, and structural order increase with α , consistent with the role of cellular flexibility in allowing for reorientation and redistribution and thus achieving efficient packing. In contrast, aspect ratio

drops with increasing α , consistent with greater alignment among the population causing the microcolony to elongate in the predominant direction of growth.

The value of γ had limited impact on these geometric microcolony features. Decrease in γ corresponds to a slight increase in aspect ratio, suggesting that constraints on cell growth in the interior of the microcolony may lead to a rounder microcolony shape, with growth in any “narrow” direction less inhibited than growth along the major “crowded” axis.

Higher values of α correlate with increased growth, implying that cell flexibility is beneficial for growth. Likewise, increased values of γ , which corresponds to reduced growth limitation, also correlate with increased growth rates. The deviation from exponential growth shows a small dependence on γ , reflecting the effect of inhibition of growth in the interior of the microcolony. This dependence becomes more pronounced as microcolony size increases.

We carried out parameter inference using the ABC-SMC approach (details in Methods) starting with uniform prior distributions as described in Table 5.2. Six rounds of calibration (timing: 48 hours on an Intel Core i7-7700 CPU @ 3.60 GHz, 32 GB RAM) produced the posterior distribution shown in Figure 5.4 (γ : mean = 128, s.d. = 3; α : mean = 3.5, s.d = 1.2; n = 40). This γ value is consistent with previous use of CellModeller (Kan, 2023; Yáñez Feliú et al., 2020). In contrast, the inferred α value differs from the package default of $1/\gamma$ (Rudge et al., 2012). The posterior distribution for γ has a higher variance than that of α , consistent with the fact that the chosen summary statistics are less sensitive to changes in γ (Figure 5.3).

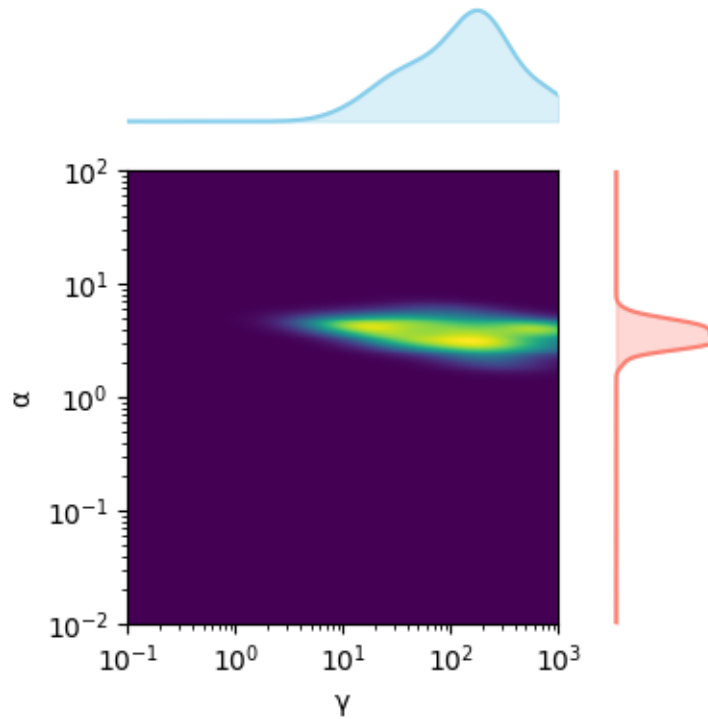


Figure 5.4. Posterior probability densities for calibrated parameter values of γ and α after six rounds of ABC calibration.

We assessed the calibrated model (Figure 5.5) by running 24 simulations with logarithmic sampling from the final posterior distribution in Figure 5.4. For comparison, we ran 24 simulations with logarithmic sampling from the prior distributions (Table 5.2). Predictions of the order parameter, colony density, exponential growth fit, and normalized population growth rate were improved using the calibrated model, i.e. they show improved agreement with the experimental microcolony features, compared to naive sampling from prior distributions (Figure 5.6). However, the calibrated model did not capture the observed median or variance in microcolony aspect ratio, nor the variance in convexity. This result can be partly attributed to the strong dependence of these statistics on α and their insensitivity to γ (Figure 5.3). This singular dependency on α , together with the tight range of α values in the posterior, results in feature distributions with a smaller variance. In addition, these discrepancies may be partly due to the dependence of these features on the rare (and unmodelled) appearance of filamentous cells. We conjecture that summary statistics collected from larger colonies, which

generally take on rounder shapes (Rudge et al., 2012), would reduce the measured variance. We also conjecture that the sensitivity of the orientational order parameter can be increased by including higher rank order parameters, which would contain more information about the orientational distribution of cells (Yankova et al., 2012).

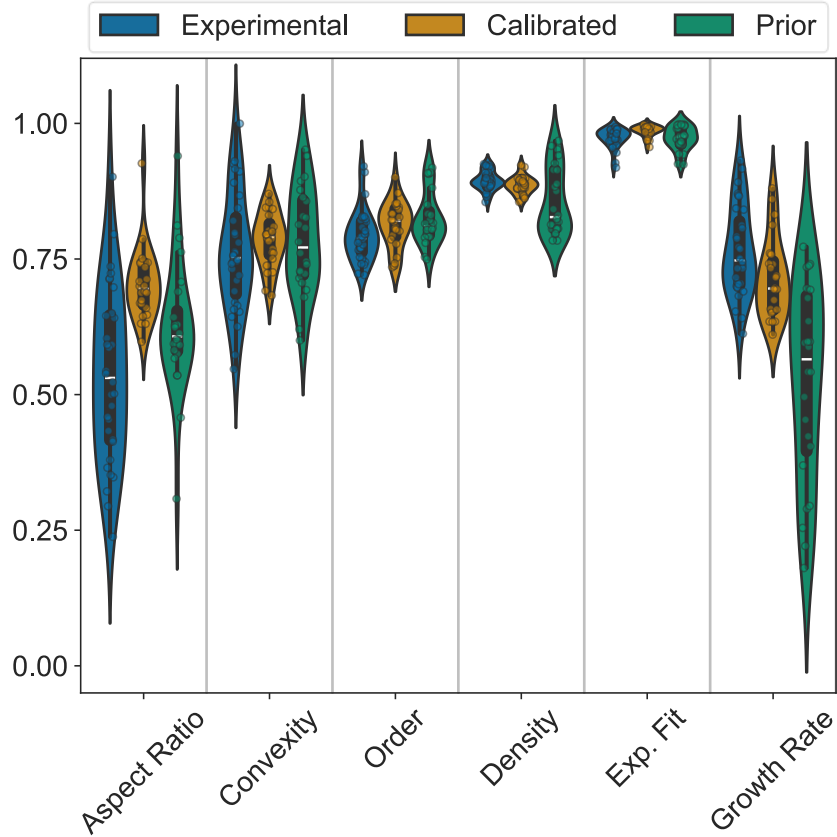


Figure 5.5. Summary statistics for experimental data and simulations sampled from prior distributions and final posterior distributions ($n = 24$). Lower box boundary: 25th percentile; upper box boundary: 75th percentile; line inside box: median; whiskers: $1.5 \times$ interquartile range.

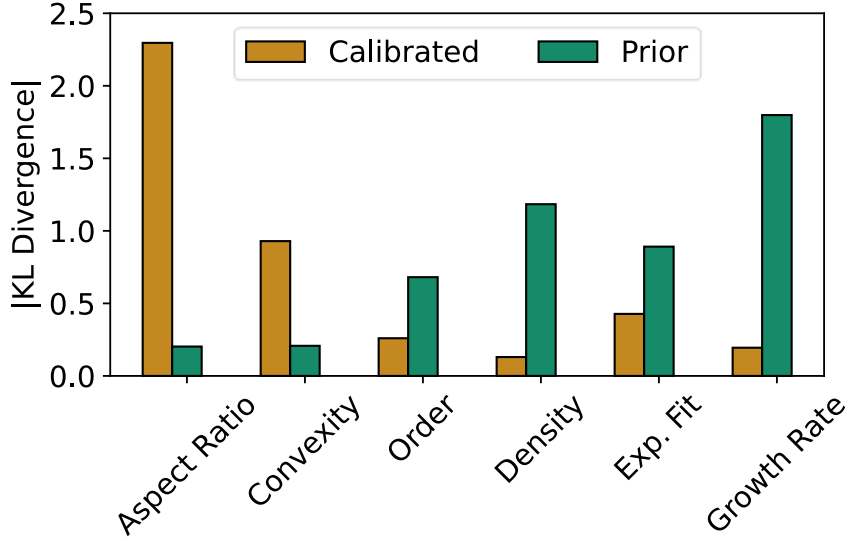


Figure 5.6. KL divergence between experimental data and simulations sampled from prior distributions and final posterior distributions shown in Figure 5.5.

The simple (two-parameter) inference task illustrated here allowed for a comprehensive sensitivity analysis. For higher-dimensional calibration tasks, more efficient approaches to global sensitivity analysis would be called for, with corresponding techniques for interpretation of the results (Hamby, 1994). Likewise, the simple illustration of the calibrated posterior distribution (Figure 5.4) would need to be treated with more sophisticated data visualization techniques in such cases.

5.5 Conclusion

We demonstrated a pipeline to calibrate spatio-temporal ABMs of bacterial microcolony formation against single-cell resolution time-lapse microscopy measurements. The calibration methodology was motivated by the standardized POM framework adopted in the ecology community. Using an ABC-SMC approach, we calibrated biophysical parameters in CellModeller to reproduce our experimental observations, resulting in reasonably tight posterior distributions for the two inferred parameter values. We found that the calibrated model adequately reproduced most features characterizing the shape, interior structure, and growth of *E. coli* microcolonies. However, even for this simple task, we conjecture that

refinement of the feature set, such as incorporation of additional temporal information, could likely lead to improved performance. The calibration pipeline illustrated here lays a foundation for systematic POM-based calibration of more complex microbial population dynamics, with associated features that occur at both single-cell and population-level scales, generalizable to the complex systems and patterns that occur in microbial communities.

5.6 Acknowledgements

This work was supported by funding from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We thank Sheng Fang, Sohaib Nadeem, and Henry Zhang for assisting with preliminary development of models and software. This work was carried out on the Haldimand Tract, the land granted to the Six Nations that includes six miles on each side of the Grand River.

5.7 References

- Ahmadi, A., Courtney, M., Ren, C., Ingalls, B., 2024. A benchmarked comparison of software packages for time-lapse image processing of monolayer bacterial population dynamics. *Microbiology Spectrum* 12, e00032-24. <https://doi.org/10.1128/spectrum.00032-24>
- Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software (TOMS)* 22, 469–483.
- Bradski, G., Kaehler, A., others, 2000. OpenCV. Dr. Dobb's journal of software tools 3.
- Cutler, K.J., Stringer, C., Lo, T.W., Rappez, L., Stroustrup, N., Brook Peterson, S., Wiggins, P.A., Mougous, J.D., 2022. Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nat Methods* 19, 1438–1448. <https://doi.org/10.1038/s41592-022-01639-4>
- Dell'Arciprete, D., Blow, M., Brown, A., Farrell, F., Lintuvuori, J.S., others, 2018. A growing bacterial colony in two dimensions as an active nematic. *Nature Communications* 9, 4190.
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., others, 2006. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* 198, 115–126.
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., others, 2005. Pattern-Oriented Modeling of Agent-Based Complex Systems: Lessons from Ecology. *Science* 310, 987–991.
- Hamby, D.M., 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment* 32, 135–154.
- Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., others, 2020. Array programming with NumPy. *Nature* 585, 357–362.
- Kan, A., 2023. Propagation of engineered cellular properties to colony scales in bacterial populations (PhD Thesis).
- Klinger, E., Rickert, D., Hasenauer, J., 2018. pyABC: distributed, likelihood-free inference. *Bioinformatics* 34, 3591–3593. <https://doi.org/10.1093/bioinformatics/bty361>
- Kullback, S., Leibler, R.A., 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22, 79–86.
- Lee, E.D., Aurand, E.R., Friedman, D.C., EBRC Microbiomes Roadmapping Working Group, 2020. Engineering Microbiomes—Looking Ahead. *ACS Synthetic Biology* 9, 3181–3183.
- Matyjaszkiewicz, A., Fiore, G., Annunziata, F., Grierson, C.S., Savery, N.J., others, 2017. BSim 2.0: An Advanced Agent-Based Cell Simulator. *ACS Synthetic Biology* 6, 1969–1972.
- Meister, A.L.F., 1769. *Generalia de genesi figurarum planarum et inde pendentibus earum affectionibus*.
- Perez-Cruz, F., 2008. Kullback-Leibler divergence estimation of continuous distributions, in: 2008 IEEE International Symposium on Information Theory. pp. 1666–1670.

- Rudge, T.J., Steiner, P.J., Phillips, A., Haseloff, J., 2012. Computational Modeling of Synthetic Microbial Biofilms. *ACS Synth. Biol.* 1, 345–352. <https://doi.org/10.1021/sb300031n>
- Schneider, C.A., Rasband, W.S., Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* 9, 671–675.
- Scott, S.R., Din, M.O., Bittihn, P., Xiong, L., Tsimring, L.S., Hasty, J., 2017. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nature Microbiology* 2, 1–9.
- Smith, W.P.J., Vettiger, A., Winter, J., Ryser, T., Comstock, L.E., others, 2020. The evolution of the type VI secretion system as a disintegration weapon. *PLOS Biology* 18, 1–26.
- Stirling, D.R., Swain-Bowden, M.J., Lucas, A.M., Carpenter, A.E., Cimini, B.A., Goodman, A., 2021. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* 22, 433. <https://doi.org/10.1186/s12859-021-04344-9>
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Volfson, D., Cookson, S., Hasty, J., Tsimring, L.S., 2008. Biomechanical ordering of dense cell populations. *Proceedings of the National Academy of Sciences* 105, 15346–15351.
- Wehrens, M., Ershov, D., Rozendaal, R., Walker, N., Schultz, D., others, 2018. Size laws and division ring dynamics in filamentous *Escherichia coli* cells. *Current Biology* 28, 972–979.
- Welzl, E., 2005. Smallest enclosing disks (balls and ellipsoids), in: *New Results and New Trends in Computer Science: Graz, Austria, June 20–21, 1991 Proceedings*. Springer, pp. 359–370.
- Yáñez Feliú, G., Vidal, G., Muñoz Silva, M., Rudge, T.J., 2020. Novel tunable spatio-temporal patterns from a simple genetic oscillator circuit. *Frontiers in Bioengineering and Biotechnology* 8, 893.
- Yankova, T.S., Bobrovsky, A.Yu., Vorobiev, A.Kh., 2012. Order Parameters $\langle P_2 \rangle$, $\langle P_4 \rangle$, and $\langle P_6 \rangle$ of Aligned Nematic Liquid-Crystalline Polymer As Determined by Numerical Simulation of Electron Paramagnetic Resonance Spectra. *J. Phys. Chem. B* 116, 6010–6016. <https://doi.org/10.1021/jp301170b>
- Yip, A., Smith-Roberge, J., Haghayegh Khorasani, S., Aucoin, M.G., Ingalls, B.P., 2022. Calibrating spatiotemporal models of microbial communities to microscopy data: A review. *PLOS Computational Biology* 18, e1010533. <https://doi.org/10.1371/journal.pcbi.1010533>
- Young, J.W., Locke, J.C., Altinok, A., Rosenfeld, N., Bacarian, T., others, 2012. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature Protocols* 7, 80–88.

Chapter 6

Characterizing bacterial conjugation dynamics in microfluidic environments

Chapter 3 demonstrated that pFAST-PETase-cis could successfully transfer to microbes found in wastewater and that isolated transconjugants could produce functional FAST-PETase to cause PET degradation. The enhancement of PET degradation in modified microbial communities is proportional to the amount of enzyme produced, which is directly related to the size of the transconjugant population. Assuming the genetic vector is functional in all transconjugants, the enhancement in PET degradation capacity will depend on: 1) propagation of the conjugative plasmid through horizontal and vertical gene transfer, and 2) production rates of FAST-PETase. More generally, these two points determine the extent to which any genetic vector will impact the composition and function of a microbial community. Therefore, the potential impact that a conjugative plasmid can have on a microbial community can be partially determined by assessing how efficiently the plasmid spreads. In light of this, being able to predict plasmid spread in realistic environments would certainly be useful for designing effective genetic bioaugmentation interventions.

6.1 Abstract

Conjugative plasmids are a vehicle for spread of genetic information between bacteria. These plasmids can contain auxiliary functions that may be beneficial for the bacterial host, such as catabolic enzymes, resistance to antibiotics and heavy metals, and RNA-guided nucleases. Because conjugation is a contact-based action, spatial arrangement of donors and recipients in structured environments has a considerable effect on plasmid spread. Here, we track the spread of a fluorescently labelled, RP4-based conjugative plasmid through populations growing in microfluidic traps. Using fluorescence time-lapse microscopy, we measured population dynamics and occurrence of individual conjugation events. We found that growth-induced alignment in the direction of the trap exit limited mixing between populations, where the

frequency of conjugation events dropping accordingly. We then utilized single-cell measurements from the experimental data set to inform development of a preliminary conjugation model implemented in the agent-based model CellModeller. This preliminary model formulation demonstrated the need to include a parameter for cell death in the donor and transconjugant populations to reproduce experimental observations. We then refined the model choice and parameter values by comparing simulated and experimental summary metrics in an Approximate Bayesian Computation loop. We concluded the study by using the calibrated model to investigate how spatio-temporal factors influence conjugative plasmid spread. This is the first report of an agent-based model for conjugation that has been systematically calibrated to single-cell, time-lapse microscopy data. The conjugation model for gene transfer in enclosed environments provides a foundation for elucidating mechanisms driving the spread of mobile genetic elements in more complicated environments like the mammalian gut, soil, plastic surfaces, and other porous media.

6.2 Introduction

Bacterial conjugation is a contact-based mechanism of horizontal gene transfer where a plasmid is transferred from one bacterium to another host. This process, and horizontal gene transfer more generally, has shaped the evolution of bacterial genomes throughout time¹⁶². Naturally occurring conjugating plasmids have been observed to carry auxiliary functionalities like virulence, catabolism, resistance to environmental stress, and concerningly, resistance to antibiotics^{85,163}. Unsurprisingly, a sizeable portion of the literature on conjugation focuses on the spread of conjugative plasmids in the context of antibiotic resistance spread amongst microbes and ways to limit that spread^{128,164,165}. On the other hand, conjugation has also been used as a gene delivery tool for applications such as bioremediation¹², enhancing nitrogen fixation¹⁴, and targeted removal of pathogens^{166,167}. The relevance of conjugative plasmids in healthcare, environmental, and agricultural applications provides strong motivation for understanding their spread in microbial communities.

Measuring the conjugation rate of a plasmid is a common first step to assessing how quickly a plasmid can spread through bacterial populations. Traditional microbiological

methods for quantifying conjugation rates rely on endpoint population counts of donors, recipients, and transconjugants to estimate conjugation rates ^{121,122}. One limitation of such methods is distinguishing between plasmid propagation from conjugation and clonal cell growth (i.e. vertical gene transfer). Vertical gene transfer can be estimated by using differential equations to model population dynamics. Such models make assumptions about cell growth states, spatial mixing of populations, and the nature of the conjugation process itself ^{121,122,125}. Assumptions made by these models do not always conform to the conditions of the experiment. For example, the assumption of spatial homogeneity is violated when matings are performed on filters with a low initial cell density, and the assumption of exponential growth is violated when endpoint measurement times are too long ¹⁶⁸. Furthermore, standard mating procedures used to measure conjugation rates are far from representative of real environments. Hotspots for conjugation like biofilms are substantially different compared to mating conditions in suspension or on a filter plate ^{166,169}; ecological and spatial factors governing population dynamics such as nutrient gradients and competition for space are not considered in mathematical models used in standard mating experiments. Notably, conjugation efficiencies measured *in vitro* have been observed to be considerably different than in mouse models of the gut ²⁸, confirming that there are limitations to extrapolating from endpoint measurements to more complex conditions. Additionally, it has been demonstrated that conjugation rates can be temporarily upregulated in newly formed transconjugants ^{94,170}, which is contrary to the assumption of a constant conjugation rate used in most models.

One alternative to de-coupling vertical and horizontal gene transfer with differential equations is to directly observe conjugation events occurring. This has been accomplished by using time-lapse microscopy to observe spread of conjugative plasmids carrying a gene for fluorescent protein expression ^{98,128,129,132,133,171}. This technique has elucidated nuances of the conjugation process itself, such as localization of conjugative plasmids ¹³¹, sex pili ¹⁷², ATPases ¹³², as well as timing of intracellular dynamics involved in plasmid transfer ^{133,171}. Single-cell analysis of donor-recipient pairs in physical contact has provided insights into the role of donor-recipient orientations, cellular growth rate, and transitory de-repression in the spread of conjugative plasmids ⁹⁸. Mating experiments performed in microfluidic traps have

also allowed direct measurement of population dynamics and gene transfer in physically constrained environments ^{128,129}. Microfluidic traps enable high-throughput measurement of conjugation events, and they provide a controlled microenvironment to investigate questions about horizontal gene transfer in microbial communities growing in structured microenvironments.

The type of data produced by time-lapse microscopy provides opportunities for calibrating predictive models of conjugation in spatially heterogeneous environments. Agent-based models (ABMs) are particularly suited for simulating horizontal gene transfer in microbial communities because they capture heterogeneity in populations, space, single-cell properties, and intercellular interactions. The simplest models have considered bacterial conjugation as a random event that occurs with a constant frequency between donor-recipient pairs in physical contact ^{130,143,144,155,156,158,173}. More detailed models have correlated the donor growth rate to conjugation frequency to reproduce observations that conjugative plasmids do not penetrate deep into biofilms ^{170,174}. Regardless of the model chosen, parameters governing the baseline conjugation frequency on a per contact basis are needed. This idea of using time-lapse microscopy data to estimate conjugation frequency at single-cell level was recently introduced by Liu *et al.* ¹³⁰. Building on their work, we argue that developing an ABM for conjugation that is methodically calibrated to single-cell measurements will make progress toward confident characterization of horizontal gene transfer in spatial environments, which will be helpful for applications like precision microbiome editing and mitigating antibiotic-resistance spread.

We present a high-throughput methodology for measuring plasmid spread in microbial communities at single-cell level. The method relies on performing matings in microfluidic traps that constrain growth to two dimensions, with analysis performed using state-of-the-art image processing tools, a custom lineage tracking pipeline to obtain single-cell measurements, and lineage tree traversal algorithms to de-couple horizontal and vertical gene transfer (Figure 3.2). We use this methodology to measure the spread of an RP4-based conjugative plasmid into an isogenic microbial population. We complement our experimental analysis by

developing an ABM with a detailed conjugation model, which we fit to experimental data using our previously developed model calibration pipeline ¹⁷⁵. The ABM enabled us to investigate biophysical questions about plasmid spread in dense microbial populations which could not be practically achieved through experimentation.

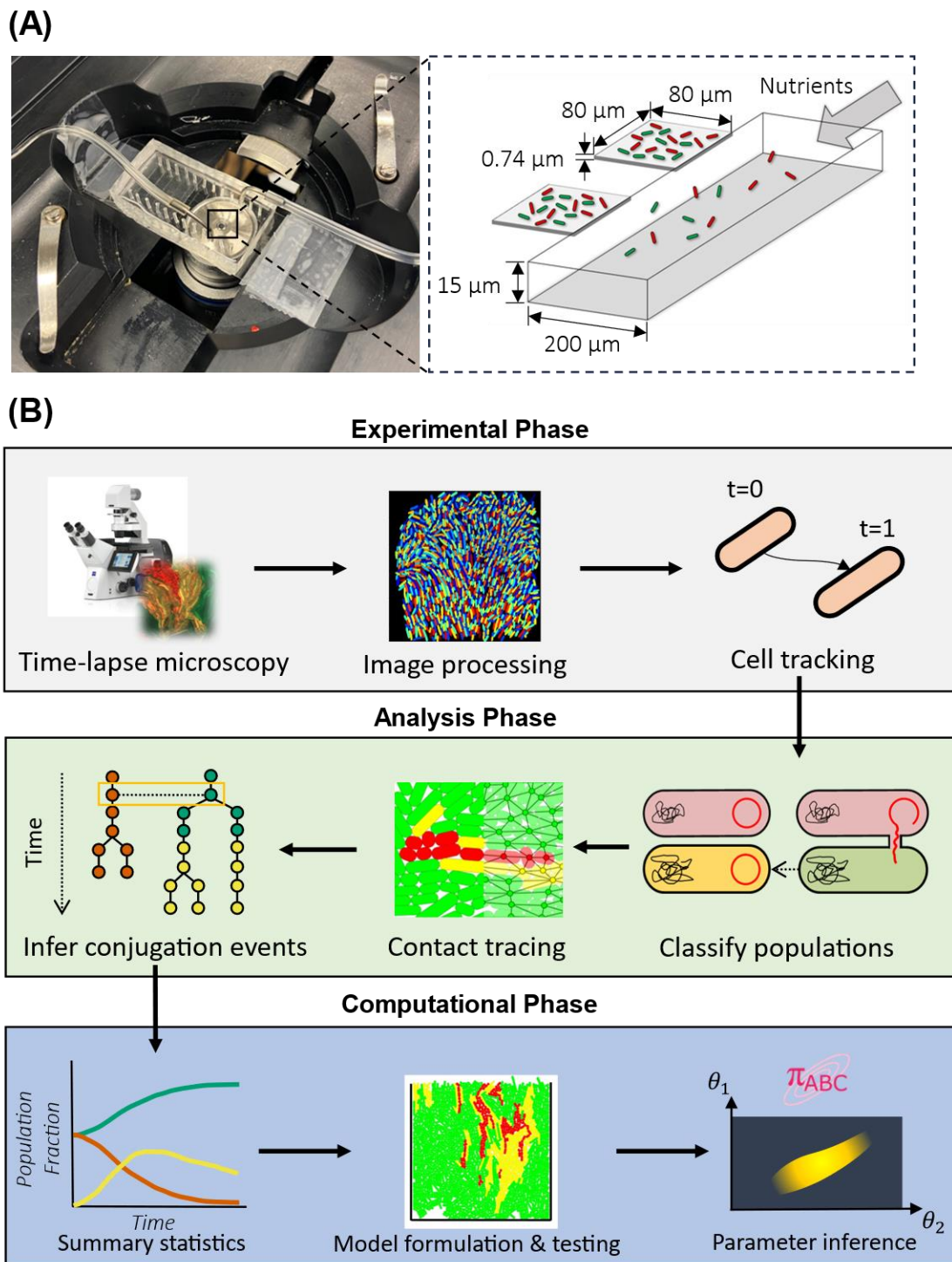


Figure 6.1. Methodology for high-throughput single-cell analysis of plasmid spread in microfluidic devices and agent-based model (ABM) development. (A) Microfluidic chip loaded onto an inverted microscope. Nutrients are fed through the inlet into the main flow

channel. Bacteria grow in a monolayer inside traps on the sides of the main flow channel. (B) Overview of the analysis pipeline. Time-lapse microscopy: fluorescence images are recorded to monitor donor, recipient, and transconjugant populations. Image processing: individual cells are distinguished through a segmentation algorithm. Cell tracking: cell lineages are tracked through time. Classify populations: cells are classified as donors, recipients, or transconjugants based on fluorescence. Contact tracing: all physical contacts between cells are monitored in each image. Infer conjugation events: timing of conjugation events is inferred based on the most recent common ancestor of transconjugants in the same family. Summary statistics: quantitative measures of population dynamics, conjugation efficiency, and intermixing are calculated. Model formulation & testing: an ABM is formulated followed by sensitivity analysis. Parameter inference: Model parameter values are calibrated to experimentally measured summary statistics using an Approximate Bayesian Computation-Scheduled Monte Carlo (ABC-SMC) algorithm built into pyABC ¹⁷⁶.

6.3 Results and Discussion

We performed mating experiments in a microfluidic device with the “hallway” design ¹⁷⁷. Our design consists of several traps ($80 \times 80 \times 0.74 \mu\text{m}$) that have three closed sides and one open side connecting to the main flow channel (Figure 3.2A). We chose this design due to its simplicity and utility in measuring population dynamics in synthetic microbial communities ^{177–180}. The trap also has some resemblance to microenvironments encountered in applications where horizontal gene transfer is relevant, such as crypts in the mammalian gut, microchannels in soil, grooves in rough surfaces, and porous materials.

The mating experiment consisted of a donor and recipient strain of *Escherichia coli*. The donor strain (*E. coli* NEB10- β) carried an RP4-based conjugative plasmid, pFAST-PETase-cis ¹⁸¹, which codes for an mCherry fluorescent protein expressed under a constitutive promoter. The recipient strain (*E. coli* DH5 α) carried a non-mobilizable plasmid coding for expression of a GFP fluorescent marker. We randomly seeded traps with a co-culture containing both populations and took phase contrast, red fluorescence, and green fluorescence images every 5 min for 16 h (Supplementary Video 6.1; details in “Methods”).

6.3.1 Delay between conjugation events and evidence of conjugation

In previous work analyzing intracellular dynamics of the F-plasmid, ssDNA to dsDNA conversion was estimated to be around 10 min, and gene expression levels have been found to

peak approximately 30-90 min after ssDNA to dsDNA conversion¹³³. Thus, we expected a reasonable delay between the occurrence of conjugation events and observation of a functional protein product. Considering that the time for 50% of mCherry proteins to mature in *E. coli* is between 15 and 37 min^{182,183}, we estimated this delay would be between 50-140 min.

Expectedly, we found that there was a substantial delay between occurrence of a conjugation event and visual detection of the fluorescent marker—we detected mCherry expression in recipient cells that had no physical contacts to donors or other transconjugants at the time of detection (Supplementary Video 6.2). We manually confirmed that these cells were in contact with at least one donor prior to detecting the conjugation event. To reduce this delay between detection and occurrence of conjugation, we augmented the brightness of the red fluorescence channel to manually label detection events at the earliest point of detection, which we then used to train a machine learning model in Ilastik¹⁸⁴ that would automatically detect transconjugants in unprocessed images (Supplementary Video 6.2). The trained Ilastik model identified 97.6% of transconjugants and had a false positive rate of 3.6% ($n = 9$ frames; see “Methods”; Appendix B). Based on visual inspection, false positives tended to occur in areas of high background fluorescence and when there was background fluorescence from dust or PDMS outside of the focus plane. The algorithm failed to identify transconjugants when their fluorescence remained consistently below the threshold that the algorithm was trained at (~1000 A.U. for a 14-bit image). We corrected false positives if the red fluorescence intensity of the cell was 95% below the threshold used to classify a cell as a donor.

As expected, division events sometimes occurred during the delay period between conjugation and the appearance of the corresponding fluorescence signal. As a result, we frequently observed near-simultaneous signal detection in related cells. By manually identifying the most recent common ancestor of a family of transconjugants, we obtained an estimate for the maximum delay between occurrence of a conjugation event and detection of the transconjugants resulting from the conjugation event (Figure 6.2; mean: 99 min; s.d.: 36 min; $n = 39$ unique events). After incorporating this mean delay into our cell type classification pipeline to approximate the timing of conjugation events, we processed time-lapse images to

obtain lineage tree data structures for each population by following the remaining steps depicted in the “Experimental Phase” and “Analysis Phase” in Figure 3.2B (details in “Methods”).

We note that the software that we used to fix tracking errors (Trackrefiner, written by Atiyeh Ahmadi) fixed roughly two thirds of tracking errors that it identified (Appendix B), although manual inspection of the tracking performance indicated that some errors (e.g., swapped cell IDs) that were not identified; the number of these unidentified errors are impractical to count and therefore not reported. Expectedly, the majority of tracking errors occurred when the trap was filled with cells and at shallower depths in microfluidic traps (Appendix B), where cells move the fastest. Some of the subsequent analysis involves filtering steps to mitigate the effects of these tracking errors.

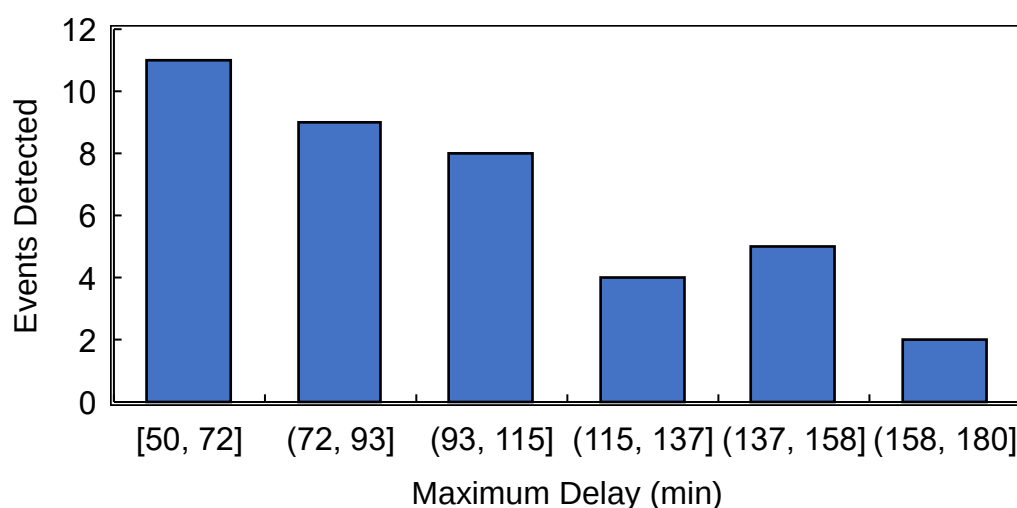


Figure 6.2. Maximum delay between occurrence and detection of conjugation events.

6.3.2 Competition and reduced intermixing limit spread of conjugative plasmids

On average, the microfluidic traps would completely fill with cells between 5-8 h after starting the experiment (Figure 6.3). This timing indicates when cells will begin to leave the trap due

to competition for space from cells deeper within the trap. In the majority of traps, the recipient population tended to outcompete the donor and transconjugant populations (Figure 6.4B). This could be due to the fitness burden that the conjugative plasmid imposes on donors and transconjugants (see Section 1.1.1) and intrinsic differences in relative growth rates of the two different strains. The transconjugant population peaked between 5-10 hours after seeding and then slowly declined for the remainder of the experiment (Figure 6.4B).

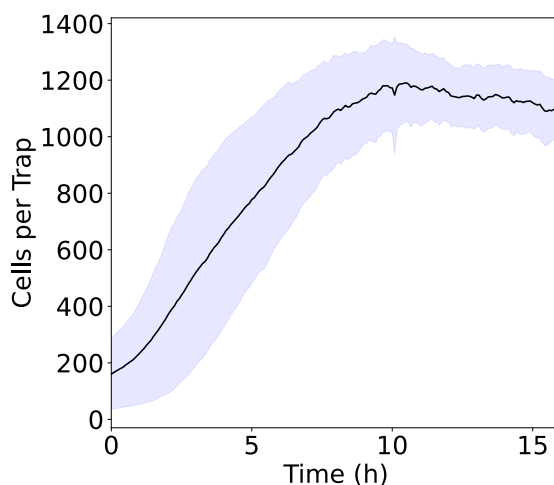


Figure 6.3. Total cells per trap. Solid line represents the mean over $n = 17$ traps; shaded area represents $\pm$ s.d. All traps are from the same experiment.

We quantified intermixing using a normalized entropy score commonly used in quantifying heterogeneity in landscapes (see “Methods”) ¹⁸⁵. Intermixing between populations began to drop after about 5 hours (Figure 6.4C), which is similar to the time it takes for the traps to begin completely filling with cells (Figure 6.3). The frequency of conjugation events also declined after the traps filled with cells (Figure 6.4D). We attribute this fact to the cells adopting a streamlined orientation towards the exit of the microfluidic trap over time (e.g., Figure 6.4A). This transition is apparent when the orientational order parameter—a scalar measure of isotropy within the population which ranges between 0 (complete disorder) and 1 (complete alignment)—is plotted over time (Figure 6.5). This parameter increases rapidly at the same time when the trap fills up 5-8 h after the start of the experiment.

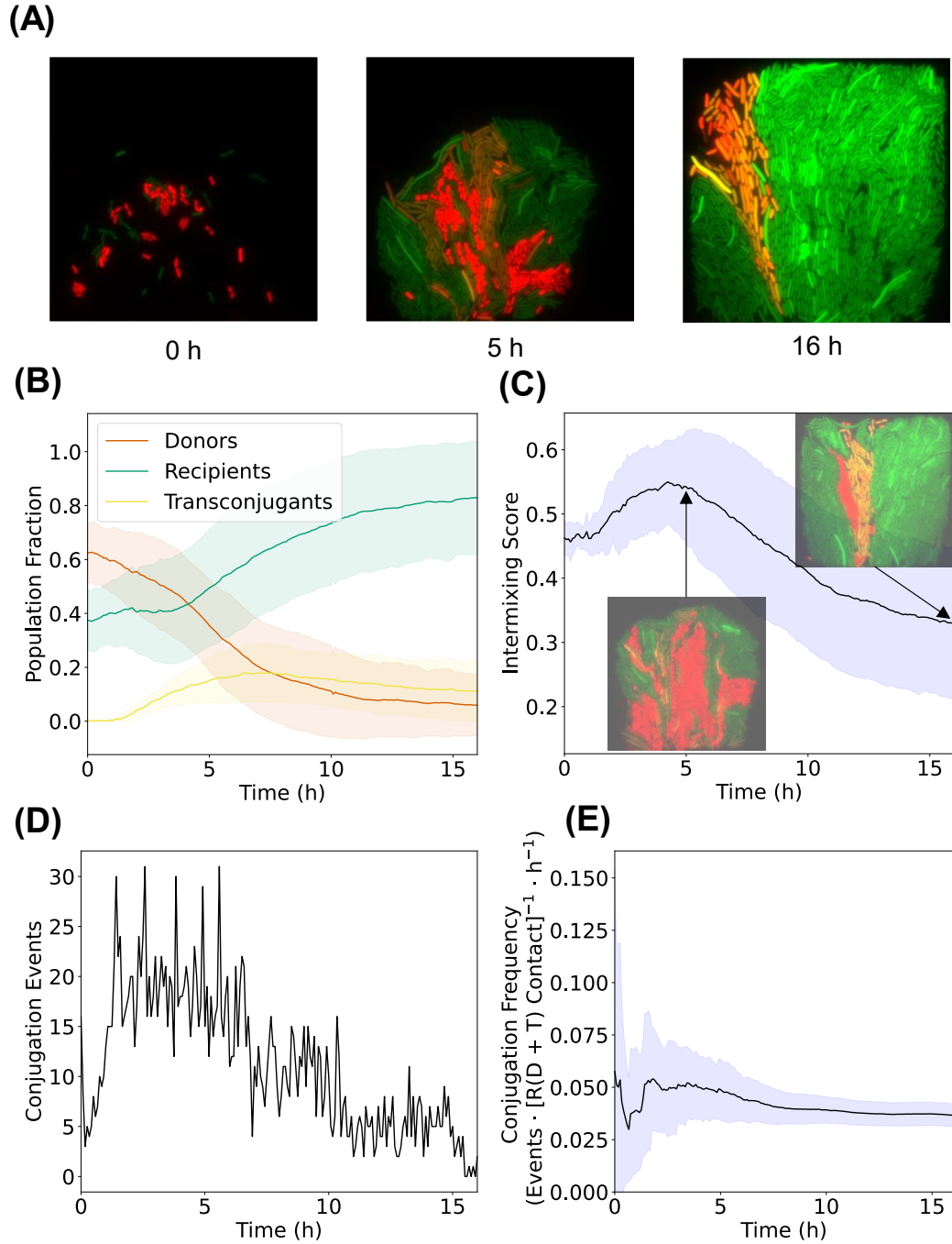


Figure 6.4. Population-level features relevant to conjugation in microfluidic traps. Red, green, and yellow populations represent donors, recipients, and transconjugants, respectively. (A) Images of donors, recipients, and transconjugants growing in a representative microfluidic trap at times 0, 5, and 16 h. (B) Population dynamics for each cell type. (C) Intermixing between

populations quantified by an entropy-based intermixing score. (D) Conjugation events measured in each frame summed over all traps. (E) Conjugation frequency measured as the ratio of the cumulative sum of conjugation events and D-R plus D-T contacts per hour, per trap. Conjugation efficiency: ratio of transconjugants and contacts between D-R plus T-R contacts. Solid lines represent the means over $n = 17$ traps in graphs with shaded areas; shaded areas represent $\pm$ s.d. All traps are from the same experiment.

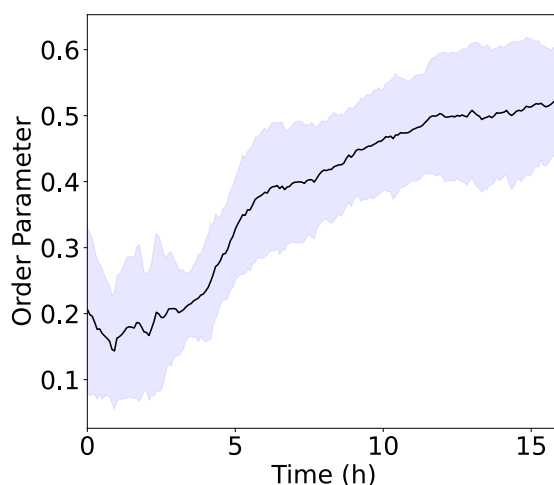


Figure 6.5. Orientation of cells over time measured by a global order parameter. Solid line represents the mean over $n = 17$ traps; shaded area represents $\pm$ s.d. All traps are from the same experiment.

We measured the average conjugation rate as the cumulative number of conjugation events per contact between donors-recipients plus transconjugants-recipients per hour (see “Methods”). Given that this metric is cumulative, we expected this measurement to converge over time. Indeed, we observed this metric converge to a stable value approximately after 8 h (Figure 6.4E). In Sections 6.3.4 and 6.3.5, this single-cell-based conjugation rate will later serve as a useful metric for calibrating parameters in an ABM related to transfer gene activation frequency.

Our observations suggest that cells that do not reach the deepest part of the trap prior to the trap filling get pushed out due to constraints on intermixing in densely growing populations, indicating that there is some priority effect in the hallway trap geometry. These observations about competition for space support the notion that uncontrolled dispersion and

random orientation of cells play a key role in the final relative size of each population. We explore these ideas further in Section 6.3.6.

6.3.3 Fitness burden associated with the conjugative plasmid

In comparison to endpoint measurements, time-lapse single-cell microscopy has the unique ability to provide information about the fate of individual cells. From the time-lapse videos (Supplementary Video 6.1), we observed that cells carrying the conjugative plasmid would eventually stop growing or die at a noticeable rate. To account for noise in elongation rate measurements, we used the window-averaged elongation rate (window = 20 min) to classify the cell growth state. We omitted analyzing any cell cycles containing large spikes in instantaneous elongation rates ($10 < \Delta L / \Delta t < -4 \mu\text{m/h}$) and any cells who left the trap (defined as having the cell centroid at $y > 75 \mu\text{m}$ above the base of the trap at any point in the cell cycle). We calculated the instantaneous death rate of each population by monitoring when individual cells ceased to grow. Based on the observed bimodality in cell elongation rates (Figure 6.6), we classified recipients and transconjugants whose minimum elongation rate reached below $1 \mu\text{m/h}$ as dead; the threshold was set to $0.5 \mu\text{m/h}$ for the donors due to their shorter division length (Figure 6.7). The results are shown in Figure 6.8 and Table 6.1. The instantaneous death rates were steady after the initial phase of the experiment (Figure 6.8). We speculate that the initial spike in measured cell deaths could be due to the stress imparted on cells from the handling steps required to load cells into microfluidic traps. Expectedly, the mean recipient death rate was lower than the death rate of the donors (Table 6.1). We noticed that the instantaneous death rate for transconjugants was noisy in the first 4 h of the experiment, suggesting that the transconjugants were dying shortly after acquiring the conjugative plasmid.

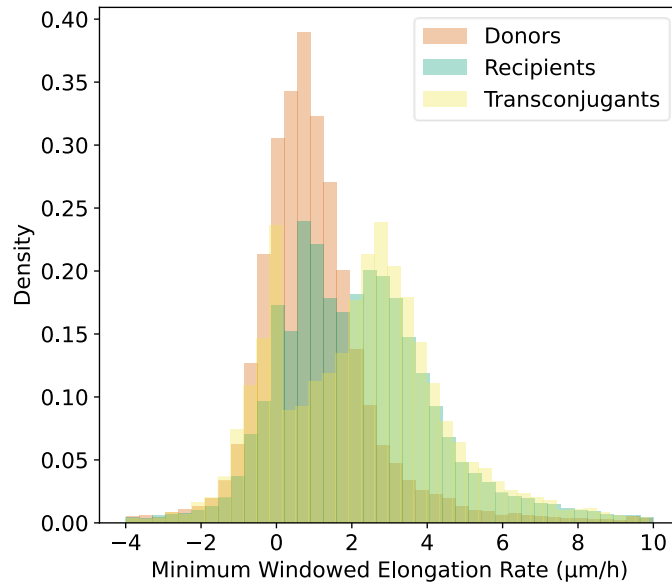


Figure 6.6. Minimum windowed-averaged elongation rate recorded for each population in $n = 17$ traps over 16 h (donors: 24127; recipients: 55475; transconjugants: 9305).

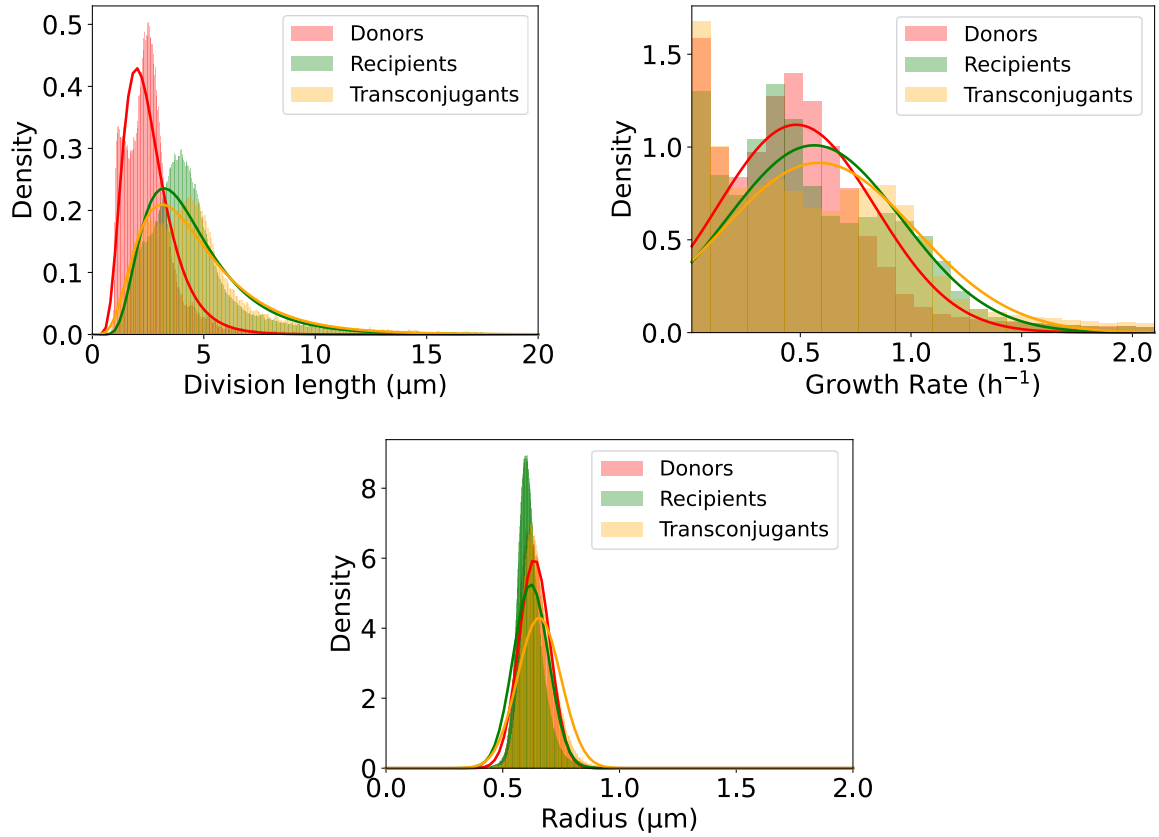


Figure 6.7. Probability density for single-cell growth rate, radius, and division length. Solid lines indicate probability density function fits for a normal distribution (growth rate, radius) or lognormal distribution (division length).

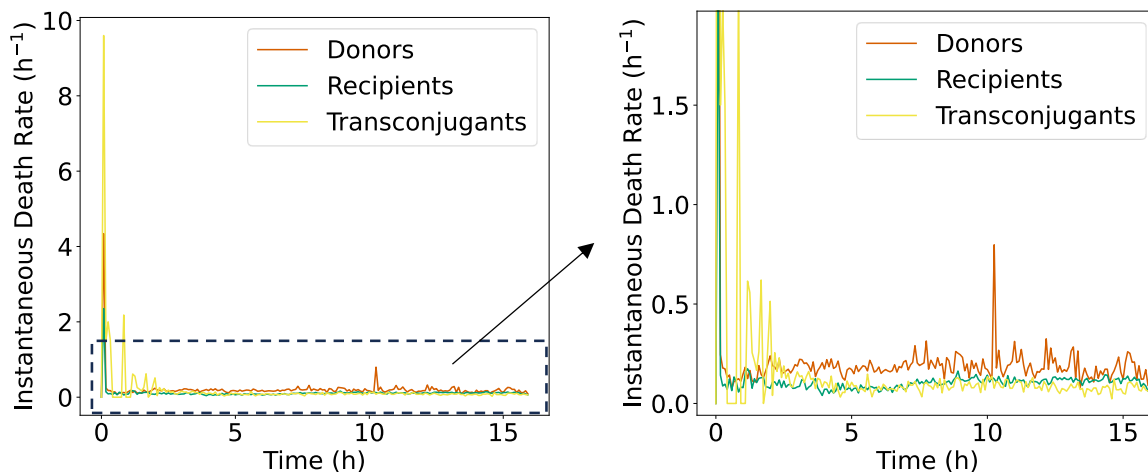


Figure 6.8. Instantaneous death rates for each population over all traps. The boxed area on the left panel is shown in higher detail in the right panel.

Table 6.1. Death rates for each population averaged over 16 h (192 time steps) across 17 traps.

Cell Type	Death Rate	
	Mean	S.d.
Donors	0.203	0.307
Recipients	0.114	0.164
Transconjugants	0.185	0.731

We investigated transconjugant death rates further by analyzing changes in elongation rates of cells relative to the time after receiving the conjugative plasmid. We only considered conjugation events that occurred within the first 4 h of the experiment to reduce chances of a conjugation event being considered due to tracking errors. We confirmed that a substantial portion of transconjugants stopped growing after receiving the conjugative plasmid (total transconjugants analyzed: 140; transconjugants that stopped growing: 51; Figure 6.9). The majority of transconjugants that stopped growing reached their minimum elongation rate between 1-2 hours after receiving the conjugative plasmid. This suggests that there may be a transient metabolic burden after receiving the plasmid.

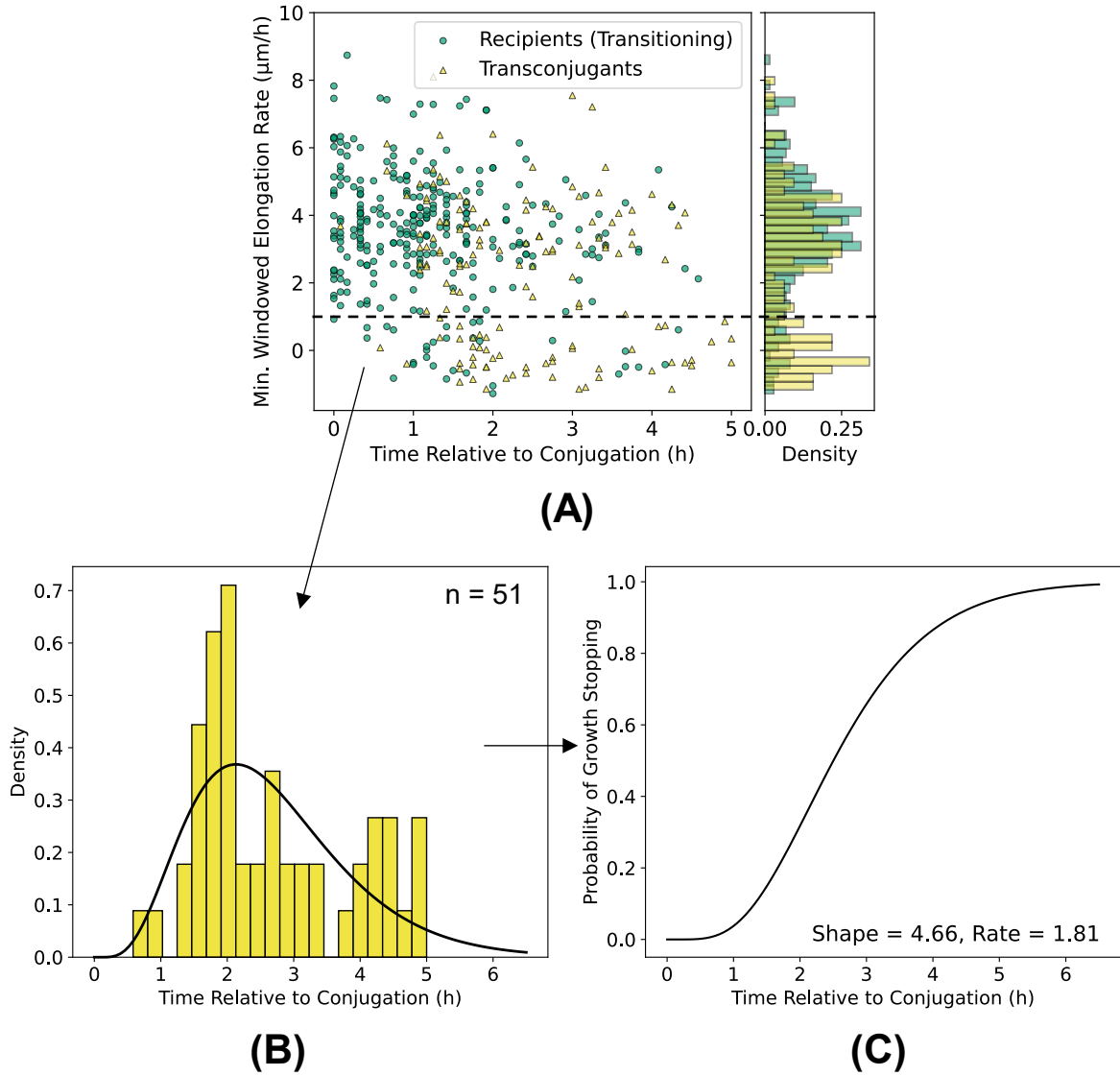


Figure 6.9. Analysis of death rates in transconjugants arising from conjugation events that occurred in the first 4 h of the experiment. (A) Minimum window-averaged elongation rate of cells involved in lineages receiving the conjugative plasmid. Transconjugant points indicate that the cell has been identified as a transconjugant by the image processing algorithm. (B) Histogram for transconjugant cells with windowed-averaged elongation rates below 1 $\mu\text{m/h}$ with a gamma distribution fit. (C) Cumulative distribution of the gamma distribution with shape and rate parameters.

Transconjugant cells dying after receiving the conjugative plasmid could be due to several reasons. One possibility is the strong P_{bs} promoter on the conjugative plasmid driving high expression levels of mCherry, a fluorophore previously shown to have higher cytotoxic effects in *E. coli* relative to EGFP¹⁸⁶. Another reason could be due to transitory de-repression of transfer genes after recircularization of the plasmid, which can be lethal effects when overexpressed⁹⁵. While this observation of cell death might be specific to this plasmid construct, it gives an example of how our method can reveal additional factors that influence plasmid spread that are not observable with endpoint methods.

6.3.4 Formulating an agent-based model for bacterial conjugation

We used the ABM CellModeller¹⁴¹ to simulate conjugative plasmid spread in microbial populations growing in microfluidic traps. We ignored the effect that nutrient gradients within the chamber have on cell growth; previous studies have demonstrated *in vitro* and *in silico* that the nutrient gradients in microfluidic traps with similar geometries and operating conditions to our experiments are rather negligible^{187,188}. Our ABM simulates growth and division characteristics of each population using probability density functions derived from single-cell analysis of our experiment (Figure 6.7; Table 6.2), which enables a direct comparison between our simulations and experiment.

Previous studies used the simplest formulation of a conjugation model, where the donors have a constant transfer gene activation frequency (f_{conj})^{130,143,144,155,156,158,173}. When transfer genes are activated in a donor, that donor selects a random neighbour in physical contact, and if the cell is a “recipient” type, then the recipient becomes a transconjugant that can immediately serve as a functional donor. We will denote this as model variant 0 (M0). This model formulation also assumes that the conjugation process is instantaneous and that cells must be in direct physical contact for conjugation to occur (see “Methods”). We performed a parametric sweep over f_{conj} to assess whether this simple model could capture experimentally observed trends. For each parameter value, we performed simulations where cell states were initialized to match the initial conditions from the experiment ($n = 17$ traps), which would ensure a quantitative comparison between simulated and experimental summary statistics. We

selected three summary statistics to investigate: conjugation rate (conjugation events per recipient contact with donors and transconjugants per hour), intermixing score, and the transconjugant fraction (in the recipient population, given as $T/(R+T)$). These three summary statistics capture the key features of the experiment: horizontal gene transfer, population growth, and spatial mixing between populations.

Table 6.2. Parameters for probability distributions fit to single-cell properties.

Cell Type	n	Property	Distribution	μ	σ	Shift
Donor	387442	Growth rate	Normal	0.495	0.375	-
Donor	387442	Division length	Lognormal	2.406	0.426	-0.042
Donor	387442	Radius	Normal	0.635	0.067	-
Recipient	1530262	Growth rate	Normal	0.563	0.401	-
Recipient	1530262	Division length	Lognormal	3.816	0.504	0.241
Recipient	1530262	Radius	Normal	0.620	0.077	-
Transconjugant	292525	Growth rate	Normal	0.581	0.413	-
Transconjugant	292525	Division length	Lognormal	4.653	0.470	-0.224
Transconjugant	292525	Radius	Normal	0.656	0.093	-

The conjugation rate summary statistic measured in the simulations was within the experimental range when f_{conj} was 0.2 h^{-1} (Figure 6.10A). Interestingly, the conjugation rate was consistently proportional to approximately 1/5 of f_{conj} across each value tested, indicating that each time transfer genes were activated, there was only a 1/5 chance of conjugating to an available recipient. This makes sense since the plasmid donors choose a neighbour at random when activating conjugation genes, so the chances of a successful conjugation event decrease with the number of neighbouring recipient cells. The factor of 1/5 is then indicative of the average number of contacts between plasmid donors and recipients in the microfluidic trap. At

$f_{conj} = 0.2 \text{ h}^{-1}$, the predicted intermixing score and endpoint transconjugant fraction were over-estimated (Figure 6.10B, C). Upon inspection of all population fractions, we found that this over-estimate was caused by the donor population persisting in simulations, whereas in the experiments the donor population would steadily decline (Figure 6.10D). This indicated that the model needed additional parameters to control the growth of donors and transconjugants.

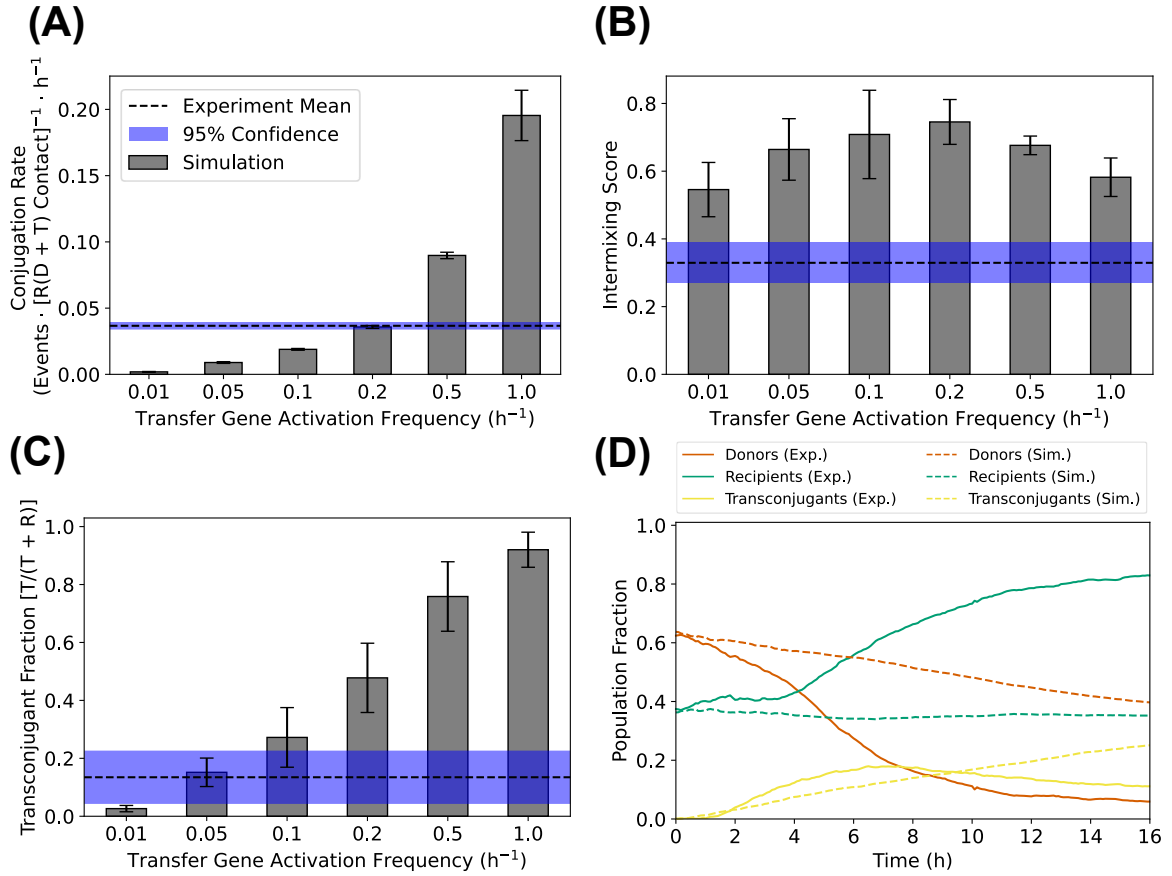


Figure 6.10. Parametric sweep over the transfer gene activation frequency for the basic conjugation model M0. Bar plots show simulated endpoint measurements for: **(A)** Conjugation rate, **(B)** intermixing score, and **(C)** Transconjugant fraction $[T/(T+R)]$. **(D)** Population fractions for the experiment (Exp.; solid lines) and simulations (Sim.; dashed lines) for $f_{conj} = 0.2 \text{ h}^{-1}$. Error bars represent 95% confidence interval for $n = 17$ simulations with initial cell states set to initial conditions from the experiment.

We formulated a new model variant (M1) by implementing a death rate (f_{death}) for donors and transconjugants. For simplicity, we modelled probability of death using an exponential distribution, where cells have a constant chance at dying at any time step. We ran

a parametric sweep over the parameter f_{death} , keeping f_{conj} constant at a value of 0.2 h^{-1} . When f_{death} was equal to 0.2 deaths per cell per hour, all endpoint summary statistics tested were within experimental uncertainty (Figure 6.11A-C). The average population dynamics predicted by M1 also closely resembled the experimental average (Figure 6.11D). This suggests that the lethality associated with the conjugative plasmid has a substantial impact on the fate of the donor and transconjugant populations. The simulated frequency of 0.2 deaths per cell per hour is in agreement with the average death rate measured for the donors (Table 6.1). A more refined probability distribution function for assigning growth rates might improve estimates for the amount of slow- or non-growing donors and transconjugants present in the trap, which would in turn reduce the death frequency parameter. The substantial differences in the results from M0 and M1 indicate that the metabolic burden of the conjugative plasmid is an important factor determining propagation of the conjugative plasmid.

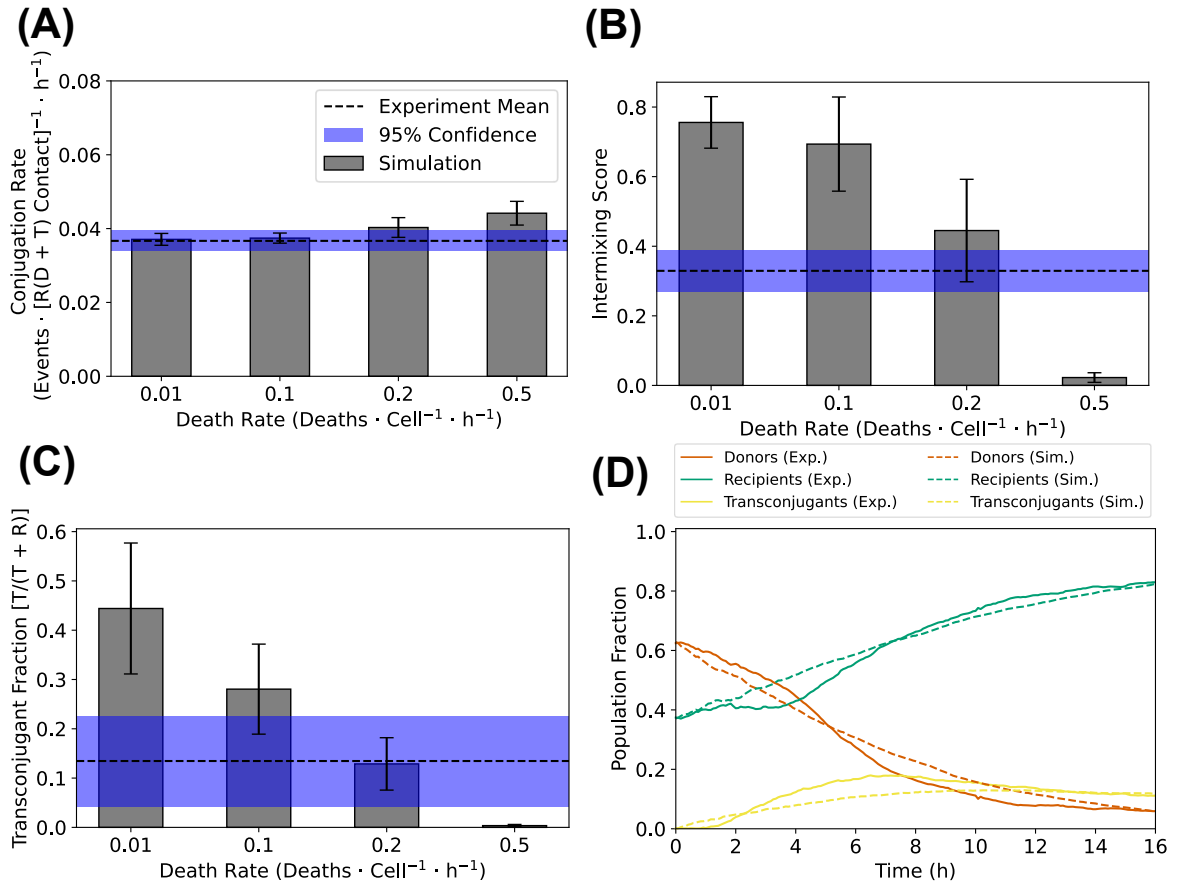


Figure 6.11. Parametric sweep over the death rate parameter for the basic conjugation model with an additional parameter governing death rates of both donors and transconjugants (M1). Bar plots show simulated endpoint measurements for: **(A)** Conjugation rate, **(B)** intermixing score, and **(C)** Transconjugant fraction [T/(T+R)]. **(D)** Population fractions for the experiment (Exp.; solid lines) and simulations (Sim.; dashed lines) for $f_{death} = 0.2$ deaths $\cdot$ cell $^{-1} \cdot$ h $^{-1}$. f_{conj} was held constant at 0.2 h $^{-1}$. Error bars represent 95% confidence interval for $n = 17$ simulations with initial cell states set to initial conditions from the experiment.

6.3.5 Calibration of ABM reproduces plasmid conjugation dynamics

Although M1 made reasonable predictions of experimental summary statistics, this relatively simple model formulation can be further refined to include more mechanisms involved in the conjugation process. We formulated a more complex model (M2, see “Methods”) which included sub-models to account for the following observations from the literature and our experiment:

1. When transfer genes are expressed, a sex-pilus is extended in a random direction within approximately 1 μm of the donor ¹⁷². If a mating-pair with a recipient is formed, then conjugation proceeds. Success of conjugation is therefore proportional to the overlapping area between the donor's pilus range and recipient.
2. There is a delay between a recipient receiving the plasmid and being able to donate the plasmid again (transconjugant maturation time) ^{133,189}.
3. Transconjugants experience temporary, initial upregulation of transfer genes after the transconjugant maturation time (i.e., transitory derepression) ^{98,190}. This temporary upregulation in transfer genes decreases at approximately an exponential rate ¹⁹⁰.
4. Donor and transconjugant cells could die at separate rates (this experiment). We assumed that these death rates were constant.

The lack of information about which donor was responsible for a conjugation event plus the delay between occurrence and observation of conjugation events made it prohibitive to measure transconjugant maturation time and transitory derepression. Hence, the goal here is to simultaneously infer ranges for these parameters and demonstrate use of the model calibration pipeline presented in Chapter 5 for calibrating a more complex model. We used the Approximate Bayesian Computation Scheduled Monte Carlo (ABC-SMC) method in pyABC ¹⁷⁶ to calibrate parameters that characterize the conjugation process and single-cell behaviours starting from uninformative prior distributions (see “Methods”). We calibrated simulated values of the conjugation rate, intermixing score, and the transconjugant fraction in the recipient population after 8 h of simulated time to the same summary statistics in the experiment at this time point. This time point was chosen instead of the actual experiment endpoint to reduce computation time. The selected summary statistics appear to stabilize in the experiment after 8 h (i.e., approximately when the trap becomes full), so we deemed this a reasonable time point to calibrate against. After three rounds of calibration, we obtained posterior distributions for the conjugation frequency and death rates that were within similar ranges to M1 (Figure 6.12).

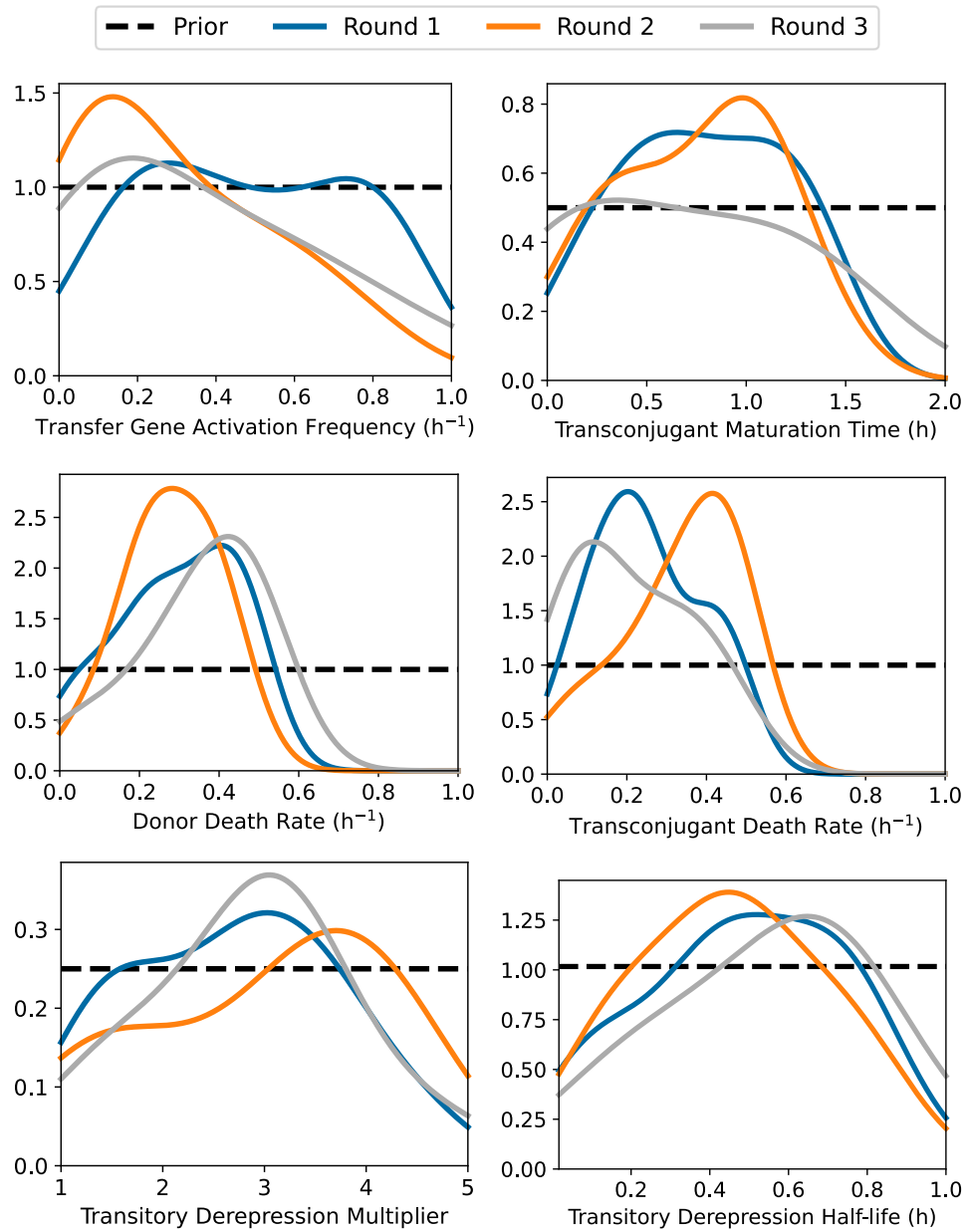


Figure 6.12. Kernel density estimates for ABM parameters after each calibration round from the ABC-SMC algorithm.

To verify the model predictions, we replicated the microfluidics experiment *in silico* with M2 using the mean parameter values obtained from the ABC-SMC calibration (Table 6.3). This model adequately reproduced observed trends and endpoints in relevant experimental summary statistics (Figure 6.13). Despite overestimating the conjugation rate

(Figure 6.13A), the model predicted similar transconjugant fractions to the experiment (Figure 6.13C). We speculate that this from calibrating the model to two out of three summary statistics that depend on the transconjugant population, which gives more weight to parameters that predict populations (e.g., death rates) rather than parameters that predict frequency of conjugation events. The predicted transconjugant fraction appears to be overestimated in the first two hours compared to the experiment, but when the delay between recipients receiving the conjugative plasmid and expressing mCherry is accounted for in the experiment, this initial “overestimate” disappears (Figure 6.13C). In that case, the model underpredicts the transconjugant fraction, which could be caused by overestimating the recipient population, which did not have a death rate assigned to it. Assigning a death rate to the recipient population in the model would likely improve predictions during the initial transient period. The predicted intermixing score was overestimated by the simulations, particularly in the first 4-6 h. Nonetheless, the predicted intermixing score followed a similar trend to the experiment and was in agreement with the experimental values near the endpoint (Figure 6.13B). The donor death rate was also slightly overestimated compared to the experimental value, which is apparent because the donor population fraction dropped faster than the experimental value in the first 4-6 hours (Figure 6.13D).

Table 6.3. Statistics for third round posterior distributions from ABC-SMC calibration.

Name	Symbol	Mean	Coefficient of Variation	Units
Transfer gene activation frequency	f_{conj}	0.354	0.720	μm
Transconjugant maturation time	t_m	0.678	0.611	h
Donor death rate	$f_{death,D}$	0.351	0.336	h^{-1}
Transconjugant death rate	$f_{death,T}$	0.213	0.535	h^{-1}
Transitory derepression multiplier	m	2.866	0.384	-
Transitory derepression half-life	$t_{1/2}$	0.545	0.404	h

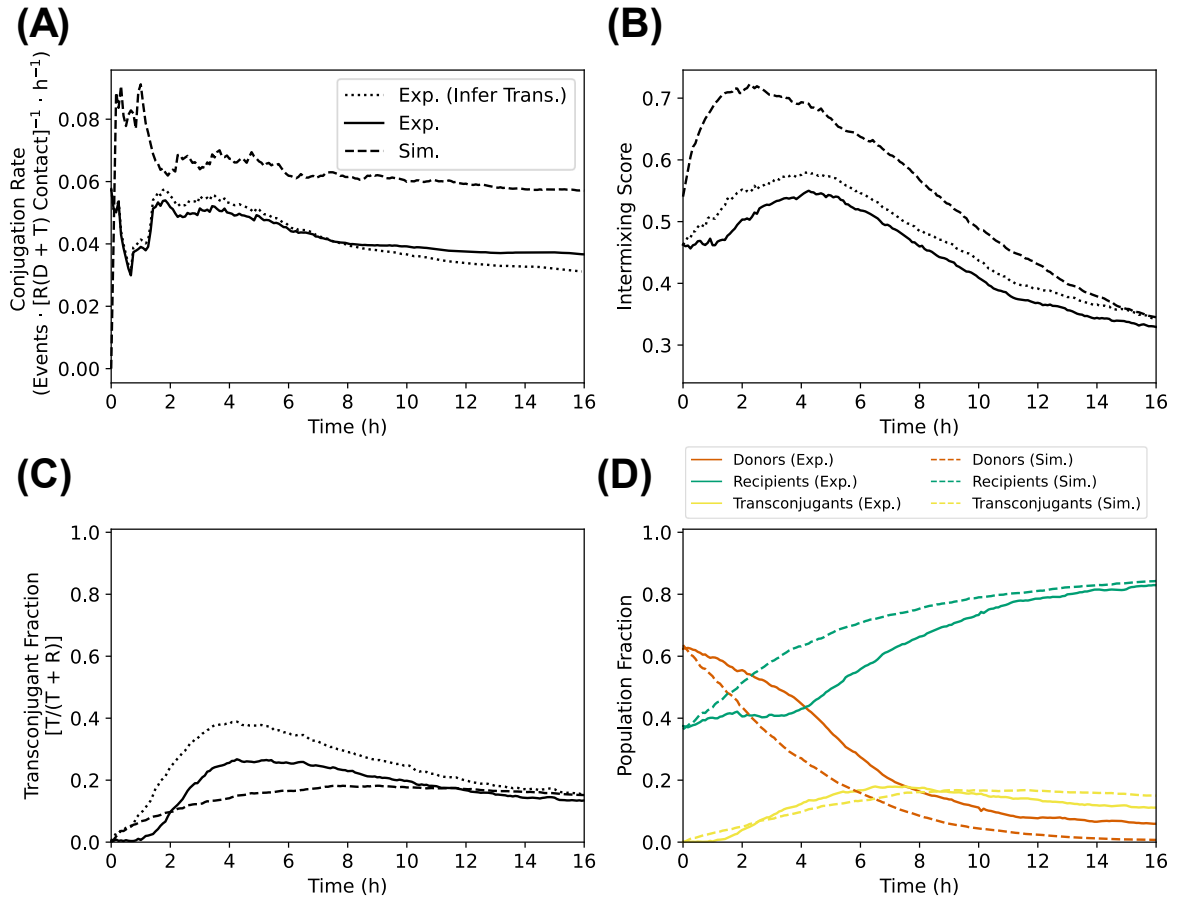


Figure 6.13. Verification of predictions from M2 with calibrated parameter values. $n = 17$ simulations with initial cell states set to initial conditions from the experiment. Means for **(A)** conjugation rate, **(B)** intermixing score, **(C)** transconjugant fraction $[T/(T+R)]$, and **(D)** population fractions for the experiment (Exp.; solid lines) and simulations (Sim.; dashed lines). Dotted lines show experimental summary statistics plotted when recipients that have received the conjugative plasmid are considered as transconjugants prior to detecting fluorescence.

6.3.6 Initial conditions influence gene delivery outcomes

Given that M2 gave reasonable predictions of our experimental observations, we used the model to explore factors affecting the overall fraction of recipients converted to transconjugants (i.e., conjugation efficiency) that would be otherwise resource-intensive to investigate experimentally. We simulated three experimental techniques that could influence the fraction of recipients converted to transconjugants (i.e., conjugation efficiency) in 80×80

μm microfluidic traps: varying the initial donor-to-recipient ratio (Case 1), varying the initial cell density (Case 2), and sequential loading of donors and recipients (Case 3; details in “Methods”). We measured summary statistics after 16 h of simulation time.

In Case 1, we ran simulations with initial D:R ratios varying from 1:10 to 10:1, with the initial number of cells kept constant at 100. In general, increasing the D:R ratio increased the endpoint transconjugant fraction, which is sensible given that this increases donor-to-recipient contacts (Figure 6.14A). However, the number of total conjugation events appeared to stop increasing after the D:R ratio was above 2:1. This could be due to the cells streamlining their orientation towards the trap exit as observed in the microscopy experiment, which then limits the number of conjugation events to the resulting number of population sectors in the trap; the low intermixing scores at the endpoint also support this observation. We note that previous work on characterizing the effect of the initial D:R ratio on plasmid spread in filter-mating experiments suggested that a 1:1 ratio was the optimal for maximizing plasmid spread¹⁹¹, but in our work we see that a ratio closer to 2:1 maximized the total number of conjugation events. This likely has to do with our donor and transconjugant populations having a higher death rate compared to the recipient population, which translates into a substantial fitness disadvantage, whereas in simulations from Zhong *et al.*¹⁹¹, the relative fitness between donors and recipients was considered as equal.

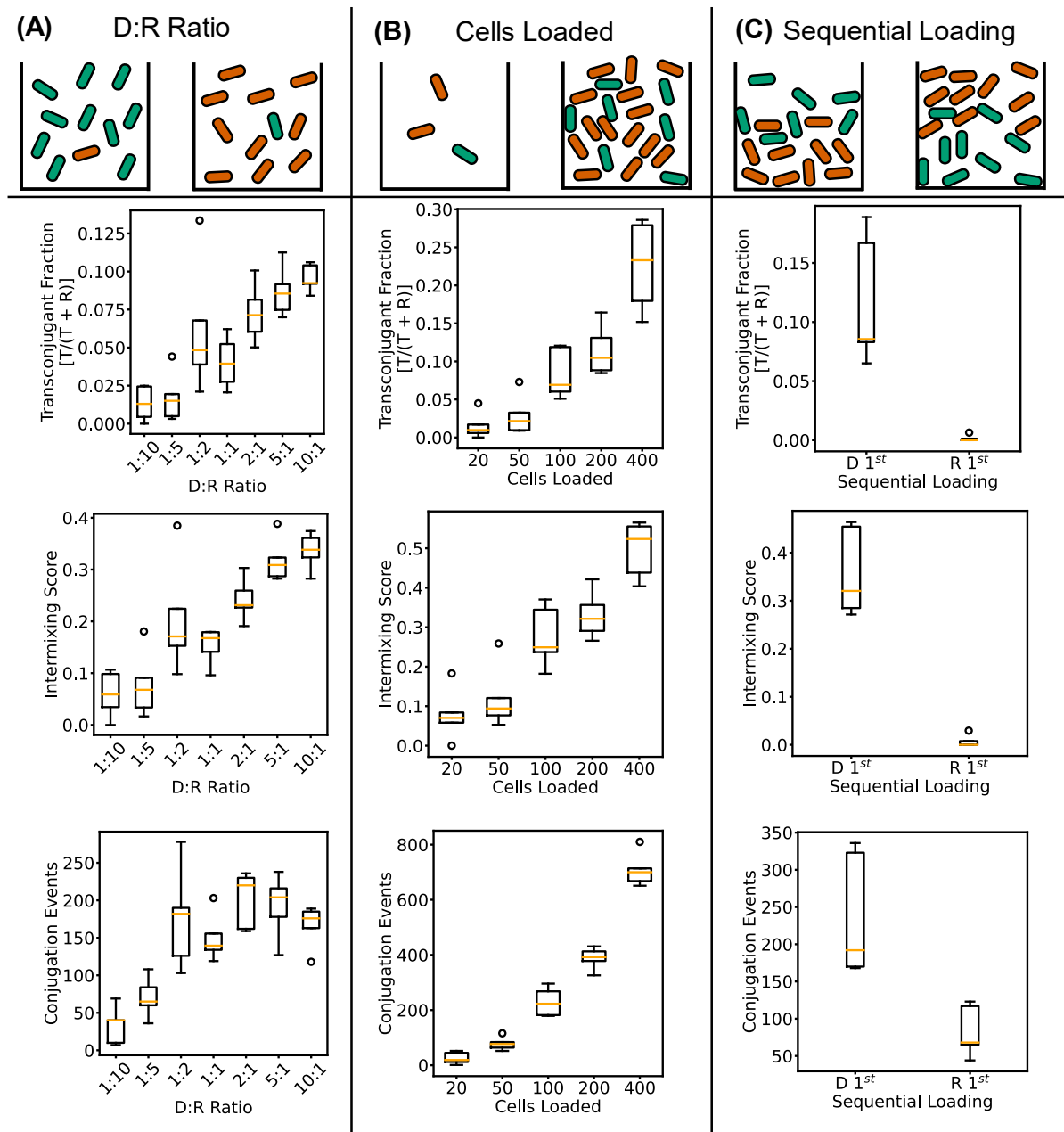


Figure 6.14. Simulated case studies for how different initial conditions in microfluidic traps influence end-point transconjugant fraction $[T/(T+R)]$, intermixing, and total conjugation events. Conditions that were tested varied were: **(A)** Initial D:R ratio, **(B)** initial number of cells loaded in the trap, and **(C)** sequential loading of different populations. Summary statistics presented are at 16 h after the simulation start time ($n = 5$ simulation replicates per condition).

In Case 2, we ran simulations with the initial number of cells ranging from 20 to 400 with a constant 2:1 D:R ratio. Increasing the number of cells loaded into the trap increased the

endpoint transconjugant fraction, intermixing, and total number of conjugation events (Figure 6.14B). This result is sensible as the cells are loaded randomly into the trap, which results in higher intermixing prior to the trap filling, increasing the likelihood of conjugation during this initial period. As an extreme comparison, a trap loaded with a single donor and recipient cell would result in two micro-colonies intermixing only at the intersection of the colonies, which is not conducive for efficient conjugative plasmid spread. This correlation between initial density and increase in the transconjugant population is consistent with population endpoint measurements from filter-mating experiments ¹⁹¹.

In Case 3, we simulated sequential loading by initializing the first population in the deeper two-thirds of the trap and the second population in the upper two-thirds, with an initial cell count of 100 cells and a 2:1 D:R ratio. When the donor population was seeded in the deeper two-thirds of the trap (i.e., seeded first), we obtained endpoint measurements for transconjugant fraction, intermixing score, and total conjugation events that were comparable to Cases 1 and 2 that had the same initial cell density and D:R ratio (Figure 6.14C). However, when the recipients were seeded in the deeper two-thirds of the trap, this resulted in endpoint measurements for the transconjugant fraction and intermixing score at zero in most replicates. Conjugation still occurred in this scenario, but both the endpoint transconjugant fraction and intermixing score indicated that the donor and transconjugant populations were washed out earlier on in the simulation. These results indicate that reaching the deepest part of the trap enables a population to compete for space, even if the population is less fit than others. This artificial fitness advantage in the space near the base of the trap is consistent with previous population dynamics studies in the hallway trap design ¹⁷⁷. In the context of conjugation, this highlights the importance of ensuring that the donor population is seeded deep into the trap if achieving a high population of transconjugants is the primary objective.

Overall, these case studies shed some insights on designing gene delivery interventions into microbial communities living in structured microenvironments. Case 1 highlighted that there is an upper limit to how much increasing the initial D:R ratio will benefit propagation of a conjugative plasmid, which is due to limitations in intermixing in unbalanced populations.

This optimum ratio will likely depend on the relative fitness between plasmid donors and recipients and could be a consideration in choosing how many donor cells to introduce. Case 2 also indicated that the initial cell density plays a major role in determining plasmid spread, suggesting that if donors can be evenly dispersed within a microbiome, densely populated microbial communities will be more susceptible to modification by conjugation compared to less populated microbiomes. However, achieving even dispersion of donors within pre-established biofilms will likely be a challenge; Case 3 demonstrated that if the recipients have established themselves in the deepest part of the trap prior to the donors, the donors will be outcompeted and pushed out of the trap.

One possible strategy for navigating around priority effects in confined spaces would be to perturb the native population (e.g., using antibiotics, bacteriocins, or mechanical agitation) prior to seeding the donors to allow donors to penetrate deeper into the space. Alternatively, the donors and transconjugants could be allowed to temporarily engraft by creating a metabolic niche for these populations. Creation of a metabolic niche to improve probiotic engraftment in the mammalian gut has been previously accomplished using porphyran (a marine polysaccharide) *in vivo* ¹⁹². This type of metabolic selection could be incorporated into a gene delivery intervention by including a metabolic gene cluster that on the conjugative plasmid that enables donors and transconjugants to consume a supplemented nutrient. Alternatively, engineered adhesion to walls could also improve the longevity of transconjugants by giving them a local advantage at colonizing the edges of structured environments ^{193–195}. Another possible strategy would be to select a donor strain that has cooperative ecological interactions (e.g., cross-feeding, syntrophy, or cross-protection) with native members in the microbiome since cooperative interactions between bacteria improve intermixing ^{196,197}.

6.4 Concluding Remarks

We devised a high-throughput methodology for measuring plasmid spread in microbial communities at single-cell level, which allows automated de-coupling horizontal and vertical gene transfer through use of standard image processing tools and lineage tracking. This

approach enables estimating the timing of individual conjugation events and the resulting conjugation rate on a -per contact basis. While the inability to identify donors responsible for specific conjugation events in our experiment imposed limitations on analyzing single-cell features about mating-pairs, our methodology is useful for investigating conjugative plasmid spread in plasmids that are labelled with fluorescent protein markers, which are a standard for experiments requiring cell sorting or cytometry^{86,87,126}. Hence, with our approach, spread of these plasmids can still be investigated with time-lapse microscopy without the need for further molecular cloning. Identification of donors and recipients involved in a mating-pair without delay could be addressed by fusing a fluorescent protein to single-stranded binding protein, which allows visualizing ssDNA transfer from donor to recipient¹³³. Measurements of transconjugant maturation time could be estimated using a labelling system where recipients express a fluorescently-labelled protein that binds to a site on the plasmid upon dsDNA synthesis, such as the *parS*/ParB system^{133,198}. These improvements can be made in future work.

Our results highlight how spatial constraints limit the spread of conjugative plasmids in densely populated microenvironments. Moreover, single-cell observations indicated that the conjugative plasmid pFAST-PETase-cis was lethal to a portion of the cells carrying it. ABM simulations indicated that this fitness burden substantially reduced plasmid spread through the population and was necessary to include in the model to reproduce experimentally observed conjugation dynamics.

We applied our previously developed systematic model calibration approach¹⁷⁵ to demonstrate its use in fitting a more complex ABM model with six tunable parameters. Calibration using the ABC-SMC method yielded satisfactory agreement between the model predictions and experimental observations at the population-level. We highlighted the utility of the calibrated model by exploring how modifications to the experimental protocol could influence conjugative plasmid spread. The simulation results indicated that the initial conditions could exacerbate the constraints that spatial factors impose on conjugative plasmid spread. Given the modular nature of ABMs, this model could be applied to model conjugation

in more complex microbial consortia or environments to gauge plasmid spread efficiency in other kinds of scenarios, such as for design of *in situ* gene delivery interventions or for estimating spread of antibiotic-resistance genes. We envision that this pipeline of using microscopy data to generate predictive models will be useful for characterizing other kinds of interactions present in microbial communities.

6.5 Methods

6.5.1 Strains, growth conditions, and plasmid construction

Escherichia coli DH5 α (peB1-gfpmut2, Addgene #103980) were used as recipients in conjugation experiments. *E. coli* NEB10- β (pFAST-PETase-cis) was used as donors¹⁸¹. The recipient strain was grown in Luria-Bertani (LB) medium with kanamycin (50 μ g/mL), and the donor strain was grown in LB medium with 0.2% glucose and 50 μ g/mL gentamicin. For growth in microfluidic traps, co-cultures were grown in LB medium with 0.1% Tween-20 with 0.2% glucose added to suppress expression of FAST-PETase.

Table 6.4. Strains and plasmids used in this study.

Name	Strain and Genotype	Description	Resistance
AY29	<i>E. coli</i> DH5 α (peB1-gfpmut2)	Recipient for pFAST-PETase-cis mating experiment	Kanamycin
AY77	<i>E. coli</i> NEB10- β (pFAST-PETase-cis)	Donor for pFAST-PETase-cis mating experiment	Gentamycin, ampicillin

6.5.2 Preparation of microfluidic devices

Molds for making microfluidic devices were made using the multilayer soft lithography technique. The layer for the cell traps was created by applying SU8 photoresist (2000.5, Microchem) onto a silicon wafer and spin-coating at 700 rpm, resulting in a layer thickness of 0.74 μ m according to a Dektak 8 stylus profilometer. A second layer of SU8 photoresist (2015, Microchem) for the main flow channel was applied onto the silicon wafer and spin-coated at 3000 rpm to achieve a layer height of 15 μ m. For each layer, features were patterned using a photomask and UV light to cross-link the photoresist.

Polydimethylsiloxane (PDMS) (Sylgard 184) was mixed with curing agent (Sylgard 184) at a 10:1 mass ratio. The mixture was degassed under vacuum until no bubbles were visible (~30 min). The PDMS was then poured over the mold and cured at 95 °C for one hour. After removing the chips from the mold, holes for the inlets and outlets were punched using a 1.2 mm biopsy corer (Harris Uni-Core No. 7093508). A 24 × 60 mm cover slip (Fisher Scientific, 12545M) was cleaned with 70% ethanol and bonded to the PDMS chip by exposing both surfaces to a plasma wand for one minute, followed by placing the chip onto the cover slip. The surfaces were then bonded at 90 °C for one hour.

6.5.3 Microscopy

Time-lapse microscopy was performed using an automated Zeiss Axio Observer inverted microscope controlled by Zen Blue 2.6 software. Imaging was done using a 63x/1.4 oil phase objective (product 420781-9910-000, Zeiss) and Axiocam 506 mono camera (Zeiss). Phase-contrast and fluorescence images were collected using a GFP filter (Chroma 49002) 470 nm/40 nm (ex.), 525 nm/50 nm (em.) and an mCherry filter (Chroma 49008) 560 nm/40 nm (ex.), 630 nm/75 nm (em.) and X-Cite 120LEDmini (Excelitas Technologies). Focus was maintained using the software's auto-focus routine using the phase-contrast channel as reference. Imaging settings are provided in Table 6.5.

Table 6.5. Microscopy imaging settings.

Setting	EGFP	mCherry	Phase
Light source intensity	10%	20%	50%
Exposure time (ms)	50	50	50

6.5.4 Microfluidic experiments

Donor and recipient strains were each inoculated from a frozen glycerol stock into 3 mL of LB media with appropriate antibiotics and grown overnight at 37 °C with shaking. Overnight cultures were diluted 1:100 into LB with appropriate antibiotics and grown at 37 °C with shaking until early exponential phase (OD₆₀₀ ~0.5). Then, 1.5 mL of each culture was then transferred to a microcentrifuge tube and washed twice with LB supplemented with 0.1%

Tween-20. The pellets were each resuspended in 100 μ L of LB supplemented with 0.1% Tween-20. Donor and recipient cultures were mixed at a 1:1 ratio and injected into the device. A 10 mL syringe containing LB supplemented with 0.1% Tween-20 and outlet tubing was connected to the device using ND-100-65 Tygon PVC tubing (product 8349T11, McMaster-Carr; inner diameter: 1/32"; outer diameter: 3/32") and stainless-steel tubing (product 5560K82, McMaster-Carr; inner diameter: 0.033"; outer diameter: 0.05") with a 90° bend. The device was secured inside of a GeneFrame incubation chamber (Pecon Incubator P S Compact) attached to the microscope stage. Cells were loaded into the traps by flicking the inlet tubing until cells approximately reached the bottom third of a trap. Media was injected continuously at 2 μ L/min using a syringe pump (GenieTouch, Kent Scientific). Images were collected every 5 min.

6.5.5 Image analysis

Time-lapse videos were manually registered to compensate for stage drift. Background subtraction in the fluorescence channels was performed by subtracting the average intensity in the main flow channel at the first time step from the entire image. Images were then cropped to contain only the area of the microfluidic trap. Overlaid fluorescence channel images were converted into single-cell data using an Omnipose-CellProfiler pipeline ¹⁹⁹, where cell segmentation was performed using Omnipose ²⁰⁰, followed by cell tracking using an overlap-based method in CellProfiler 4.2.6 ²⁰¹. Neighbour identities for each cell were estimated by expanding a segmented cell by 6 pixels (0.86 μ m), and the identities of cells overlapping with the expanded cell were recorded.

Appearance of transconjugants was detected by observing cells appear in images (with no background subtractions) from the red fluorescence channel. A subset of the images was used to train a supervised machine learning model in Ilastik 1.4.0 ¹⁸⁴ to create a mask over newly formed transconjugant cells. This mask was then shrunk by a radius of 3 pixels and applied over the cell segmentation images for the tracking step to detect the occurrence of a recipient becoming a transconjugant. The Trackrefiner package (written by Atiyeh Ahmadi; [Github](#)) was then used to correct lineage tracking errors. Cells detected in Ilastik were

considered as false positives if their red fluorescence intensity was 95% less than the donor threshold and were corrected to recipients. To supplement the machine-learning-based detection of transconjugants, a thresholding measure was used to detect transconjugants that the algorithm detected. All descendants of detected transconjugants were then labelled as transconjugants to correct for situations where Ilastik would miss identifying the same transconjugant in multiple consecutive frames. Background subtraction was performed on RFP and GFP channels using the rolling ball radius method (radius = 20 pixels) to facilitate with classifying donors and recipients. Classification thresholds for these modified intensities were determined by measuring fluorescence intensities for donors (644 A.U.) and recipients (366 A.U.) in areas of both low and high cell density. These thresholds were determined manual measurements of mean fluorescence intensity measurements of donor and recipient cells ($n = 20$ donors; $n = 23$ recipients; Appendix B). Cells with red fluorescence intensity greater than $1.5\times$ the threshold and green fluorescence $1.25\times$ the threshold were labelled as transconjugants. All cells that were not identified as transconjugants by Ilastik that had intensities below the threshold multipliers in both channels were assigned to the cell type corresponding to the channel with the highest intensity normalized to the determined thresholds.

6.5.6 Validation of automated conjugation detection

A total of nine time points from four different microfluidic traps not in the training set were inspected to validate the machine-learning-based detection of transconjugants. Transconjugant cells in selected frames were annotated by eye by the same person. A “miss” was counted if the algorithm did not identify a given transconjugant within ± 3 frames, and a false positive was counted if the algorithm highlighted a cell that was verified to not be a transconjugant over the course of the time-lapse. Annotation was done by the same person.

6.5.7 Converting tracking data to lineage tree data structure

Tracking data outputted by TrackRefiner was converted into a multi-dimensional graph using NetworkX ²⁰². These multi-dimensional graphs contain information about cell lineages and physical contacts. Each node represents a cell and its properties at a specific point in time. Two

types of edges were used to connect nodes and create the graph. Cell lineages were constructed by connecting nodes with the same ID sequentially with “lineage” edges; contact networks at each time point were created using “contact” edges to connect nodes where cell IDs were listed in the database of physical contacts from CellProfiler.

6.5.8 Measurement of single-cell properties

Cell cycles were grouped by cell ID and cell type (D, R, or T). Mean single-cell properties were calculated over the cell cycle. Lineages were filtered out if the following properties were not within the following ranges:

$$0.01 < \mu_{max} < 2.1 \text{ (h}^{-1}\text{)}$$

$$0.01 < r < 2 \text{ (}\mu\text{m)}$$

$$1 < L_{div} < 20 \text{ (}\mu\text{m)}$$

Normal distributions were fit to the growth rate and radius, whereas a lognormal distribution was fit to the division length. These fits were performed using the SciPy library ²⁰³.

6.5.9 Intermixing

An intermixing score was calculated using a normalized entropy measure that is similar to contagion indices used for landscape analysis ¹⁸⁵. The entropy was calculated based on the number of types of cell type pairs at a given point in time,

$$E = - \sum_{q=1}^{n_a} p_q \ln p_q \quad (6.1)$$

where E is the entropy of attribute adjacency, p_q is the fraction of cell type pair q , and n_a is the number of possible unique cell type. For three cell types, there are six unique cell pairs possible ($\{D,D\}$, $\{R,R\}$, $\{T,T\}$, $\{D,R\}$, $\{D,T\}$, $\{R,T\}$). The maximum entropy (E_{max}) was defined as the entropy where all attribute adjacencies were equally likely,

$$E_{max} = \ln(c^2 + c) - \ln(2) \quad (6.2)$$

where c is the number of cell types in the system. The intermixing score was then defined as E/E_{max} .

6.5.10 Isotropy

Isotropy among the cells within a microfluidic trap was quantified by the orientational order parameter (Q)²⁰⁴,

$$Q = \sqrt{\langle \cos 2\theta \rangle^2 + \langle \sin 2\theta \rangle^2} \quad (6.3)$$

where θ is the angle between the orientation of a cell and the channel direction, and the brackets indicate averaging over all cells in the trap.

6.5.11 Analysis of conjugation events

Timing of conjugation events was inferred by identifying the initial detection of a cell labelled as a transconjugant in a cell lineage, followed by finding the most recent common ancestor of all other related cells that were also identified as transconjugants. The first part of the algorithm finds all the first instances where a transconjugant was detected:

1. Obtain weakly connected components of the lineage tree data structure. Perform the following actions for each component.
2. Search through the nodes in depth-first-search order.
3. If the node is identified as a transconjugant, store the node ID in a list and do not search through the successors of the node.

Once a list of transconjugants was formulated, the occurrence of the event was inferred based on an estimate of a 100 min delay between occurrence of conjugation and detection of the resulting transconjugant. The algorithm is as follows for each identified transconjugant:

1. Create a lineage sub-tree of the ancestors and predecessors of transconjugant i that spans 100 min prior to detection and 25 min after detection.
2. Compile a set of cells S that have been identified as transconjugants within this lineage sub-tree.

3. Find the most recent common ancestor all cells in S .

The conjugation event was assumed to have happened within the cell cycle of the lowest common ancestor. The following edge cases were considered:

1. If only one transconjugant was detected in the lineage sub-tree, then the lowest common ancestor was chosen as the cell's ancestor 100 min prior.
2. If the lowest common ancestor was detected less than 50 min prior to detection of the conjugation event, the event was omitted since such a delay shorter than 50 min was not observed experimentally, so occurrence of this in the analysis would likely occur when there is a tracking error in the transconjugant lineage.

6.5.12 Conjugation rate

The conjugation rate was calculated as the number of conjugation events per contact between donors-recipients and transconjugants-recipients per hour. At time t , the cumulative conjugation rate (r_{conj}) for a given trap at time point k is calculated as follows:

$$r_{conj} = \sum_{i=1}^k \frac{Events_i}{(C_{DR_i} + C_{TR_i})\Delta t} \quad (6.4)$$

where $Events_i$ is the number of conjugation events that occurred in frame i , C_{DR_i} and C_{TR_i} are the number of current physical contacts between donors-recipients and transconjugant-recipients at frame i , respectively, and Δt is the imaging frequency.

6.5.13 Death rate

Lineage trees were obtained by removing all “contact” edges from the multi-dimensional graph representation of each time-lapse. For this analysis, the elongation rate of newly born cells was set to the average elongation rate of the parent cell and itself in the frame after birth. Average instantaneous elongation rates were calculated using a windowed average (window size: 20 min). Cells containing any large spikes in instantaneous elongation rates ($10 < \Delta L/\Delta t < -4$ $\mu\text{m}/\text{h}$) and any cells who left the trap (defined as having the cell centroid at $y > 75$ μm above

the base of the trap at any point in the cell cycle) were filtered out of the analysis. A cell's minimum windowed average elongation rate needed to drop below a specific threshold to be considered dead. This threshold was 1 $\mu\text{m/h}$ for recipients and transconjugants, and 0.5 $\mu\text{m/h}$ for donors; these thresholds were determined *ad hoc* from observing distributions of elongation rates in each population. These cells were also ensured to have no daughter cells, confirming that the lineage ended due to cell death or tracking error. The instantaneous death rate at each time was calculated as the ratio of dead cells counted over the size of the population.

6.5.14 Agent-based model

Simulations were performed in the ABM CellModeller¹⁴¹, a multi-scale physics engine for simulating growth of rod-shaped bacteria in time and space. Bacteria were modelled to grow uniaxially at a rate proportional to the exponential growth law,

$$\frac{dL}{dt} = \mu(L + 2r) \quad (6.5)$$

where L is the length of the cylindrical part of the bacterium, t is time, r is the radius of the cylinder, and μ is the cell growth rate. The additional factor of $2r$ was included since the experimental growth rate was measured using the major-axis length of each cell. When a bacterium reached its assigned division length threshold (L_{div}), it divided into two equally sized cells that inherited state variables of the parent. After this division event, a new division length threshold and growth rate were re-assigned to the daughter cells by sampling from statistical distributions fit to single-cell measurements (Table 6.2). A minimum growth rate of 0 $\mu\text{m/h}$ and a minimum division length threshold of 1.5 μm were enforced in this sampling. The biophysical parameter γ , which determines the growth penalty imposed on a cell due to physical constraints from neighbouring cells, was set to 200 (determined in Chapter 5). The regularization parameter α , which determines how permissive the numerical solver is to cell-cell overlaps, was kept at the default value of $1/\gamma$.

Cells were initialized to have the same initial position, orientation, and other cell state variables as cells in the first frame in the experiments. Cell radius was kept constant at 0.6 μm

given the low variance measured in experiments. The domain was set to $80 \times 80 \mu\text{m}$, with three wall boundary conditions and one open boundary condition. Cells were removed from the simulation when their centroid crossed the open boundary.

Conjugation was modelled as a stochastic process that was comprised of two steps: 1) activating conjugation-associated genes, and 2) selecting a neighbouring recipient. In M0, conjugation genes were activated in donors and transconjugants at a constant frequency (f_{conj}). At each time step (duration of Δt), each donor and transconjugant had a probability of activating conjugation genes equal to $f_{conj}\Delta t$. Neighbours in physical contact with the donor cell were detected using the *CompNeighbours* function in CellModeller. When conjugation genes were activated, a neighbour in physical contact with the donor was randomly selected, and if the selected cell was a “recipient” cell type, then the recipient cell was converted into a transconjugant cell and could immediately serve as a functional donor. In M1, a constant death rate (f_{death}) was assigned to donor and transconjugant cells. At each time step, each donor and transconjugant had a probability of dying (i.e. removed from the simulation) equal to $f_{death}\Delta t$.

M2 was similar to M1, but with some added complexity for selection of neighbouring recipients and some distinctions between donor and transconjugant behaviour. A pilus range was set to $1 \mu\text{m}$ based on single-cell observations of F-pili localization¹⁷² and studies on the structure of P-like pili^{205,206}. Selection of neighbouring recipients was weighted by the overlapping bounding box areas between the conjugative plasmid host and recipient (Figure 6.15). A hexagonal bounding box was used to approximate the pilus range from the periphery of the donor cell. Recipient areas were approximated using a bounding box that spanned the length and width of the cell. Selection of a neighbouring cell was then weighted proportionally to the fraction of area overlap between the donor and recipient. The option of choosing no cell (i.e., a miss) was also proportional to the area of a donor’s bounding box that was unoccupied by neighbouring cells. Transfer gene activation frequency in transconjugants ($f_{conj,T}$) was calculated as follows,

$$f_{conj,T} = f_{conj} \left(1 + m \cdot \exp \left(\frac{(t - t_m) \ln(2)}{t_{1/2}} \right) \right) \quad (6.6)$$

where f_{conj} is the baseline conjugation frequency, m is a multiplier for enhancing transfer gene activation frequency, t is the time since receiving the conjugative plasmid, t_m is the transconjugant maturation time, and $t_{1/2}$ is the half-life of the transitory derepression period. Death rates were also assigned separately to donors and transconjugants.

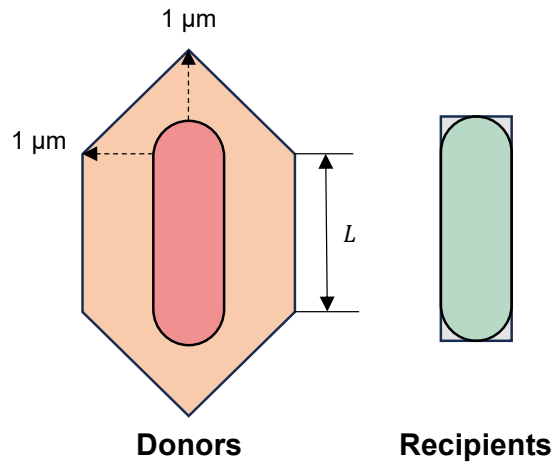


Figure 6.15. Dimensions for bounding boxes of conjugative plasmid hosts (donors) and recipients.

All simulations were performed on a Ubuntu 20.04 LTS operating system and an Intel Core i7-7700 CPU at 3.60 GHz. The time step was set to 1 min. Cell state variables were saved every 5 min of simulation time. Simulations in parametric sweep experiments of M0 and M1 and simulations for gene delivery interventions with M2 were run for 16 h of simulated time. Other simulation parameters are listed in Table 6.6.

Table 6.6. CellModeller simulation settings.

Name	Symbol	Value	Units
Domain width	L_x	80	μm
Domain depth	L_y	80	μm
Cell radius	r	0.6	μm
Division axis noise	ζ	$U(-0.001, 0.001)$	rad

Solver tolerance	ϵ_c	1×10^{-5}	-
Max. contacts per cell	M_{cts}	36	-
Cell grid element size	L_{grid}	5	μm
Time step	Δt	1	min
Growth penalty factor	γ	200	$\text{Pa}\cdot\text{s}\cdot\text{m}$
Regularization parameter	α	1/200	-

6.5.15 Model calibration

Posterior distributions for model parameter values were fit using the Approximate Bayesian Computation Scheduled Monte Carlo (ABC-SMC) method implemented in pyABC ¹⁷⁶. ABC-SMC algorithm parameters are listed in Table 6.7. Initial conditions for simulations were randomly chosen from one of the 17 traps in the pFAST-PETase-cis mating experiment. Simulations were ran for 8 hours of simulated time and compared to experimental summary statistics at that time using the Euclidean distance. The total distance was the sum of Euclidean distances for each summary statistics (i.e., all summary statistics were weighted equally). The conjugation rate, intermixing score, and transconjugant fraction $[T/(T+R)]$ from the final frame were used for the distance calculation. The error tolerance (ϵ) for accepting simulated runs was sequentially decreased with each calibration round by setting ϵ to the median error of the previous calibration round. The initial error tolerance was chosen as the median error of 40 simulations with parameters sampled from the prior distributions. Uninformative prior distributions were chosen for each parameter set; ranges are listed in Table 6.7. A calibration round was terminated when 40 simulations were accepted. Kernel density estimate plots were made using the *plot_kde_1d* function in pyABC with default settings (gaussian kernel). Parameter values for subsequent use of Model 2 were obtained from the weighted average of the probability density function for each parameter.

Table 6.7. Prior distributions used in ABC-SMC calibration.

Parameter	Range	Source
Transfer gene activation frequency (h^{-1})	[0, 1]	This study
Transconjugant maturation time (h)	[0, 2]	¹⁸⁹

Donor death rate (h^{-1})	[0, 0.5]	This study
Transconjugant death rate (h^{-1})	[0, 0.5]	This study
Transitory derepression multiplier	[1, 5]	¹⁹⁰
Transitory derepression half-life (h)	[0, 1]	This study

6.5.16 Simulation case studies on initial conditions

The three case studies investigated were: 1) changing the initial D:R ratio, 2) changing the initial cell density, and 3) sequential loading of donor and recipient populations. These will be referred to as Case 1 through 3. All simulations were performed with Model 2 using the same simulation settings described in Table 6.6. In all Cases, cell division lengths and growth rates were initialized from experimentally measured probability distributions (Table 6.2), and the initial cell length was set to one-half the division length. Cells were seeded with random orientations with a minimum distance of 2 μm between cell centroids to avoid cells overlapping upon initialization. Simulations were run for 16 h of simulated time. Five replicates per simulation setting were performed. Settings for each Case are summarized in Table 6.8.

Table 6.8. Initialization settings in simulation case studies.

Setting	Case 1	Case 2	Case 3
Initialization of cell positions	Random	Random	Random in upper and lower 2/3 rd s of domain
Initial number of cells	100	Variable	100
Initial D:R ratio	Variable	2:1	2:1

6.6 Acknowledgements

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC). A.Y. was also supported by funding from the Ken O'Driscoll Graduate Scholarship in Polymer Engineering/Science, and Murray Moo Young Biochemical Engineering Scholarship. We are grateful to Atiyeh Ahmadi for assistance in installing and using the CellProfiler-Omnipose image processing modules, as well as supplying the Trackrefiner lineage correction software. We are also thankful for assistance from Vincent

Demont (Visiting Master's student, Université Paul Sabatier Toulouse, France) in early-phase testing of microfluidic chips and protocols.

Chapter 7

Discussion

Bacterial conjugation is a vehicle for gene delivery into microbiomes that offers flexibility with host range, genetic cargo size, conjugation frequency, and efficiency in solid/liquid environments. With continued advancements in developing genetic parts and gene-editing tools for non-model organisms^{15,86,207}, there are increasing possibilities for expanding the host-range and specificity of delivered functionalities. However, predicting the outcomes of *in situ* manipulation remains a daunting task given the biotic and abiotic complexity in microbiomes. The purpose of this work was to develop tools for microbiome engineering through conjugation and investigate the efficacy of these tools against problems relevant to global challenges like plastic pollution.

7.1 Engineering microbiomes to degrade plastics

Advances in engineering plastic-depolymerizing enzymes like FAST-PETase and MHETase variants has enabled depolymerization of PET in a matter of days. In principle, genes encoding for these engineered enzymes could be delivered into microbiomes to enable rapid biodegradation of plastics. Chapter 3 explored the feasibility of this idea, by delivering FAST-PETase into a municipal wastewater microbiome via conjugation. The proof-of-concept study was successful, whereby several transconjugant isolates could express FAST-PETase and release the enzyme into the extracellular environment, resulting in depolymerization of commercial and post-consumer PET.

Several challenges need to be overcome to achieve *in situ* degradation of plastics and bring the technology closer to application, such as biological barriers to gene delivery, gene expression, selection and maintenance of engineered microbes, maintaining operating conditions that support enzyme activity, and addressing biological and ecological safety concerns. In Chapter 3, the diversity of isolated transconjugants was limited due to the manual selection method on minimal glucose media and the reliance on a single promoter for selection

by expression of a fluorescent protein. Using a library of ribosome binding sites and promoters for selection markers coupled with fluorescence activated cell sorting (FACS) could improve the observed diversity of transconjugants and would facilitate screening for functional promoters for expression of PETase. Using diverse donor strains and conjugative plasmid systems could also improve conjugation efficiency to different members of the microbiome. The isolated transconjugants lost the plasmid without selection, so additional measures such as genomic integration or temporary selection pressure could be taken to maintain the gene of interest.

The activity of FAST-PETase was highly dependent on the media in which the depolymerization assay was performed in. Abiotic conditions would likely be infeasible to control in environments where microplastics are present like wastewater treatment plants, soil, and open water bodies. Hence, further testing of PETase activity in environmentally relevant conditions will be necessary prior to in-field applications. Protein engineering could be leveraged to optimize PETase activity at ambient temperatures and in specific environmental conditions ²⁰⁸. Cell-surface display of PETases could be useful to implement in future work because cell-surface display has been shown to improve PET degradation compared to using PETase in solution ^{67,74,75}. Another question that needs addressing is the efficacy of genetic bioaugmentation on plastic degradation *in situ*; instead of testing degradation using microbial isolates, future work should investigate plastic degradation after *in situ* gene delivery to assess performance of the proposed technology.

Given the plethora of plastic-degrading microbes that have been documented ⁵⁰, it may be possible to move towards degrading other kinds of plastics using genetic bioaugmentation. Future discoveries of mechanisms for biological degradation of plastics can be leveraged by adding the corresponding gene clusters responsible to the gene delivery vector. For plastics that lack functional groups (and are hence less amenable to enzymatic attack), it would be interesting to explore the introduction of cleavable comonomer additives ²⁰⁹. These links in the polymer could enable deconstruction of polymers into molecules that could be amenable to further enzymatic degradation. Or, a combined bio-chemical process could be used to

biologically deconstruct collected plastics and chemically upcycle the oligomers. Additional advancements through protein engineering could also improve the efficacy of genetic bioaugmentation for bioremediation of plastics in the environment.

Finally, biosafety will be important to consider in application of the technology. Antibiotic resistance genes should be removed from the gene delivery vector prior to any field studies. This can be achieved by using auxotrophic- and essential gene-based selection can be used to select and maintain engineered plasmids in transformants ²¹⁰. The majority of transconjugant isolates were opportunistic pathogens; escape of these microbes that could have a fitness advantage by consuming plastic as a carbon and energy source would be undesirable. Delivery to select microbes that are Generally Recognized As Safe (GRAS) could be achieved using CRISPR associated transposases ¹⁵, whereby the gene of interest can only integrate if a desired sequence is present in the host genome. Another option for biocontainment would be to use a mobilizable plasmid instead of a conjugative plasmid to reduce the possibility of additional propagation of the plasmid by transconjugants. Ideally, engineered bacteria would only persist in the presence of microplastics due to the fitness advantage of using the plastic monomers as a carbon or energy source; further investigation on any potential fitness advantage and changes to microbial community composition will help address some of these ecological questions.

7.2 Gene delivery into microbial communities

For *in situ* microbiome engineering applications, the success of genetic bioaugmentation relies on the propagation of the genetic payload through both horizontal and vertical gene transfer. While propagation of a conjugative plasmid is always constrained by biological factors like the plasmid's conjugation frequency and the recipient's ability to uptake and maintain the plasmid, propagation of a conjugative plasmid in spatially heterogeneous environments is further constrained by intermixing between populations and competition for space. Together, these biological and spatial challenges complicate the task of predicting the outcomes of a genetic bioaugmentation operation. Effects of these spatial constraints were directly observed in a mating experiment performed in microfluidic traps in Chapter 6. The conjugative plasmids

investigated posed a substantial metabolic burden on their hosts, causing them to get outcompeted by the faster-growing recipient population. However, this trend was not universal in all traps due to some donors and transconjugants reaching the deepest part of the trap, which effectively prevents the population from getting washed out.

The pipeline used to analyze conjugation in microfluidic traps could be used to simultaneously measure HGT between numerous species, so long as different bacterial species can be uniquely identified (e.g., through fluorescence or morphology). This could be useful for simultaneously measuring conjugation frequency between individual members as well as how conjugative plasmids propagate in more complex microbial consortia. Such pair-wise estimates of conjugation frequencies between donor and recipient species could then be implemented in ABMs to infer plasmid spread in more complex environments.

Matings were performed by seeding empty traps with donors and recipients, which helps eliminate priority effects in competition for space. The impact of priority effects on plasmid propagation could be investigated through sequential loadings of the recipient and donor strains into the microfluidic device. Microbiomes in real environments would already be established, so the donor cells would likely need engineered assistance to temporarily engraft long enough to deliver the genetic payload. This could be implemented by creating a metabolic niche for the donor cells or by improving their ability to compete for space through engineered expression of bacteriocins^{192,211}.

For *in situ* microbiome engineering applications, microfluidic chips offer a more refined level of abstraction compared to growth on agarose pads. The presence of a constant nutrient source and the small micro-channel have some resemblance to more complex environments where horizontal gene transfer is relevant, such as villi in the mammalian gut or channels in soil. The abstract representation that microfluidic environments provide is a first step to developing more complex *in vitro* models for investigating genetic bioaugmentation. The microfluidic environment is still very much an abstraction of real microenvironments due to the trap's simple 2D square geometry and chemically inert walls. The geometric complexity

could be increased by changing the trap to allow growth of cells in 3D, and would require frequent imaging with a confocal microscope and more sophisticated image processing methods. The environmental complexity could also be increased by having mass transfer at the trap boundaries, which would be implemented through using a “gut on a chip” platform²¹², or perhaps by infusing small molecules into the material of the microfluidic chip that would diffuse into the trap upon seeding.

Given the vast number of combinations for species and plasmids that could be involved in conjugation, it would be useful to have catalogue conjugation rates for plasmid-donor-recipient combinations for choosing parameters in a conjugation model. This catalogue should include information about donor and recipient species, the conjugation system, the mating conditions, and the resulting conjugation rate. In the long-term, an expansive catalogue could lead towards more predictive mechanistic models and could also lead to understanding new patterns governing gene transfer frequencies in microbial communities. The challenge with generating such a catalogue is that historically, conjugation has been quantified using numerous metrics, many of which are reported as end-point population ratios rather than as horizontal gene transfer rates. A guide has been recently published for standardizing reports of conjugation rates¹²⁵, which should allow fair comparisons between experiments across different labs.

7.3 Agent-based modelling of microbial communities

ABMs provide enough flexibility to allow consideration of both biological factors and spatial factors, making them an attractive tool for exploring scientific questions about horizontal gene transfer and anticipating outcomes of genetic bioaugmentation. In Chapter 5, a POM-inspired pipeline for calibrating ABMs to microscopy data was applied to reproduce patterns in 2D micro-colony formation on flat surfaces. This pipeline was then applied in Chapter 6 to reproduce dynamics of a plasmid propagating through populations in a microfluidic environment. The calibrated model enabled rapid and controlled exploration of how spatio-temporal factors influence conjugative plasmid spread in constrained microenvironments. This

type of thorough exploration would have been substantially more laborious to perform experimentally, which emphasizes the practical value of well-calibrated, mechanistic models.

Chapters 5 and 6 highlight a methodology to develop calibrated, well-documented models for simulating applications in synthetic biology and microbiome engineering. The conjugation model was calibrated using ABC, which is an inference-based method for calibrating models to observations. This is an improvement over less rigorous calibration methods described in Chapter 4 that are commonly used in developing ABMs of microbial communities. The use of ABC aligns well with the recommended POM strategy for creating predictive ABMs. Unlike other methods like Maximum Likelihood estimation and Markov Chain Monte Carlo methods, ABC enables performing likelihood-free inference¹⁷⁶. This is useful for calibrating ABMs that often have likelihood functions that are unknown or infeasible to compute. Furthermore, ABC simultaneously achieves parameter fitting and uncertainty quantification in the same step, whereas other optimization methods like Gradient Descent and Simulated Annealing only perform the parameter fitting step. More generally, use of a quantitative fitting procedure forces the model developer to quantify only important “patterns” that the model must reproduce; the ABC framework is flexible enough to enable weighting different patterns based on their relative importance to the developer. ABC is a reproducible algorithm, but still requires some human judgement in selection of which features to calibrate to, selection of prior distributions, and sample sizes. While the latter two decisions remain up to personal discretion, there could be future opportunities to automate feature selection using deep learning.

The advantage that ABMs have over statistical and machine learning models is that ABMs are mechanistic, which enables understanding emerging features in the data from defined rules in the model. This makes ABMs a useful tool for asking scientific questions and understanding general trends. The model calibration pipeline presented in this thesis can start moving ABMs toward predictive purposes by allowing quantitative reproduction of spatio-temporal patterns. The long-term vision here is to advance ABMs to a point where they can be used for engineering microbiomes in real applications, in a similar fashion to how

computational fluid dynamics models are advanced enough to make predictions for engineering design of airplanes, heat exchangers, distillation columns, etc. One of the challenges in bringing ABMs to that point is the array of biological nuances that create uncertainty in model parameterizations. For example, in the context of this work, the conjugation frequency is governed by factors such as the type of plasmid, relatedness between donor and recipient, temperature, cell metabolism, pH, salinity, etc. The relationship between conjugation frequency and abiotic conditions could be addressed through empirical correlations, but more sophisticated approaches (e.g., neural networks) may be necessary for describing the relationship between conjugation frequency and biological factors like plasmid-donor-recipient triplet combinations. The trade-off between experimental effort to gather the necessary data for the model and the general utility of the ABM will likely dictate how quickly ABMs can move towards making predictions about complex microbial systems.

The question remains open about how the ABM developed in this thesis could help with using genetic bioaugmentation for degrading plastics. Although the ABM for conjugation developed in this thesis is strictly parameterized for conjugation of an IncP plasmid between lab strains of *E. coli*, the utility of this model comes in its ability to predict gene transfer in structured environments. CellModeller is also capable of modelling mass transport, which can be further augmented using finite-element method software¹⁵². Together, the conjugation and mass transport modules could be used for estimating how degradation of plastic surfaces influences gene transfer (Figure 7.1). Likewise, local gradients of PETase and PET monomers arising from the bioaugmentation treatment could be predicted to determine if local conditions created in a biofilm enhance or diminish the surface reaction rate, and to quantify the rate of degradation products released into the environment. Such simulations could enhance understanding the fitness advantage and local niche that the PET monomers would provide for engineered or natural microbes, which could facilitate with designing microbial communities for efficient plastic degradation and with ecological risk analysis.

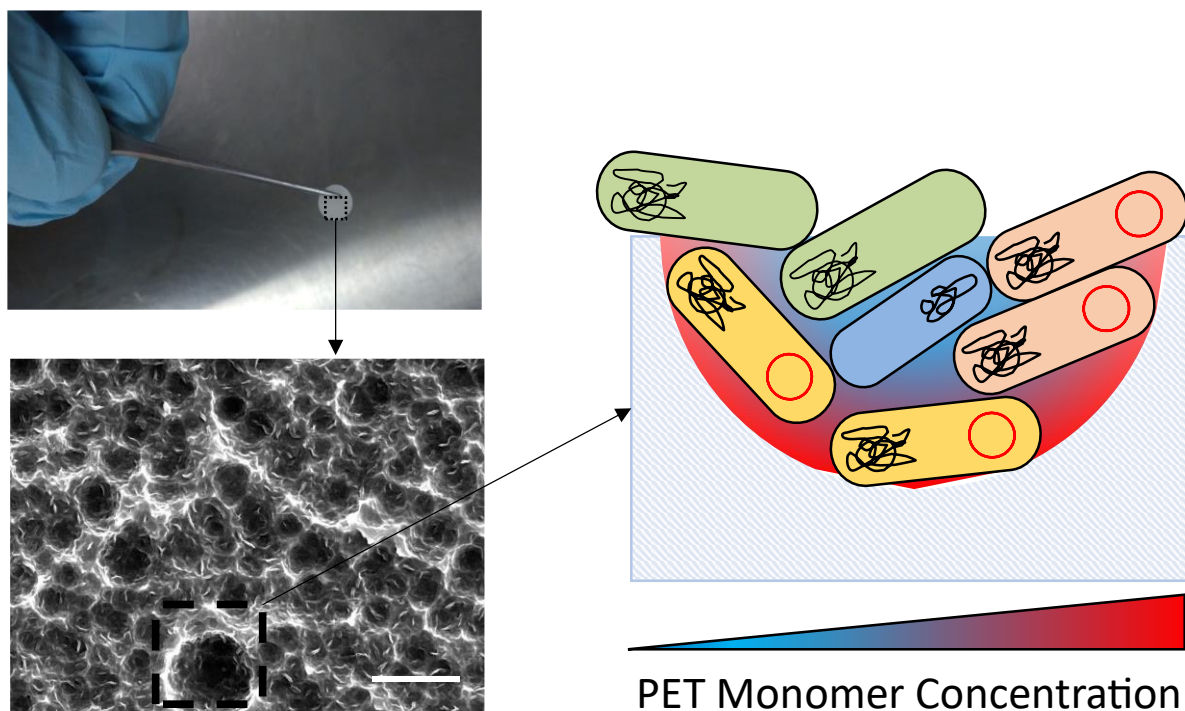


Figure 7.1. An example of how a calibrated agent-based model of horizontal gene transfer in microbial communities could be used to make predictions about the efficacy of genetic bioaugmentation for plastic degradation. The degraded piece of PET from herbs packaging degraded by FAST-PETase (top left) has a pitted surface structure (bottom left) which could lead to local micro-environments for biofilm formation, conjugation, and additional plastic degradation. Scale bar: 10 μm .

7.4 Conclusions

This thesis presented several novel tools for manipulating microbial communities through delivery of conjugative plasmids: a conjugative plasmid that enables diverse bacteria to degrade PET, a POM-inspired methodology for calibrating spatio-temporal models to microscopy data, a microfluidics and bioinformatics platform to quantify the fate of conjugative plasmids in heterogeneous environments, and an ABM for predicting manipulations to microbial communities through conjugation. Together, I used these tools to answer specific questions about the feasibility of using genetic bioaugmentation for degrading PET, and more general questions about predicting spread of conjugative plasmids in spatially heterogeneous environments.

Some bacteria in wastewater were amenable to *in situ* genetic modification and could release FAST-PETase into the extracellular environment that would depolymerize post-consumer PET. Importantly, this work demonstrates the feasibility of using genetic bioaugmentation for biodegradation of petroleum-based plastics. Moreover, this proof-of-concept study opens opportunities for building diverse microbiomes to autonomously biodegrade PET plastics and other recalcitrant materials. I then demonstrated a high-throughput, single-cell-based approach for investigating conjugation dynamics in microfluidic traps. I calibrated an ABM against this data set, and the model was able to reproduce observed patterns in plasmid spread through the microbial community, as well as characterize spatio-temporal factors influencing plasmid spread.

Predicting outcomes of genetic bioaugmentation is challenging, but predictive ABMs can facilitate investigation during the early phases of a project and narrow down the design space prior to laboratory testing. They can also enable further investigation of micro-scale phenomena that cannot be tested experimentally. Importantly, this thesis demonstrates that conjugative plasmids are a suitable vehicle for efficiently delivering novel functionalities into microbial communities, and that well-documented, calibrated ABMs can predict conjugative plasmid spread in complex microenvironments.

References

- (1) Cohen, S. N., Chang, A. C. Y., Boyer, H. W., and Helling, R. B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proceedings of the National Academy of Sciences* 70, 3240–3244.
- (2) Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. D. (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences* 76, 106–110.
- (3) Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943.
- (4) Lawson, C. E., Harcombe, W. R., Hatzenpichler, R., Lindemann, S. R., Löffler, F. E., O'Malley, M. A., García Martín, H., Pfleger, B. F., Raskin, L., Venturelli, O. S., Weissbrodt, D. G., Noguera, D. R., and McMahon, K. D. (2019) Common principles and best practices for engineering microbiomes. *Nat Rev Microbiol* 17, 725–741.
- (5) Brenner, K., You, L., and Arnold, F. H. (2008) Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology* 26, 483–489.
- (6) Tsoi, R., Dai, Z., and You, L. (2019) Emerging strategies for engineering microbial communities. *Biotechnology Advances* 37, 107372.
- (7) Ma, H., Zhao, Y., Yang, K., Wang, Y., Zhang, C., and Ji, M. (2022) Application oriented bioaugmentation processes: Mechanism, performance improvement and scale-up. *Bioresource Technology* 344, 126192.
- (8) Albright, M. B. N., Louca, S., Winkler, D. E., Feeser, K. L., Haig, S.-J., Whiteson, K. L., Emerson, J. B., and Dunbar, J. (2022) Solutions in microbiome engineering: prioritizing barriers to organism establishment. *ISME J* 16, 331–338.
- (9) Chettri, D., Verma, A. K., and Verma, A. K. (2023) Bioaugmentation: an approach to biological treatment of pollutants. *Biodegradation* 35, 117–135.
- (10) Bhatt, P., Bhandari, G., Bhatt, K., Maithani, D., Mishra, S., Gangola, S., Bhatt, R., Huang, Y., and Chen, S. (2021) Plasmid-mediated catabolism for the removal of xenobiotics from the environment. *Journal of Hazardous Materials* 420, 126618.
- (11) Löffler, F. E., and Edwards, E. A. (2006) Harnessing microbial activities for environmental cleanup. *Current Opinion in Biotechnology* 17, 274–284.
- (12) French, K. E., Zhou, Z., and Terry, N. (2020) Horizontal ‘gene drives’ harness indigenous bacteria for bioremediation. *Sci Rep* 10, 15091.
- (13) Peters, M., Heinaru, E., Talpsep, E., Wand, H., Stottmeister, U., Heinaru, A., and Nurk, A. (1997) Acquisition of a deliberately introduced phenol degradation operon, pheBA, by different indigenous *Pseudomonas* species. *Applied and Environmental Microbiology* 63, 4899–4906.
- (14) Brophy, J. A. N., Triassi, A. J., Adams, B. L., Renberg, R. L., Stratis-Cullum, D. N., Grossman, A. D., and Voigt, C. A. (2018) Engineered integrative and conjugative

- elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat Microbiol* 3, 1043–1053.
- (15) Rubin, B. E., Diamond, S., Cress, B. F., Crits-Christoph, A., Lou, Y. C., Borges, A. L., Shivram, H., He, C., Xu, M., Zhou, Z., Smith, S. J., Rovinsky, R., Smock, D. C. J., Tang, K., Owens, T. K., Krishnappa, N., Sachdeva, R., Barrangou, R., Deutschbauer, A. M., Banfield, J. F., and Doudna, J. A. (2022) Species- and site-specific genome editing in complex bacterial communities. *Nat Microbiol* 7, 34–47.
 - (16) Vo, P. L. H., Ronda, C., Klompe, S. E., Chen, E. E., Acree, C., Wang, H. H., and Sternberg, S. H. (2021) CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat Biotechnol* 39, 480–489.
 - (17) Bikard, D., and Barrangou, R. (2017) Using CRISPR-Cas systems as antimicrobials. *Current Opinion in Microbiology* 37, 155–160.
 - (18) Ramachandran, G., and Bikard, D. (2019) Editing the microbiome the CRISPR way. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374, 20180103.
 - (19) Rao, V. B., and Zhu, J. (2022) Bacteriophage T4 as a nanovehicle for delivery of genes and therapeutics into human cells. *Current Opinion in Virology* 55, 101255.
 - (20) Cazares, A., Moore, M. P., Hall, J. P. J., Wright, L. L., Grimes, M., Emond-Rhéault, J.-G., Pongchaikul, P., Santanirand, P., Levesque, R. C., Fothergill, J. L., and Winstanley, C. (2020) A megaplasmid family driving dissemination of multidrug resistance in *Pseudomonas*. *Nat Commun* 11, 1370.
 - (21) Jain, A., and Srivastava, P. (2013) Broad host range plasmids. *FEMS Microbiology Letters* 348, 87–96.
 - (22) Musovic, S., Oregaard, G., Kroer, N., and Sørensen, S. J. (2006) Cultivation-Independent Examination of Horizontal Transfer and Host Range of an IncP-1 Plasmid among Gram-Positive and Gram-Negative Bacteria Indigenous to the Barley Rhizosphere. *Applied and Environmental Microbiology* 72, 6687–6692.
 - (23) Bates, S., Cashmore, A. M., and Wilkins, B. M. (1998) IncP Plasmids Are Unusually Effective in Mediating Conjugation of *Escherichia coli* and *Saccharomyces cerevisiae*: Involvement of the Tra2 Mating System. *Journal of Bacteriology* 180, 6538–6543.
 - (24) Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., Brunson, J. K., Valas, R. E., Deerinck, T. J., Jablanovic, J., Gillard, J. T. F., Beeri, K., Ellisman, M. H., Glass, J. I., Hutchison III, C. A., Smith, H. O., Venter, J. C., Allen, A. E., Dupont, C. L., and Weyman, P. D. (2015) Designer diatom episomes delivered by bacterial conjugation. *Nat Commun* 6, 6925.
 - (25) Molin, S., and Tolker-Nielsen, T. (2003) Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. *Current Opinion in Biotechnology* 14, 255–261.
 - (26) Bikard, D., Euler, C. W., Jiang, W., Nussenzweig, P. M., Goldberg, G. W., Duportet, X., Fischetti, V. A., and Marraffini, L. A. (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nat Biotechnol* 32, 1146–1150.
 - (27) Voorhees, P. J., Cruz-Teran, C., Edelstein, J., and Lai, S. K. (2020) Challenges & opportunities for phage-based *in situ* microbiome engineering in the gut. *Journal of Controlled Release* 326, 106–119.

- (28) Neil, K., Allard, N., Grenier, F., Burrus, V., and Rodrigue, S. (2020) Highly efficient gene transfer in the mouse gut microbiota is enabled by the IncI2 conjugative plasmid TP114. *Commun Biol* 3, 1–9.
- (29) Arias-Andres, M., Klümper, U., Rojas-Jimenez, K., and Grossart, H.-P. (2018) Microplastic pollution increases gene exchange in aquatic ecosystems. *Environmental Pollution* 237, 253–261.
- (30) Zhou, Y., Zhang, G., Zhang, D., Zhu, N., Bo, J., Meng, X., Chen, Y., Qin, Y., Liu, H., and Li, W. (2024) Microplastic biofilms promote the horizontal transfer of antibiotic resistance genes in estuarine environments. *Marine Environmental Research* 202, 106777.
- (31) Luo, T., Dai, X., Wei, W., Xu, Q., and Ni, B.-J. (2023) Microplastics Enhance the Prevalence of Antibiotic Resistance Genes in Anaerobic Sludge Digestion by Enriching Antibiotic-Resistant Bacteria in Surface Biofilm and Facilitating the Vertical and Horizontal Gene Transfer. *Environ. Sci. Technol.* 57, 14611–14621.
- (32) Liu, X., Xu, R., Wu, H., Xu, K., Zhang, W., Wang, Z., and Gong, W. (2023) Nanoplastics promote the dissemination of antibiotic resistance through conjugative gene transfer: implications from oxidative stress and gene expression. *Environmental Science: Nano* 10, 1329–1340.
- (33) Chattopadhyay, I. (2022) Role of microbiome and biofilm in environmental plastic degradation. *Biocatalysis and Agricultural Biotechnology* 39, 102263.
- (34) Nguyen, P. Q., Huang, X., Collins, D. S., Collins, J. J., and Lu, T. (2023) Harnessing synthetic biology to enhance ocean health. *Trends in Biotechnology*.
- (35) Gustafsson, K., and Jansson, J. K. (1993) Ecological risk assessment of the deliberate release of genetically modified microorganisms. *Ambio* 22, 236–242.
- (36) de Lorenzo, V. (2022) Environmental Galenics: large-scale fortification of extant microbiomes with engineered bioremediation agents. *Philosophical Transactions of the Royal Society B: Biological Sciences* 377, 20210395.
- (37) Geyer, R., Jambeck, J. R., and Law, K. L. (2017) Production, use, and fate of all plastics ever made. *Science* 3, e1700782.
- (38) Chamas, A., Moon, H., Zheng, J., Qiu, Y., Tabassum, T., Jang, J. H., Abu-Omar, M., Scott, S. L., and Suh, S. (2020) Degradation rates of plastics in the environment. *ACS Sustainable Chem. Eng.* 8, 3494–3511.
- (39) Frias, J. P. G. L., and Nash, R. (2019) Microplastics: Finding a consensus on the definition. *Marine Pollution Bulletin* 138, 145–147.
- (40) Yuan, Z., Nag, R., and Cummins, E. (2022) Human health concerns regarding microplastics in the aquatic environment - From marine to food systems. *Science of The Total Environment* 823, 153730.
- (41) Wang, W., Gao, H., Jin, S., Li, R., and Na, G. (2019) The ecotoxicological effects of microplastics on aquatic food web, from primary producer to human: A review. *Ecotoxicology and Environmental Safety* 173, 110–117.
- (42) Xu, S., Ma, J., Ji, R., Pan, K., and Miao, A.-J. (2020) Microplastics in aquatic environments: Occurrence, accumulation, and biological effects. *Science of The Total Environment* 703, 134699.

- (43) Holmes, L. A., Turner, A., and Thompson, R. C. (2012) Adsorption of trace metals to plastic resin pellets in the marine environment. *Environmental Pollution* 160, 42–48.
- (44) Leslie, H. A., van Velzen, M. J. M., Brandsma, S. H., Vethaak, A. D., Garcia-Vallejo, J. J., and Lamoree, M. H. (2022) Discovery and quantification of plastic particle pollution in human blood. *Environment International* 163, 107199.
- (45) Ragusa, A., Svelato, A., Santacroce, C., Catalano, P., Notarstefano, V., Carnevali, O., Papa, F., Rongioletti, M. C. A., Baiocco, F., Draghi, S., D’Amore, E., Rinaldo, D., Matta, M., and Giorgini, E. (2021) Plasticenta: First evidence of microplastics in human placenta. *Environment International* 146, 106274.
- (46) Vogt, B. D., Stokes, K. K., and Kumar, S. K. (2021) Why is Recycling of Postconsumer Plastics so Challenging? *ACS Appl. Polym. Mater.* 3, 4325–4346.
- (47) Ali, S. S., Abdelkarim, E. A., Elsamahy, T., Al-Tohamy, R., Li, F., Kornaros, M., Zuurro, A., Zhu, D., and Sun, J. (2023) Bioplastic production in terms of life cycle assessment: A state-of-the-art review. *Environmental Science and Ecotechnology* 15, 100254.
- (48) Lau, W. W. Y., Shiran, Y., Bailey, R. M., Cook, E., Stuchtey, M. R., Koskella, J., Velis, C. A., Godfrey, L., Boucher, J., Murphy, M. B., Thompson, R. C., Jankowska, E., Castillo Castillo, A., Pilditch, T. D., Dixon, B., Koerselman, L., Kosior, E., Favoino, E., Gutberlet, J., Baulch, S., Atreya, M. E., Fischer, D., He, K. K., Petit, M. M., Sumaila, U. R., Neil, E., Bernhofen, M. V., Lawrence, K., and Palardy, J. E. (2020) Evaluating scenarios toward zero plastic pollution. *Science* 369, 1455–1461.
- (49) Fuhs, G. W. (1961) Der mikrobielle Abbau von Kohlenwasserstoffen. *Archiv. Mikrobiol.* 39, 374–422.
- (50) Gambarini, V., Pantos, O., Kingsbury, J. M., Weaver, L., Handley, K. M., and Lear, G. (2022) PlasticDB: a database of microorganisms and proteins linked to plastic biodegradation. *Database* 2022, baac008.
- (51) Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., and Oda, K. (2016) A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* 351, 1196–1199.
- (52) Sulaiman, S., Yamato, S., Kanaya, E., Kim, J.-J., Koga, Y., Takano, K., and Kanaya, S. (2012) Isolation of a Novel Cutinase Homolog with Polyethylene Terephthalate-Degrading Activity from Leaf-Branch Compost by Using a Metagenomic Approach. *Applied and Environmental Microbiology* 78, 1556–1562.
- (53) Joo, S., Cho, I. J., Seo, H., Son, H. F., Sagong, H.-Y., Shin, T. J., Choi, S. Y., Lee, S. Y., and Kim, K.-J. (2018) Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat Commun* 9, 382.
- (54) Harwood, C. S., and Parales, R. E. (1996) THE β -ketoadipate pathway and the biology of self-identity. *Annual Review of Microbiology* 50, 553–590.
- (55) Zhang, J., Wang, H., Luo, Z., Yang, Z., Zhang, Z., Wang, P., Li, M., Zhang, Y., Feng, Y., Lu, D., and Zhu, Y. (2023) Computational design of highly efficient thermostable MHET hydrolases and dual enzyme system for PET recycling. *Commun Biol* 6, 1–18.
- (56) Son, H. F., Cho, I. J., Joo, S., Seo, H., Sagong, H.-Y., Choi, S. Y., Lee, S. Y., and Kim, K.-J. (2019) Rational protein engineering of thermo-stable PETase from *Ideonella sakaiensis* for highly efficient PET degradation. *ACS Catal.* 9, 3519–3526.

- (57) Lu, H., Diaz, D. J., Czarnecki, N. J., Zhu, C., Kim, W., Shroff, R., Acosta, D. J., Alexander, B. R., Cole, H. O., Zhang, Y., Lynd, N. A., Ellington, A. D., and Alper, H. S. (2022) Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* 604, 662–667.
- (58) Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y., Mitra, R., Han, J., Li, C., Han, X., Liu, W., Chen, Q., Wei, W., Wang, X., Du, W., Tang, S., Xiang, H., Liu, H., Liang, Y., Houk, K. N., and Wu, B. (2021) Computational redesign of a PETase for plastic biodegradation under ambient condition by the GRAPE strategy. *ACS Catal.* 11, 1340–1350.
- (59) Bell, E. L., Smithson, R., Kilbride, S., Foster, J., Hardy, F. J., Ramachandran, S., Tedstone, A. A., Haigh, S. J., Garforth, A. A., Day, P. J. R., Levy, C., Shaver, M. P., and Green, A. P. (2022) Directed evolution of an efficient and thermostable PET depolymerase. *Nat Catal* 5, 673–681.
- (60) Shi, L., Liu, P., Tan, Z., Zhao, W., Gao, J., Gu, Q., Ma, H., Liu, H., and Zhu, L. (2023) Complete Depolymerization of PET Wastes by an Evolved PET Hydrolase from Directed Evolution. *Angewandte Chemie International Edition* 62, e202218390.
- (61) Cui, Y., Chen, Y., Sun, J., Zhu, T., Pang, H., Li, C., Geng, W.-C., and Wu, B. (2024) Computational redesign of a hydrolase for nearly complete PET depolymerization at industrially relevant high-solids loading. *Nat Commun* 15, 1417.
- (62) Erickson, E., Shakespeare, T. J., Bratti, F., Buss, B. L., Graham, R., Hawkins, M. A., König, G., Michener, W. E., Miscall, J., Ramirez, K. J., Rorrer, N. A., Zahn, M., Pickford, A. R., McGeehan, J. E., and Beckham, G. T. (2022) Comparative performance of PETase as a function of reaction conditions, substrate properties, and product accumulation. *ChemSusChem* 15, e202101932.
- (63) Zhang, Y., and Hess, H. (2017) Toward Rational Design of High-efficiency Enzyme Cascades. *ACS Catal.* 7, 6018–6027.
- (64) Knott, B. C., Erickson, E., Allen, M. D., Gado, J. E., Graham, R., Kearns, F. L., Pardo, I., Topuzlu, E., Anderson, J. J., Austin, H. P., Dominick, G., Johnson, C. W., Rorrer, N. A., Szostkiewicz, C. J., Copié, V., Payne, C. M., Woodcock, H. L., Donohoe, B. S., Beckham, G. T., and McGeehan, J. E. (2020) Characterization and engineering of a two-enzyme system for plastics depolymerization. *Proceedings of the National Academy of Sciences* 117, 25476–25485.
- (65) Seo, H., Kim, S., Son, H. F., Sagong, H.-Y., Joo, S., and Kim, K.-J. (2019) Production of extracellular PETase from *Ideonella sakaiensis* using sec-dependent signal peptides in *E. coli*. *Biochemical and Biophysical Research Communications* 508, 250–255.
- (66) Huang, X., Cao, L., Qin, Z., Li, S., Kong, W., and Liu, Y. (2018) Tat-Independent Secretion of Polyethylene Terephthalate Hydrolase PETase in *Bacillus subtilis* 168 Mediated by Its Native Signal Peptide. *J. Agric. Food Chem.* 66, 13217–13227.
- (67) Chen, Z., Wang, Y., Cheng, Y., Wang, X., Tong, S., Yang, H., and Wang, Z. (2020) Efficient biodegradation of highly crystallized polyethylene terephthalate through cell surface display of bacterial PETase. *Science of The Total Environment* 709, 136138.

- (68) Kim, J. W., Park, S.-B., Tran, Q.-G., Cho, D.-H., Choi, D.-Y., Lee, Y. J., and Kim, H.-S. (2020) Functional expression of polyethylene terephthalate-degrading enzyme (PETase) in green microalgae. *Microb Cell Fact* 19, 97.
- (69) Moog, D., Schmitt, J., Senger, J., Zarzycki, J., Rexer, K.-H., Linne, U., Erb, T., and Maier, U. G. (2019) Using a marine microalga as a chassis for polyethylene terephthalate (PET) degradation. *Microb Cell Fact* 18, 171.
- (70) Yan, F., Wei, R., Cui, Q., Bornscheuer, U. T., and Liu, Y.-J. (2021) Thermophilic whole-cell degradation of polyethylene terephthalate using engineered *Clostridium thermocellum*. *Microbial Biotechnology* 14, 374–385.
- (71) Arnal, G., Anglade, J., Gavalda, S., Tournier, V., Chabot, N., Bornscheuer, U. T., Weber, G., and Marty, A. (2023) Assessment of four engineered PET degrading enzymes considering large-scale industrial applications. *ACS Catal.* 13, 13156–13166.
- (72) Gui, J., Sun, Y., Wang, J., Chen, X., Zhang, S., and Wu, D. (2021) Microplastics in composting of rural domestic waste: abundance, characteristics, and release from the surface of macroplastics. *Environmental Pollution* 274, 116553.
- (73) Hanefeld, U., Gardossi, L., and Magner, E. (2009) Understanding enzyme immobilisation. *Chemical Society Reviews* 38, 453–468.
- (74) Gercke, D., Furtmann, C., Tozakidis, I. E. P., and Jose, J. (2021) Highly Crystalline Post-Consumer PET Waste Hydrolysis by Surface Displayed PETase Using a Bacterial Whole-Cell Biocatalyst. *ChemCatChem* 13, 3479–3489.
- (75) Han, W., Zhang, J., Chen, Q., Xie, Y., Zhang, M., Qu, J., Tan, Y., Diao, Y., Wang, Y., and Zhang, Y. (2024) Biodegradation of poly(ethylene terephthalate) through PETase surface-display: From function to structure. *Journal of Hazardous Materials* 461, 132632.
- (76) Chen, Z., Duan, R., Xiao, Y., Wei, Y., Zhang, H., Sun, X., Wang, S., Cheng, Y., Wang, X., Tong, S., Yao, Y., Zhu, C., Yang, H., Wang, Y., and Wang, Z. (2022) Biodegradation of highly crystallized poly(ethylene terephthalate) through cell surface codisplay of bacterial PETase and hydrophobin. *Nat Commun* 13, 7138.
- (77) Thomsen, T. B., Hunt, C. J., and Meyer, A. S. (2022) Influence of substrate crystallinity and glass transition temperature on enzymatic degradation of polyethylene terephthalate (PET). *New Biotechnology* 69, 28–35.
- (78) Zhu, B., Ye, Q., Seo, Y., and Wei, N. (2022) Enzymatic Degradation of Polyethylene Terephthalate Plastics by Bacterial Curli Display PETase. *Environ. Sci. Technol. Lett.* 9, 650–657.
- (79) Qi, X., Ma, Y., Chang, H., Li, B., Ding, M., and Yuan, Y. (2021) Evaluation of PET degradation using artificial microbial consortia. *Frontiers in Microbiology* 12.
- (80) Gao, R., and Sun, C. (2021) A marine bacterial community capable of degrading poly(ethylene terephthalate) and polyethylene. *Journal of Hazardous Materials* 416, 125928.
- (81) Roberts, C., Edwards, S., Vague, M., León-Zayas, R., Scheffer, H., Chan, G., Swartz, N. A., and Mellies, J. L. (2020) Environmental consortium containing pseudomonas and bacillus species synergistically degrades polyethylene terephthalate plastic. *mSphere* 5, e01151-20.

- (82) Bao, T., Qian, Y., Xin, Y., Collins, J. J., and Lu, T. (2023) Engineering microbial division of labor for plastic upcycling. *Nat Commun* 14, 5712.
- (83) Lederberg, J., and Tatum, E. L. (1946) Gene Recombination in *Escherichia Coli*. *Nature* 158, 558–558.
- (84) Abbas, A., Barkhouse, A., Hackenberger, D., and Wright, G. D. (2024) Antibiotic resistance: A key microbial survival mechanism that threatens public health. *Cell Host & Microbe* 32, 837–851.
- (85) Virolle, C., Goldlust, K., Djermoun, S., Bigot, S., and Lesterlin, C. (2020) Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. *Genes* 11, 1239.
- (86) Ronda, C., Chen, S. P., Cabral, V., Yaung, S. J., and Wang, H. H. (2019) Metagenomic engineering of the mammalian gut microbiome in situ. *Nat Methods* 16, 167–170.
- (87) Klümper, U., Riber, L., Dechesne, A., Sannazzarro, A., Hansen, L. H., Sørensen, S. J., and Smets, B. F. (2015) Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. *ISME J* 9, 934–945.
- (88) Pansegrau, W., Lanka, E., Barth, P. T., Figurski, D. H., Guiney, D. G., Haas, D., Helinski, D. R., Schwab, H., Stanisich, V. A., and Thomas, C. M. (1994) Complete Nucleotide Sequence of Birmingham IncPα Plasmids: Compilation and Comparative Analysis. *Journal of Molecular Biology* 239, 623–663.
- (89) Zatyka, M., and Thomas, C. M. (1998) Control of genes for conjugative transfer of plasmids and other mobile elements. *FEMS Microbiology Reviews* 21, 291–319.
- (90) Young, C., Prince, A. S., and Figurski, D. H. (1985) *korA* function of promiscuous plasmid RK2: an autorepressor that inhibits expression of host-lethal gene *kilA* and replication gene *trfA*. *Proceedings of the National Academy of Sciences* 82, 7374–7378.
- (91) Frost, L. S., and Koraimann, G. (2010) Regulation of Bacterial Conjugation: Balancing Opportunity With Adversity. *Future Microbiology* 5, 1057–1071.
- (92) Chiu, C.-M., Manzoor, S. E., Batt, S. M., Muntaha, S. tul, Bingle, L. E. H., and Thomas, C. M. (2008) Distribution of the partitioning protein *KorB* on the genome of IncP-1 plasmid RK2. *Plasmid* 59, 163–175.
- (93) Bingle, L. E. H., Zatyka, M., Manzoor, S. E., and Thomas, C. M. (2003) Co-operative interactions control conjugative transfer of broad host-range plasmid RK2: full effect of minor changes in *TrbA* operator depends on *KorB*. *Molecular Microbiology* 49, 1095–1108.
- (94) Miyakoshi, M., Ohtsubo, Y., Nagata, Y., and Tsuda, M. (2020) Transcriptome Analysis of Zygotic Induction During Conjugative Transfer of Plasmid RP4. *Front. Microbiol.* 11.
- (95) Thomson, V. J., Jovanovic, O. S., Pohlman, R. F., Chang, C. H., and Figurski, D. H. (1993) Structure, function, and regulation of the *kilB* locus of promiscuous plasmid RK2. *Journal of Bacteriology* 175, 2423–2435.
- (96) Simonsen, L. (1990) Dynamics of plasmid transfer on surfaces. *Microbiology* 136, 1001–1007.

- (97) Christensen, B. B., Sternberg, C., and Molin, S. (1996) Bacterial plasmid conjugation on semi-solid surfaces monitored with the green fluorescent protein (GFP) from *Aequorea victoria* as a marker. *Gene* 173, 59–65.
- (98) Seoane, J., Yankelevich, T., Dechesne, A., Merkey, B., Sternberg, C., and Smets, B. F. (2011) An individual-based approach to explain plasmid invasion in bacterial populations. *FEMS Microbiology Ecology* 75, 17–27.
- (99) Smets, B. F., Rittmann, B. E., and Stahl, D. A. (1993) The specific growth rate of *Pseudomonas putida* PAW1 influences the conjugal transfer rate of the TOL plasmid. *Applied and Environmental Microbiology* 59, 3430–3437.
- (100) Smets, B. f., Rittmann, B. e., and Stahl, D. a. (1995) Quantification of the effect of substrate concentration on the conjugal transfer rate of the TOL plasmid in short-term batch mating experiments. *Letters in Applied Microbiology* 21, 167–172.
- (101) Fox, R. E., Zhong, X., Krone, S. M., and Top, E. M. (2008) Spatial structure and nutrients promote invasion of IncP-1 plasmids in bacterial populations. *The ISME Journal* 2, 1024–1039.
- (102) Frost, L. S., and Manchak, J. (1998) F- phenocopies: characterization of expression of the F transfer region in stationary phase. *Microbiology* 144, 2579–2587.
- (103) Sysoeva, T. A., Kim, Y., Rodriguez, J., Lopatkin, A. J., and You, L. (2020) Growth-stage-dependent regulation of conjugation. *AIChE Journal* 66, e16848.
- (104) Starčič, M., Žgur-Bertok, D., Jordi, B. J. A. M., Wösten, M. M. S. M., Gaastra, W., and van Putten, J. P. M. (2003) The Cyclic AMP-Cyclic AMP Receptor Protein Complex Regulates Activity of the traJ Promoter of the Escherichia coli Conjugative Plasmid pRK100. *Journal of Bacteriology* 185, 1616–1623.
- (105) Roer, L., Aarestrup, F. M., and Hasman, H. (2014) The EcoKI Type I Restriction-Modification System in Escherichia coli Affects but Is Not an Absolute Barrier for Conjugation. *Journal of Bacteriology* 197, 337–342.
- (106) Butler, C. A., and Gotschlich, E. C. (1991) High-frequency mobilization of broad-host-range plasmids into Neisseria gonorrhoeae requires methylation in the donor. *Journal of Bacteriology* 173, 5793–5799.
- (107) Dimitriu, T., Szczelkun, M. D., and Westra, E. R. (2024) Various plasmid strategies limit the effect of bacterial restriction–modification systems against conjugation. *Nucleic Acids Research* 52, 12976–12986.
- (108) Shaw, L. P., Rocha, E. P. C., and MacLean, R. C. (2023) Restriction-modification systems have shaped the evolution and distribution of plasmids across bacteria. *Nucleic Acids Research* 51, 6806–6818.
- (109) Johnston, C. D., Cotton, S. L., Rittling, S. R., Starr, J. R., Borisy, G. G., Dewhirst, F. E., and Lemon, K. P. (2019) Systematic evasion of the restriction-modification barrier in bacteria. *Proceedings of the National Academy of Sciences* 116, 11454–11459.
- (110) Yasui, K., Kano, Y., Tanaka, K., Watanabe, K., Shimizu-Kadota, M., Yoshikawa, H., and Suzuki, T. (2009) Improvement of bacterial transformation efficiency using plasmid artificial modification. *Nucleic Acids Research* 37, e3.

- (111) Alderliesten, J. B., Duxbury, S. J. N., Zwart, M. P., de Visser, J. A. G. M., Stegeman, A., and Fischer, E. A. J. (2020) Effect of donor-recipient relatedness on the plasmid conjugation frequency: a meta-analysis. *BMC Microbiol* 20, 135.
- (112) Sanderson, K. E., Janzer, J., and Head, J. (1981) Influence of lipopolysaccharide and protein in the cell envelope on recipient capacity in conjugation of *Salmonella typhimurium*. *Journal of Bacteriology* 148, 283–293.
- (113) Pérez-Mendoza, D., and de la Cruz, F. (2009) *Escherichia coli* genes affecting recipient ability in plasmid conjugation: Are there any? *BMC Genomics* 10, 71.
- (114) Bradley, D. E., Taylor, D. E., and Cohen, D. R. (1980) Specification of surface mating systems among conjugative drug resistance plasmids in *Escherichia coli* K-12. *Journal of Bacteriology* 143, 1466–1470.
- (115) Ishiwa, A., and Komano, T. (2000) The lipopolysaccharide of recipient cells is a specific receptor for PilV proteins, selected by shufflon DNA rearrangement, in liquid matings with donors bearing the R64 plasmid. *Mol Gen Genet* 263, 159–164.
- (116) Craig, L., Forest, K. T., and Maier, B. (2019) Type IV pili: dynamics, biophysics and functional consequences. *Nat Rev Microbiol* 17, 429–440.
- (117) Khalil, T. A., and Gealt, M. A. (1987) Temperature, pH, and cations affect the ability of *Escherichia coli* to mobilize plasmids in L broth and synthetic wastewater. *Can. J. Microbiol.* 33, 733–737.
- (118) Fernandez-Astorga, A., Muela, A., Cisterna, R., Iriberry, J., and Barcina, I. (1992) Biotic and abiotic factors affecting plasmid transfer in *Escherichia coli* strains. *Applied and Environmental Microbiology* 58, 392–398.
- (119) Johnsen, A. R., and Kroer, N. (2007) Effects of stress and other environmental factors on horizontal plasmid transfer assessed by direct quantification of discrete transfer events. *FEMS Microbiology Ecology* 59, 718–728.
- (120) Huisman, J. S., Benz, F., Duxbury, S. J. N., de Visser, J. A. G. M., Hall, A. R., Fischer, E. A. J., and Bonhoeffer, S. (2022) Estimating plasmid conjugation rates: A new computational tool and a critical comparison of methods. *Plasmid* 121, 102627.
- (121) Levin, B. R., Stewart, F. M., and Rice, V. A. (1979) The kinetics of conjugative plasmid transmission: Fit of a simple mass action model. *Plasmid* 2, 247–260.
- (122) Simonsen, L., Gordon, D. M., Stewart, F. M., and Levin, B. R. (1990) Estimating the rate of plasmid transfer: an end-point method. *Microbiology* 136, 2319–2325.
- (123) Fernandez-Lopez, R., Campo, I. del, Revilla, C., Cuevas, A., and Cruz, F. de la. (2014) Negative Feedback and Transcriptional Overshooting in a Regulatory Network for Horizontal Gene Transfer. *PLOS Genetics* 10, e1004171.
- (124) Kosterlitz, O., Tirado, A. M., Wate, C., Elg, C., Bozic, I., Top, E. M., and Kerr, B. (2022) Estimating the transfer rates of bacterial plasmids with an adapted Luria–Delbrück fluctuation analysis. *PLOS Biology* 20, e3001732.
- (125) Kosterlitz, O., and Huisman, J. S. (2023) Guidelines for the estimation and reporting of plasmid conjugation rates. *Plasmid* 126, 102685.
- (126) Malwade, A., Nguyen, A., Sadat-Mousavi, P., and Ingalls, B. P. (2017) Predictive modeling of a batch filter mating process. *Frontiers in Microbiology* 8.

- (127) Sørensen, S. J., Bailey, M., Hansen, L. H., Kroer, N., and Wuertz, S. (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol* 3, 700–710.
- (128) Lopatkin, A. J., Huang, S., Smith, R. P., Srimani, J. K., Sysoeva, T. A., Bewick, S., Karig, D. K., and You, L. (2016) Antibiotics as a selective driver for conjugation dynamics. *Nat Microbiol* 1, 1–8.
- (129) Li, B., Qiu, Y., Song, Y., Lin, H., and Yin, H. (2019) Dissecting horizontal and vertical gene transfer of antibiotic resistance plasmid in bacterial community using microfluidics. *Environment International* 131, 105007.
- (130) Liu, L., Zhang, Q., and Li, R. (2023) In Situ and Individual-Based Analysis of the Influence of Polystyrene Microplastics on Escherichia coli Conjugative Gene Transfer at the Single-Cell Level. *Environ. Sci. Technol.* 57, 15936–15944.
- (131) Lawley, T. D., Gordon, G. S., Wright, A., and Taylor, D. E. (2002) Bacterial conjugative transfer: visualization of successful mating pairs and plasmid establishment in live Escherichia coli. *Molecular Microbiology* 44, 947–956.
- (132) Carranza, G., Menguiano, T., Valenzuela-Gómez, F., García-Cazorla, Y., Cabezón, E., and Arechaga, I. (2021) Monitoring Bacterial Conjugation by Optical Microscopy. *Front. Microbiol.* 12.
- (133) Couturier, A., Virolle, C., Goldlust, K., Berne-Dedieu, A., Reuter, A., Nolivos, S., Yamaichi, Y., Bigot, S., and Lesterlin, C. (2023) Real-time visualisation of the intracellular dynamics of conjugative plasmid transfer. *Nat Commun* 14, 294.
- (134) Clarke, M., Maddera, L., Harris, R. L., and Silverman, P. M. (2008) F-pili dynamics by live-cell imaging. *Proceedings of the National Academy of Sciences* 105, 17978–17981.
- (135) Nagarajan, K., Ni, C., and Lu, T. (2022) Agent-Based Modeling of Microbial Communities. *ACS Synth. Biol.* 11, 3564–3574.
- (136) Railsback, S. F., and Grimm, V. (2019) Agent-Based and Individual-Based Modeling: A Practical Introduction 2nd ed. Princeton University Press.
- (137) Gorochofski, T. E. (2016) Agent-based modelling in synthetic biology. *Essays in Biochemistry* (Pinheiro, V. B., Ed.) 60, 325–336.
- (138) Matyjaszkiewicz, A., Fiore, G., Annunziata, F., Grierson, C. S., Savery, N. J., and others. (2017) BSim 2.0: An Advanced Agent-Based Cell Simulator. *ACS Synthetic Biology* 6, 1969–1972.
- (139) Naylor, J., Fellermann, H., Ding, Y., Mohammed, W. K., Jakubovics, N. S., Mukherjee, J., Biggs, C. A., Wright, P. C., and Krasnogor, N. (2017) Simbiotics: A Multiscale Integrative Platform for 3D Modeling of Bacterial Populations. *ACS Synth. Biol.* 6, 1194–1210.
- (140) Lardon, L. A., Merkey, B. V., Martins, S., Dötsch, A., Picioreanu, C., Kreft, J.-U., and Smets, B. F. (2011) iDynoMiCS: next-generation individual-based modelling of biofilms. *Environmental Microbiology* 13, 2416–2434.
- (141) Rudge, T. J., Steiner, P. J., Phillips, A., and Haseloff, J. (2012) Computational Modeling of Synthetic Microbial Biofilms. *ACS Synth. Biol.* 1, 345–352.
- (142) Goñi-Moreno, A., Amos, M., and Cruz, F. de la. (2013) Multicellular Computing Using Conjugation for Wiring. *PLOS ONE* 8, e65986.

- (143) Goñi-Moreno, A., and Amos, M. (2015) DiSCUS: A Simulation Platform for Conjugation Computing, in *Unconventional Computation and Natural Computation* (Calude, C. S., and Dinneen, M. J., Eds.), pp 181–191. Springer International Publishing, Cham.
- (144) Gutiérrez, M., Gregorio-Godoy, P., Pérez del Pulgar, G., Muñoz, L. E., Sáez, S., and Rodríguez-Patón, A. (2017) A New Improved and Extended Version of the Multicell Bacterial Simulator gro. *ACS Synth. Biol.* 6, 1496–1508.
- (145) Li, B., Taniguchi, D., Gedara, J. P., Gogulancea, V., Gonzalez-Cabaleiro, R., Chen, J., McGough, A. S., Ofiteru, I. D., Curtis, T. P., and Zuliani, P. (2019) NUFEB: A massively parallel simulator for individual-based modelling of microbial communities. *PLOS Computational Biology* 15, e1007125.
- (146) Agmon, E., Spangler, R. K., Skalnik, C. J., Poole, W., Peirce, S. M., Morrison, J. H., and Covert, M. W. (2022) Vivarium: an interface and engine for integrative multiscale modeling in computational biology. *Bioinformatics* 38, 1972–1979.
- (147) Granato, E. T., Smith, W. P. J., and Foster, K. R. (2023) Collective protection against the type VI secretion system in bacteria. *The ISME Journal* 17, 1052–1062.
- (148) Smith, W. P. J., Brodmann, M., Unterweger, D., Davit, Y., Comstock, L. E., Basler, M., and Foster, K. R. (2020) The evolution of tit-for-tat in bacteria via the type VI secretion system. *Nat Commun* 11, 5395.
- (149) Smith, W. P. J., Vettiger, A., Winter, J., Ryser, T., Comstock, L. E., and others. (2020) The evolution of the type VI secretion system as a disintegration weapon. *PLOS Biology* 18, 1–26.
- (150) Bottery, M. J., Passaris, I., Dytham, C., Wood, A. J., and Woude, M. W. van der. (2019) Spatial Organization of Expanding Bacterial Colonies Is Affected by Contact-Dependent Growth Inhibition. *Current Biology* 29, 3622–3634.e5.
- (151) Rudzite, M., Subramoni, S., Endres, R. G., and Filloux, A. (2023) Effectiveness of *Pseudomonas aeruginosa* type VI secretion system relies on toxin potency and type IV pili-dependent interaction. *PLOS Pathogens* 19, e1011428.
- (152) Smith, W. P. J., Davit, Y., Osborne, J. M., Kim, W., Foster, K. R., and Pitt-Francis, J. M. (2017) Cell morphology drives spatial patterning in microbial communities. *Proceedings of the National Academy of Sciences* 114, E280–E286.
- (153) Rudge, T. J., Federici, F., Steiner, P. J., Kan, A., and Haseloff, J. (2013) Cell Polarity-Driven Instability Generates Self-Organized, Fractal Patterning of Cell Layers. *ACS Synth. Biol.* 2, 705–714.
- (154) Frost, I., Smith, W. P. J., Mitri, S., Millan, A. S., Davit, Y., Osborne, J. M., Pitt-Francis, J. M., MacLean, R. C., and Foster, K. R. (2018) Cooperation, competition and antibiotic resistance in bacterial colonies. *The ISME Journal* 12, 1582–1593.
- (155) Ruan, C., Ramoneda, J., Kan, A., Rudge, T. J., Wang, G., and Johnson, D. R. (2024) Phage predation accelerates the spread of plasmid-encoded antibiotic resistance. *Nat Commun* 15, 5397.
- (156) Ma, Y., Ramoneda, J., and Johnson, D. R. (2023) Timing of antibiotic administration determines the spread of plasmid-encoded antibiotic resistance during microbial range expansion. *Nat Commun* 14, 3530.

- (157) Ruan, C., Borer, B., Ramoneda, J., Wang, G., and Johnson, D. R. (2023) Evaporation-induced hydrodynamics control plasmid transfer during surface-associated microbial growth. *npj Biofilms Microbiomes* 9, 1–10.
- (158) Ma, Y., Kan, A., and Johnson, D. R. (2024) Metabolic interactions control the transfer and spread of plasmid-encoded antibiotic resistance during surface-associated microbial growth. *Cell Reports* 43.
- (159) Kan, A., Del Valle, I., Rudge, T., Federici, F., and Haseloff, J. (2018) Intercellular adhesion promotes clonal mixing in growing bacterial populations. *Journal of The Royal Society Interface* 15, 20180406.
- (160) Yáñez Feliú, G., Vidal, G., Muñoz Silva, M., and Rudge, T. J. (2020) Novel tunable spatio-temporal patterns from a simple genetic oscillator circuit. *Frontiers in Bioengineering and Biotechnology* 8, 893.
- (161) Nuñez, I. N., Matute, T. F., Del Valle, I. D., Kan, A., Choksi, A., Endy, D., Haseloff, J., Rudge, T. J., and Federici, F. (2017) Artificial Symmetry-Breaking for Morphogenetic Engineering Bacterial Colonies. *ACS Synth. Biol.* 6, 256–265.
- (162) Davison, J. (1999) Genetic Exchange between Bacteria in the Environment. *Plasmid* 42, 73–91.
- (163) Lawrence, J. G. (1999) Gene transfer, speciation, and the evolution of bacterial genomes. *Current Opinion in Microbiology* 2, 519–523.
- (164) Lermينياux, N. A., and Cameron, A. D. S. (2019) Horizontal transfer of antibiotic resistance genes in clinical environments. *Can. J. Microbiol.* 65, 34–44.
- (165) Graf, F. E., Palm, M., Warringer, J., and Farewell, A. (2019) Inhibiting conjugation as a tool in the fight against antibiotic resistance. *Drug Development Research* 80, 19–23.
- (166) Hamilton, T. A., Pellegrino, G. M., Therrien, J. A., Ham, D. T., Bartlett, P. C., Karas, B. J., Gloor, G. B., and Edgell, D. R. (2019) Efficient inter-species conjugative transfer of a CRISPR nuclease for targeted bacterial killing. *Nat Commun* 10, 4544.
- (167) Neil, K., Allard, N., Roy, P., Grenier, F., Menendez, A., Burrus, V., and Rodrigue, S. (2021) High-efficiency delivery of CRISPR-Cas9 by engineered probiotics enables precise microbiome editing. *Molecular Systems Biology* 17, e10335.
- (168) del Campo, I., Ruiz, R., Cuevas, A., Revilla, C., Vielva, L., and de la Cruz, F. (2012) Determination of conjugation rates on solid surfaces. *Plasmid* 67, 174–182.
- (169) Hausner, M., and Wuertz, S. (1999) High Rates of Conjugation in Bacterial Biofilms as Determined by Quantitative In Situ Analysis. *Applied and Environmental Microbiology* 65, 3710–3713.
- (170) Merkey, B. V., Lardon, L. A., Seoane, J. M., Kreft, J.-U., and Smets, B. F. (2011) Growth dependence of conjugation explains limited plasmid invasion in biofilms: an individual-based modelling study. *Environmental Microbiology* 13, 2435–2452.
- (171) Babić, A., Lindner, A. B., Vulić, M., Stewart, E. J., and Radman, M. (2008) Direct Visualization of Horizontal Gene Transfer. *Science* 319, 1533–1536.
- (172) Goldlust, K., Ducret, A., Halte, M., Dedieu-Berne, A., Erhardt, M., and Lesterlin, C. (2023) The F pilus serves as a conduit for the DNA during conjugation between physically distant bacteria. *Proceedings of the National Academy of Sciences* 120, e2310842120.

- (173) García, A. P., and Rodríguez-Patón, A. (2015) BactoSim – An Individual-Based Simulation Environment for Bacterial Conjugation, in *Advances in Practical Applications of Agents, Multi-Agent Systems, and Sustainability: The PAAMS Collection* (Demazeau, Y., Decker, K. S., Bajo Pérez, J., and de la Prieta, F., Eds.), pp 275–279. Springer International Publishing, Cham.
- (174) Krone, S. M., Lu, R., Fox, R., Suzuki, H., and Top, E. M. (2007) Modelling the spatial dynamics of plasmid transfer and persistence. *Microbiology* 153, 2803–2816.
- (175) Ahmadi, A., Yip, A., Chalaturnyk, J., Aucoin, M. G., and Ingalls, B. P. (2024) A Pipeline for Calibrating Agent-Based Models of Microbial Populations: From Image Collection to Model Parameterization. *IFAC-PapersOnLine* 58, 1–6.
- (176) Klinger, E., Rickert, D., and Hasenauer, J. (2018) pyABC: distributed, likelihood-free inference. *Bioinformatics* 34, 3591–3593.
- (177) Alnahhas, R. N., Winkle, J. J., Hirning, A. J., Karamched, B., Ott, W., Josić, K., and Bennett, M. R. (2019) Spatiotemporal Dynamics of Synthetic Microbial Consortia in Microfluidic Devices. *ACS Synth. Biol.* 8, 2051–2058.
- (178) Alnahhas, R. N., Sadeghpour, M., Chen, Y., Frey, A. A., Ott, W., Josić, K., and Bennett, M. R. (2020) Majority sensing in synthetic microbial consortia. *Nat Commun* 11, 3659.
- (179) Dal Co, A., van Vliet, S., Kiviet, D. J., Schlegel, S., and Ackermann, M. (2020) Short-range interactions govern the dynamics and functions of microbial communities. *Nat Ecol Evol* 4, 366–375.
- (180) van Gestel, J., Bareia, T., Tenennbaum, B., Dal Co, A., Guler, P., Aframian, N., Puyesky, S., Grinberg, I., D’Souza, G. G., Erez, Z., Ackermann, M., and Eldar, A. (2021) Short-range quorum sensing controls horizontal gene transfer at micron scale in bacterial communities. *Nat Commun* 12, 2324.
- (181) Yip, A., McArthur, O. D., Ho, K. C., Aucoin, M. G., and Ingalls, B. P. (2024) Degradation of polyethylene terephthalate (PET) plastics by wastewater bacteria engineered via conjugation. *Microbial Biotechnology* 17, e70015.
- (182) Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., and Tsien, R. Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* 22, 1567–1572.
- (183) Balleza, E., Kim, J. M., and Cluzel, P. (2018) Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* 15, 47–51.
- (184) Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A., and Kreshuk, A. (2019) ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* 16, 1226–1232.
- (185) Riitters, K. H., O’Neill, R. V., Wickham, J. D., and Jones, K. B. (1996) A note on contagion indices for landscape analysis. *Landscape Ecol* 11, 197–202.
- (186) Shen, Y., Chen, Y., Wu, J., Shaner, N. C., and Campbell, R. E. (2017) Engineering of mCherry variants with long Stokes shift, red-shifted fluorescence, and low cytotoxicity. *PLOS ONE* 12, e0171257.
- (187) Westerwalbesloh, C., Grünberger, A., Wiechert, W., Kohlheyer, D., and von Lieres, E. (2017) Coarse-graining bacteria colonies for modelling critical solute distributions in

- picolitre bioreactors for bacterial studies on single-cell level. *Microbial Biotechnology* 10, 845–857.
- (188) Burmeister, A., Hilgers, F., Langner, A., Westerwalbesloh, C., Kerkhoff, Y., Tenhaef, N., Drepper, T., Kohlheyer, D., Lieres, E. von, Noack, S., and Grünberger, A. (2018) A microfluidic co-cultivation platform to investigate microbial interactions at defined microenvironments. *Lab Chip* 19, 98–110.
 - (189) Massoudieh, A., Mathew, A., Lambertini, E., Nelson, K. E., and Ginn, T. R. (2007) Horizontal Gene Transfer on Surfaces in Natural Porous Media: Conjugation and Kinetics. *Vadose Zone Journal* 6, 306–315.
 - (190) Lundquist, P. D., and Levin, B. R. (1986) Transitory derepression and the maintenance of conjugative plasmids. *Genetics* 113, 483–497.
 - (191) Zhong, X., Droesch, J., Fox, R., Top, E. M., and Krone, S. M. (2012) On the meaning and estimation of plasmid transfer rates for surface-associated and well-mixed bacterial populations. *Journal of Theoretical Biology* 294, 144–152.
 - (192) Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R., and Sonnenburg, J. L. (2018) An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* 557, 434–438.
 - (193) Piñero-Lambea, C., Bodelón, G., Fernández-Periáñez, R., Cuesta, A. M., Álvarez-Vallina, L., and Fernández, L. Á. (2015) Programming Controlled Adhesion of *E. coli* to Target Surfaces, Cells, and Tumors with Synthetic Adhesins. *ACS Synth. Biol.* 4, 463–473.
 - (194) Mays, Z. J. S., Chappell, T. C., and Nair, N. U. (2020) Quantifying and Engineering Mucus Adhesion of Probiotics. *ACS Synth. Biol.* 9, 356–367.
 - (195) McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L., and Foster, K. R. (2016) Host Selection of Microbiota via Differential Adhesion. *Cell Host & Microbe* 19, 550–559.
 - (196) Momeni, B., Brileya, K. A., Fields, M. W., and Shou, W. (2013) Strong inter-population cooperation leads to partner intermixing in microbial communities. *eLife* (Tautz, D., Ed.) 2, e00230.
 - (197) Kong, W., Meldgin, D. R., Collins, J. J., and Lu, T. (2018) Designing microbial consortia with defined social interactions. *Nat Chem Biol* 14, 821–829.
 - (198) Nielsen, H. J., Ottesen, J. R., Youngren, B., Austin, S. J., and Hansen, F. G. (2006) The *Escherichia coli* chromosome is organized with the left and right chromosome arms in separate cell halves. *Molecular Microbiology* 62, 331–338.
 - (199) Ahmadi, A., Courtney, M., Ren, C., and Ingalls, B. (2024) A benchmarked comparison of software packages for time-lapse image processing of monolayer bacterial population dynamics. *Microbiology Spectrum* 12, e00032-24.
 - (200) Cutler, K. J., Stringer, C., Lo, T. W., Rappez, L., Stroustrup, N., Brook Peterson, S., Wiggins, P. A., and Mougous, J. D. (2022) Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. *Nat Methods* 19, 1438–1448.
 - (201) Stirling, D. R., Swain-Bowden, M. J., Lucas, A. M., Carpenter, A. E., Cimini, B. A., and Goodman, A. (2021) CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* 22, 433.

- (202) Hagberg, A., Swart, P. J., and Schult, D. A. (2008) Exploring network structure, dynamics, and function using NetworkX.
- (203) Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17, 261–272.
- (204) Volfson, D., Cookson, S., Hasty, J., and Tsimring, L. S. (2008) Biomechanical ordering of dense cell populations. *Proceedings of the National Academy of Sciences* 105, 15346–15351.
- (205) Eisenbrandt, R., Kalkum, M., Lai, E.-M., Lurz, R., Kado, C. I., and Lanka, E. (1999) Conjugative Pili of IncP Plasmids, and the Ti Plasmid T Pilus Are Composed of Cyclic Subunits *. *Journal of Biological Chemistry* 274, 22548–22555.
- (206) Fronzes, R., Christie, P. J., and Waksman, G. (2009) The structural biology of type IV secretion systems. *Nat Rev Microbiol* 7, 703–714.
- (207) Marsh, J. W., Kirk, C., and Ley, R. E. (2023) Toward Microbiome Engineering: Expanding the Repertoire of Genetically Tractable Members of the Human Gut Microbiome. *Annual Review of Microbiology* 77, 427–449.
- (208) Zurier, H. S., and Goddard, J. M. (2023) PETase Engineering for Enhanced Degradation of Microplastic Fibers in Simulated Wastewater Sludge Processing Conditions. *ACS EST Water* 3, 2210–2218.
- (209) Kiel, G. R., Lundberg, D. J., Prince, E., Husted, K. E. L., Johnson, A. M., Lensch, V., Li, S., Shieh, P., and Johnson, J. A. (2022) Cleavable Comonomers for Chemically Recyclable Polystyrene: A General Approach to Vinyl Polymer Circularity. *J. Am. Chem. Soc.* 144, 12979–12988.
- (210) Amrofell, M. B., Rengarajan, S., Vo, S. T., Tovar, E. S. R., LoBello, L., Dantas, G., and Moon, T. S. (2023) Engineering E. coli strains using antibiotic-resistance-gene-free plasmids. *Cell Reports Methods* 3.
- (211) Kommineni, S., Bretl, D. J., Lam, V., Chakraborty, R., Hayward, M., Simpson, P., Cao, Y., Bousounis, P., Kristich, C. J., and Salzman, N. H. (2015) Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature* 526, 719–722.
- (212) Nelson, M. T., Charbonneau, M. R., Coia, H. G., Castillo, M. J., Holt, C., Greenwood, E. S., Robinson, P. J., Merrill, E. A., Lubkowicz, D., and Mauzy, C. A. (2021) Characterization of an engineered live bacterial therapeutic for the treatment of phenylketonuria in a human gut-on-a-chip. *Nat Commun* 12, 2805.

Appendix A

Supplementary Data for Chapter 3

Table A.1. Primers used in Chapter 3.

Primer	Sequence (5'-3')	Notes
AY-7	GAAATACTAGATGGTGAGCAAG GGCGAG	Gibson assembly primer to amplify mCherry from pBT1-proD-mCherry
AY-8	ATCAGCGTTATTACTTGTACAGT TCGTCCATGC	Gibson assembly primer to amplify mCherry from pBT1-proD-mCherry
AY-9	GTACAAGTAATAACGCTGATAG TGCTAG	Gibson assembly primer to amplify pNuc-trans backbone with pBs promoter
AY-10	TGCTCACCATCTAGTATTTCTCC TCTTTCTC	Gibson assembly primer to amplify pNuc-trans backbone with pBs promoter
DE-3424	GAATTTCTGCCATTCATCCGCTT ATTATCACTTATTCAGGCGTAG CACCAGGCGTTTAACGATCGTC TTGCCTTGCTCGT	Gibson assembly primer to amplify - trans conjugating plasmid with overlaps to pTA-Mob
DE-3425	GCGTCCTGCTCGTGATCGGGAG TATCTGGCTGGGCCAACGTTCC AACCGCACTCCTAGTCAAGTCC AGACTCCTGTGTAA	Gibson assembly primer to amplify - trans conjugating plasmid with overlaps to pTA-Mob
OM-1	ACTGTAGACACCATACAGAAAC AGAGGAG	Gibson assembly primer to amplify FAST-PETase from pBTK522::FAST-PETase without promoter
OM-2	CACATTTCTCCGATTACGGATTC AGTGG	Gibson assembly primer to amplify FAST-PETase from pBTK522::FAST-PETase without promoter
OM-5	TGGTAACTGTCCTCCTTGACACT GAATTTAG	Gibson assembly primer to amplify pNuc-trans-mCherry backbone
OM-6	TTCTGTATGGTGTCTACAGTGCT AGCCC	Gibson assembly primer to amplify pNuc-trans-mCherry backbone including pBAD promoter
OM-8	TCCGTAATCGGAGAAATGTGCG CGGAAC	Gibson assembly primer to amplify bla-TEM1 from pFA6a-link-yoEGFP- SpHis5

OM-9	GTCAAGGAGGACAGTTACCAAT GCTTAATCAG	Gibson assembly primer to amplify bla-TEM1 from pFA6a-link-yoEGFP- SpHis6
------	--------------------------------------	---

Appendix B

Supplementary Data for Chapter 6

Table B.1. Accuracy for detection of transconjugants with Ilastik.

Trap #	Frame	Transconjugants Counted (by human)	Transconjugants Missed (by Ilastik)	False Positives
3	30	9	0	0
3	60	77	1	3
3	120	66	3	1
5	30	19	1	0
5	60	38	2	2
5	90	35	0	1
12	30	39	0	4
12	60	135	4	5
14	30	31	0	1

Table B.2. Manual measurements of fluorescence intensity for recipients (*E. coli* DH5 α pEB1-gfpmut2) in trap # 12.

Cell	Frame	Mean GFP	s.d.	Mean mCherry	s.d.
		Intensity (A.U.)		Intensity (A.U.)	
1	1	1341	271	77	71
2	1	1367	221	202	110
3	1	1375	335	8	8
4	1	1107	159	61	132
5	1	723	85	54	63
6	1	1145	296	15	16
7	1	1849	417	17	25
8	1	1136	189	255	297
9	1	1198	210	9	10
10	1	1482	397	181	193
11	80	839	179	42	62
12	80	1714	320	109	186
13	80	1024	154	31	67
14	80	777	189	61	66
15	80	1325	283	757	592
16	80	825	159	223	150
17	80	1308	444	396	222
18	80	1168	256	29	38
19	80	929	199	71	112
20	80	1505	361	239	266
21	80	605	147	11	9
22	80	403	109	26	20
23	80	366	93	33	26

Table B.3. Manual measurements of fluorescence intensity for donors (E. coli NEB10- β pFAST-PETase-cis) in trap # 12.

Cell	Frame	Mean GFP Intensity (A.U.)	s.d.	Mean mCherry Intensity (A.U.)	s.d.
1	1	3	5	2509	476
2	1	4	6	2402	564
3	1	8	9	3493	702
4	1	9	10	2628	663
5	1	0	1	2200	340
6	1	0	0	2258	324
7	1	97	74	2173	562
8	1	12	13	2854	632
9	1	16	24	5023	1674
10	1	68	78	3872	1036
11	80	32	24	1317	289
12	80	9	9	1214	226
13	80	75	107	2287	399
14	80	37	51	1222	166
15	80	53	55	881	130
16	80	16	18	1695	366
17	80	52	32	644	158
18	80	23	29	1866	309
19	80	8	9	809	227
20	80	53	47	1504	366

Table B.4. Manual classification of cell growth state after receiving the conjugative plasmid.

Trap	Cell ID (Frame.cell_id)	Centroid (pixels)	cellType	Stopped, ground truth (1 = Y, 0 = N)	Stopped, algorithm (1 = Y, 0 = N)
1	44.419	[397, 43]	trans	1	1
	53.420	[310, 431]	trans	1	1
	46.417	[340, 454]	trans	1	1
	36.418	[334, 403]	recip	1	0
5	26.409	[73, 563]	trans	0	0
	43.1140	[51, 568]	trans	0	0
	43.1139	[79, 570]	trans	0	0
	31.374	[60, 559]	trans	0	0
	43.694	[46, 550]	trans	0	0
	39.697	[65, 544]	trans	0	0
	50.1283	[123, 435]	trans	0	0
	56.778	[101, 479]	trans	0	1
	52.788	[56, 504]	trans	0	0
17	34.164	[137, 457]	trans	1	1
	38.379	[139, 518]	trans	0	0
	46.376	[152, 508]	trans	1	1

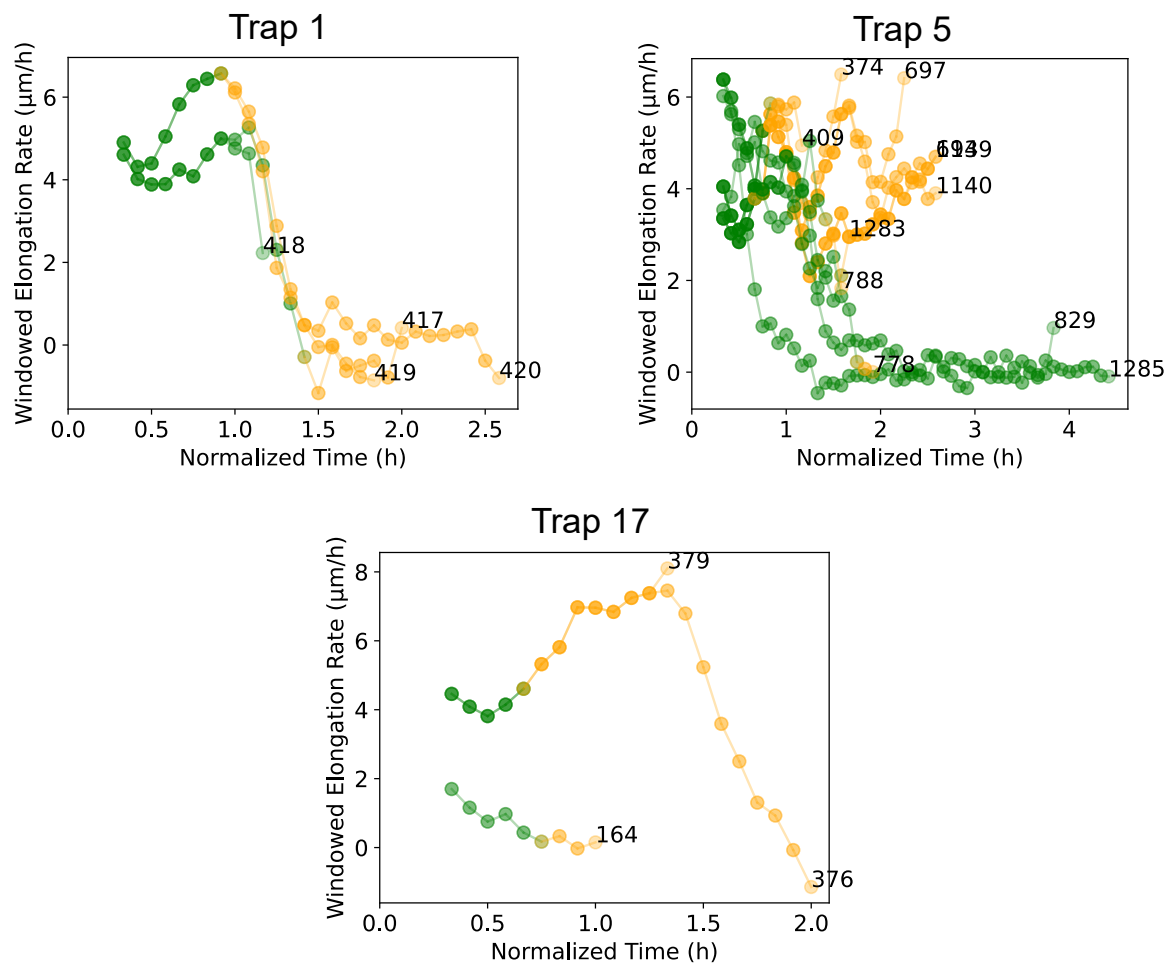


Figure B.1. Window-averaged elongation rates for select cells after receiving the conjugative plasmid. Green dots indicate the cell was classified as a recipient; yellow dots indicate classification as a transconjugant. Number annotations indicate cell ID at the end of the lineage.

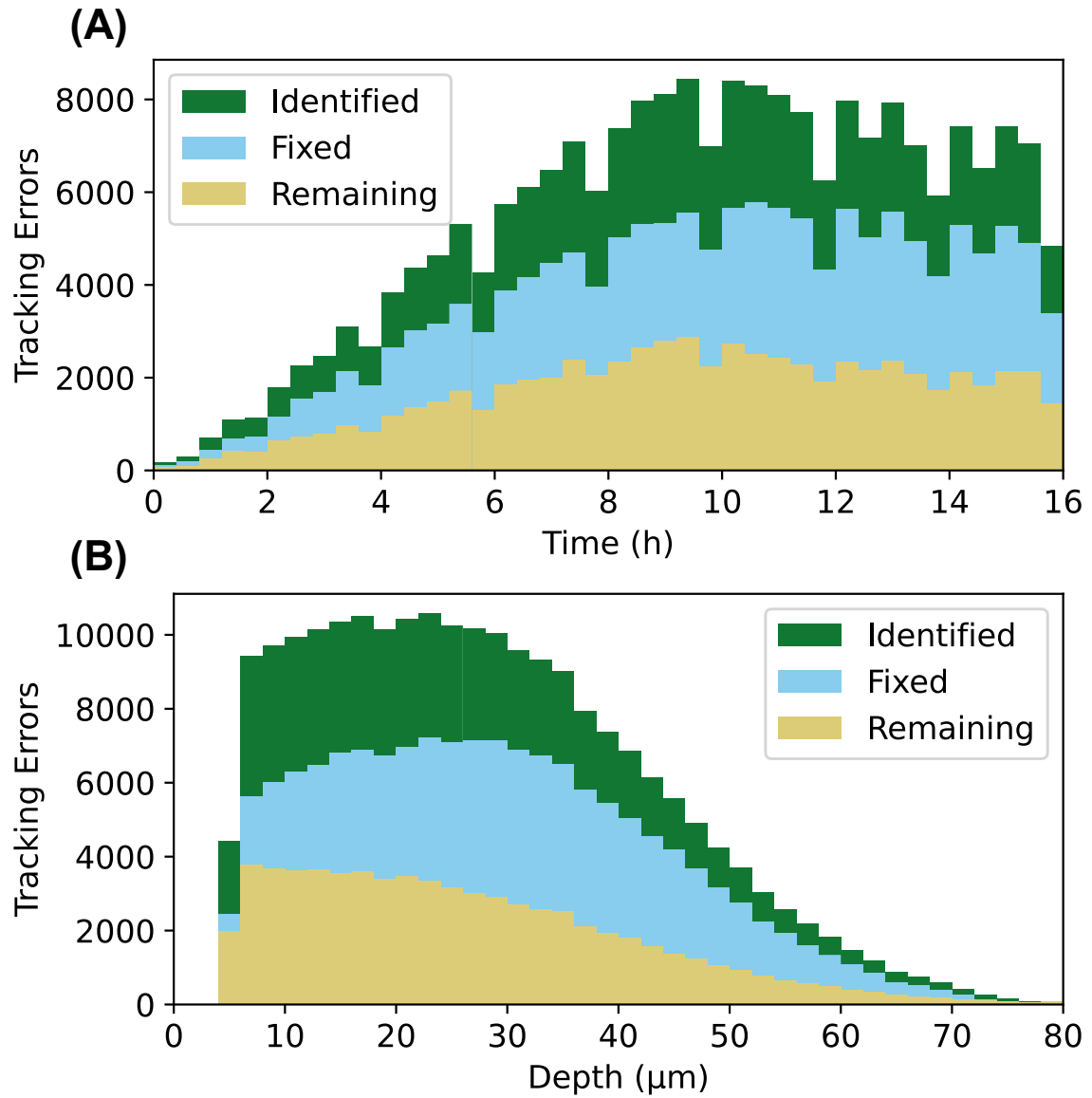


Figure B.2. Trackrefiner error tracking logs for pFAST-PETase-cis mating experiment. (A) Number of tracking errors that occurred at each time point. (B) Number of tracking errors that occurred at each depth in microfluidic traps. Errors are summed over all traps ($n = 17$). Any tracking errors that occurred within ~ 1 cell length ($5 \mu\text{m}$) of the trap exit were filtered out of the tracking error logs outputted by Trackrefiner.

Appendix C

Copyright Permission for Published Chapters

The following copyright statement has been directly from the published article in Chapter 3:

“This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes. © 2024 The Author(s).”

CC BY 4.0 License: <https://creativecommons.org/licenses/by/4.0/>

The following copyright statement has been directly from the published article in Chapter 4:

“**Copyright:** © 2022 Yip et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.”

CC BY-NC 4.0 License: <https://creativecommons.org/licenses/by-nc/4.0/>

The following copyright statement has been directly from the published article in Chapter 5:

“Copyright © The Authors. This is an open access article under the CC BY-NC-ND license.”

CC BY-NC-ND: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

Appendix D

Description of Supplementary Files

Chapter 3:

- Supplementary Data 3.1: bITS sequencing data for selected isolates.
- Supplementary Data 3.2: bITS sequencing data for all pooled isolates.
- Supplementary Data 3.4: Sequence for pFAST-PETase-cis.

Chapter 6:

- Supplementary Video 6.1: Time-lapse for microfluidic mating experiment performed with pFAST-PETase-cis.
- Supplementary Video 6.2: Time-lapse for microfluidic mating experiment performed with pFAST-PETase-cis (cells expressing GFP are false-coloured cyan).
- Supplementary Data 6.1: Trained Ilastik module for pixel classification of transconjugants.
- Supplementary Data 6.2: Trained Ilastik module for object classification of transconjugants.
- Supplementary Data 6.3: CellProfiler pipeline for image processing and cell tracking.

Appendix E

CellModeller Biophysical Model Description

The model description follows the ODD (Overview, Design concepts, Details) protocol for describing individual- and agent-based models defined in the work of Grimm and colleagues:

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., and others (2006) A standard protocol for describing individual-based and agent-based models. *Ecol Model* **198**: 115–126.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J., and Railsback, S.F. (2010) The ODD protocol: A review and first update. *Ecol Model* **221**: 2760–2768.

Grimm, V., Railsback, S.F., Vincenot, C.E., Berger, U., Gallagher, C., DeAngelis, D.L., et al. (2020) The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism. *J Artif Soc Soc Simul* **23**: 7.

Some parts in quotations are taken directly from the original publication and documentation where CellModeller was presented:

Rudge, T.J., Steiner, P.J., Phillips, A., and Haseloff, J. (2012) Computational Modeling of Synthetic Microbial Biofilms. *ACS Synth Biol* **1**: 345–352.

1. Purpose and patterns

The general purpose of CellModeller is to predict behaviour and spatial organization of biofilms from basic principles.

In Rudge et al. (2012), microcolony morphology of experimental and simulated data was qualitatively compared. Additional patterns that can be used to validate the biophysics module were surveyed in Chapter 4.

2. Entities, state variables, and scales

CellModeller is composed of three modules: biophysics, intracellular dynamics, and extracellular signalling. The biophysics module is required in all simulations, whereas intracellular dynamics and signalling are optional. The state variables required for each module are described in Table E.1. More state variables can be added to cells at the discretion of the user.

Table E.1. State Variables in CellModeller.

State Variable	Variable Name	Units	Default Value	Module
Cell radius	radius	μm	0.5	Biophysics
Cell length	length	μm	3.5	Biophysics
Centroid position	pos	[x, y, z] coordinates	N/A	Biophysics
Cell orientation	dir	[x, y, z] vectors	N/A	Biophysics
Growth rate	growthRate	time^{-1}	1	Biophysics
Divide flag	divideFlag	True/False	False	Biophysics
Cell type identifier	cellType	Integer number	0	All
Intracellular species concentration	species	(mass or moles)/ μm^3	N/A	Intracellular
Signal concentration	signals	(mass or moles)/ μm^3	N/A	Signalling

Different physiological attributes can be assigned to cells based on their *cellType*.

State variables are advanced at discrete time steps. The time step, dt , is constant and user-defined when initializing the *Simulator* object. Typical time steps that maintain numerical stability are on the order of 0.01 h. In the biophysics module alone, the spatial domain is continuous, and its size is based on the minimum and maximum coordinates of all cells at a given time step. In the signalling module, the domain is finite and represented by a grid of control volumes with dimensions defined by the user.

3. Process overview and scheduling

Process overview: A Simulator object is initialized in a batch script. Inputs to the Simulator object are the simulation module and time step. The simulation is run by calling the main loop function in the Simulator object.

Schedule: The main simulator loop is given in the *step* function in *Simulator.py*. In the following list, steps that are taken verbatim from Rudge et al. (2012) are in quotes:

1. “Call an update function to apply user-defined rules to each cell”.
2. “Divide cells that have their divideFlag set to True”.
3. “Integrate growth of cells forward by a chosen time step Δt , solve constraints to obtain new cell positions”, update biophysics state variables for each cell.
4. “Integrate the species and extra-cellular signals u of each cell forward by the time step Δt ”, update intracellular and signalling state variables for each cell.
5. Save state variables to a .pickle file if n time steps have elapsed.
6. “...Repeat from step 1”.

The simulation is terminated based on a user-defined condition in the batch file that runs the simulator. Typical termination conditions are a maximum number of cells or time steps.

4. Design concepts

Basic principles: The biophysics model is based on the standard molecular dynamics algorithm (Poschel and Schwager, 2005), with some simplifications. Cells grow according to the exponential growth law. Cells experience drag forces according to Newton’s law of viscosity,

and overlap forces are treated like ideal springs. The contact forces acting on each cell are calculated by finding a solution that minimizes the sum of squared cell-cell overlap distances.

Emergence: Spatiotemporal arrangement of single- and multi-species biofilms are expected to emerge from the basic principles and sub-models applied to individual cells. Spatiotemporal patterns in state variables other than positioning may arise.

Adaptation: By default, cells have no adaptation. Adaptation could be added by the user in the form of motility, chemotaxis, and evolution.

Objectives: By default, cells do not act to meet some objective.

Learning: By default, cells do not learn.

Prediction: None. Perhaps stochastic variation in gene level expression and bet-hedging could be implemented as a form of prediction.

Sensing: Cells are “aware” only of their internal species concentrations and the extracellular species concentrations at their position. Cells can also be aware of their neighbours that they are in physical contact with.

Interaction: There are no assumed interactions between cells aside from displacement due to cell growth. Additional interactions may be included at discretion of the user. Indirect interactions could include competition for a limited nutrient. Physical interactions could include bacterial conjugation and toxin-injection by T6SS. Longer-range interactions could include secretion of metabolites and quorum sensing molecules.

Stochasticity: Some noise is applied to the parent and daughter cell orientations at division. This simulates imperfections in cell shape and alignment and breaks alignment symmetry between dividing cells. The user has the option to include stochasticity in state variables related to cell physiology. For example, a cell’s division length and growth rate can be sampled from a probability distribution upon initialization or birth.

Collectives: Groups of cells with the same *cellType* can form patches in colonies. These collective patches of cells with the same *cellType* are an emergent property of the basic principles and submodels applied to each cell.

Observation: Data is exported to .pickle files every n time steps, where n is specified by the user. The exported data are described in Table E.2.

Table E.2. CellModeller Outputs.

Output	Description	Data Type
cellStates	State variables for each cell at $t = \text{stepNum} \times dt$	Dictionary of cellState objects
stepNum	Step number	int
lineage	Cell ids for all parents of each cell	list
moduleStr	Code for the simulation module	string
moduleName	Name of the simulation module	string
specData	Intracellular species concentrations	Dictionary of floats
sigData	Signal concentrations at cell centroids	Dictionary of floats
sigGridOrig	Origin of the signal grid	1×3 tuple
sigGridDim	Number of control volumes [x, y, z]	1×3 tuple
sigGridSize	Control volume size [x, y, z]	1×3 tuple
sigGrid	Signal concentration in each control volume	list

5. Initialization

A simulation is initialized by creating a *Simulator* object in a batch file. A *Simulator* object takes as input the CellModeller simulation module, time step size, data output frequency, and the device number where simulations will run.

The settings for each of the three desired modules are initialized at the beginning of the *setup* function of the simulation module. After setup, cells are added into the domain. The positions and orientations can be specified as deterministic or random, depending on the application. Other state variables for each cell like the division length threshold, growth rate, intracellular species concentration, and signal concentrations are then initialized in the *init* function of the simulation module. Initialization for the biophysics module is described below.

Biophysics initialization: The biophysics model is initialized by creating a *CLBacterium* object. The attributes of the class are listed in Table E.3. After initialization, walls can be defined as planes.

Table E.3. Default Attributes of the *CLBacterium* Class.

Attribute	Variable Name	Units	Default Value
Maximum sub-steps in solver	max_substeps	N/A	8
Maximum number of cells	max_cells	N/A	10000
Maximum number of contacts per cell	max_contacts	N/A	24
Maximum number of planes	max_planes	N/A	1
Maximum number of spherical boundaries	max_spheres	N/A	1
Maximum number of elements in computational search grid	max_sqs	N/A	192 ²
Grid element size	grid_spacing	μm	5
Medium viscosity	muA	$\text{kg} \cdot \mu\text{m}^{-1} \cdot \text{time}^{-1}$	1.0
Cell stiffness	gamma	$\text{kg} \cdot \text{time}^{-1}$ (viscosity * length)	10.0
Time step	dt	time	None
Solver tolerance	cgs_tol	N/A	5×10^{-3}
Jitter flag	jitter_z	boolean	True
Nonsymmetric division flag	alternate_divisions	boolean	False
Print to terminal flag	printing	boolean	True
Compute neighbours flag	compNeighbours	boolean	False

6. Input data

Simulations do not require input from external sources. However, the simulation module is flexible enough to allow the user to import external data sources, such as positions, orientations, and cell types from microscopy experiments. Physiological properties such as cell radius, division length, and growth rate can be sampled from probability distributions obtained from experimental data.

7. Submodels

User-defined rules: state variables for each cell are updated based on user-defined rules. At minimum, the condition for cell division should be specified in this function. A typical condition for cell division is to set the `divideFlag` to `True` when the cell exceeds the division length threshold.

```
def update(cells):
    #Iterate through each cell and flag cells that reach target size for division
    for (id, cell) in cells.items():
        if cell.volume > cell.targetVol:
            cell.divideFlag = True
```

Cell division:

Cell division occurs for any cell that has a `divideFlag` equal to `True`. Initially, the daughter cell inherits all state variables from the parent, and the age of the new parent and daughter are reset to zero. The position, velocity, and lengths of the new parent and daughter are set using the `divide_cell` function from the *CLBacterium* class. The parent cell divides symmetrically in half, and centroid coordinates of both cells are updated accordingly. Both the parent and daughter are given same direction. To break symmetry of orientation at division, $U(-0.001, 0.001)$ is added to the orientation vector of both the parent and daughter in the x-, y-, and z-coordinates. Linear and rotational velocities of the daughter are inherited from the parent. If the intracellular dynamics or signalling modules are active in the simulation, then the daughter inherits the species and signal concentrations from the parent. Finally, the user-defined *divide* function in the simulation module is called, where state variables of the parent and daughter can optionally be reassigned to generate stochasticity.

Cell growth:

After the division step, the length of each cell is increased by $v_{growth}\Delta t$, where v_{growth} is described by exponential growth:

$$v_{growth} = \frac{dl}{dt} = \mu_{growth}l. \quad (E.1)$$

Cells are then displaced, rotated, and grown based on the linear velocity, angular velocity, and growth velocity of the previous time step.

Resolution of overlaps:

Displacement of cells is calculated by minimizing elastic energy stored in cell-cell overlaps. Elastic energy is modelled by a Hooke model (i.e., energy is proportional to overlap distance squared). The energy equation is derived by relating the generalized impulse $\Delta\tilde{\mathbf{p}} = (\Delta\mathbf{p} \ \Delta\mathbf{L} \ \Delta\mathbf{g})$ required to displace a cell by $\Delta\mathbf{x}$ in a viscous medium. These are given as follows:

$$\Delta\mathbf{x} = \frac{\Delta\mathbf{p}}{\mu l_0} \quad (\text{E.2})$$

$$\Delta\theta = \frac{12\mathbf{P}\Delta\mathbf{L}}{\mu l_0^3} \quad (\text{E.3})$$

$$\Delta l = \frac{\Delta g}{\gamma}, \quad (\text{E.4})$$

where μ is the medium viscosity, l_0 is the length of the cell, $\mathbf{P}$ is a projection matrix used to project the rotational impulse onto the plane of the cell axis, $\mathbf{L}$ is the rotational impulse, $\mathbf{g}$ is the growth impulse, and γ is a measure of the cell rigidity. Details on derivations are provided in the original documentation for CellModeller. The assumptions required to formulate the above equations are:

- Growth occurs by extension by all parts of the cell along its major axis
- Cells experience drag from the medium according to Newton's law of viscosity
- Inertial effects on the bacterium are negligible compared to drag forces
- Bacterial capsules are approximated as slender rods with no rotational inertia
- The change in a bacterium's length in a single time step is very small compared to the total bacterium length
- Resistance of the cell to deformation is much larger than viscous drag from the medium

Cell motion due to a generalized impulse is then characterized by:

$$\Delta \tilde{\mathbf{x}}_i = \begin{bmatrix} \text{diag}\left(\frac{1}{\mu l_0}\right) & 0 & 0 \\ 0 & \frac{12\mathbf{P}}{\mu l_0^3} & 0 \\ 0 & 0 & \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{p} \\ \Delta \mathbf{L} \\ \Delta \mathbf{g} \end{bmatrix} = \mathbf{M}_i^{-1} \Delta \tilde{\mathbf{p}}_i. \quad (\text{E.5})$$

The impulses required to resolve overlap distances $\mathbf{d}$ between cells are defined by applying impulses in along the contact normal vector direction $\mathbf{B}$, resulting in the equation

$$\mathbf{B}\mathbf{M}^{-1}\Delta \tilde{\mathbf{p}} = \mathbf{A}\Delta \tilde{\mathbf{p}} = \Delta \mathbf{d}, \quad (\text{E.6})$$

which can be rewritten as $\mathbf{A}\Delta \tilde{\mathbf{p}} + \mathbf{d} = 0$ since $\Delta \mathbf{d} = -\mathbf{d}$. This system of equations is constrained by selecting a solution that minimizes the work done by the applied impulses. Applying this regularization and setting the derivative with respect to $\Delta \tilde{\mathbf{p}}$ to zero yields the equation governing resolution of cell overlaps

$$(\mathbf{A}^T \mathbf{A} + \alpha \mathbf{M})\mathbf{p} = -\mathbf{A}^T \mathbf{d}, \quad (\text{E.7})$$

where α is a regularization parameter that weights the energy cost of cell movement. This equation is solved iteratively using a conjugate gradient solver to calculate the impulses that will result in mechanical equilibrium whilst minimizing cell displacement, to within a tolerance defined by the parameter *cgs_tol*.