Exploring heterogeneous phenotypic states in bacteria.

Leandra M. Brettner

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

R. William DePaolo, Chair

Eric Klavins, Chair

Wendy Thomas

Georg Seelig

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Exploring heterogeneous phenotypic states in bacteria.

Leandra M. Brettner

Chairs of the Supervisory Committee:

Professor R. William DePaolo
College of Medicine, Division of Gasteroenterology

Professor Eric Klavins
Department of Electrical Engineering

The collective dry biomass of all bacteria on earth is estimated to be in the ballpark of 400 trillion tons. A single human body alone can harbor approximately 100 trillion bacterial cells. We are only just beginning to understand how the goings-on of these creatures affect not only our own health and biology, but the health and biology of every ecosystem on earth. The vast majority of naturally existing bacteria (>99%) have yet to be cultured in a laboratory setting, making studying them using classical microbiology techniques difficult at best. Investigating the complex community structures of these organisms becomes even more difficult still. To work around this problem, we can first construct synthetic microbial consortia that are engineered to exhibit the behaviors we would like to study. Or second, we can develop applications that can measure organism behavior directly in their natural environments, as is being seen the rapid expansion of the –omics technologies. In this exam, we use both approaches to examine how single-cell behaviors, in both isogenic and mixed bacterial communities, affect the phenotypic dynamics of the population as a whole.

ACKNOWLEDGEMENTS

As with most who go through the PhD process, I have many people to acknowledge and thank; and without whom, I would not be Dr. Leandra Brettner.

My advisers and committee: Professors Georg Seelig, Wendy Thomas, R. William DePaolo, Eric Klavins, Amy Willis, and Hao Yuan Kueh.

My collaborators: Professor Ben Kerr, Dr. Robert Egbert, Dr. David Zong, Ryan Kibler, Yue Zhang, and Matthew Hirano.

My dear friends and colleagues: Dr. Anna Kuchina, Luana Paleologu, Dr. Alberto Carignano, Dr. Alex Rosenberg, and Dr. Charlie Roco. In no particular order, Tileli Amimeur, Dr. Paul Sample, Dr. Sunny Rao, Nick Bolten, Dr. Stephanie Berger, Cody Johnson, Dr. David Younger, Meghan Flaherty, Dr. Chris Takahashi, Dr. Sifang Chen, Dr. Ayah Shevchenko, Dr. Chuhern Hwang, Dr. Randolph Lopez, Oscar Wong, Dr. Nick Bogard, Alex Baryshev, Ban Wang, Johannes Linder, Dr. Max Darnell, Dr. Sergii Pochekailov, Dr. Arjun Khakar, Dr. Sumit Mukherjee, Ladan Mukherjee, Dr. Ben Groves, Dr. Gourab Chatterjee, Dr. Yuan-Jyue Chen, Dr. Richard Muscat, Brianna Kuypers, Justin Vrana, Tamara Casper, Cami Corday, Samer Halabiya, Michelle Parks, Dr. Miles Gander, Dr. Yaoyu Yang, Dr. Shelly Jang, Dr. Kyle Havens, Dr. Jason Stevens, Dr. Jonathon Tsui, Dr. Robert Lamm, Dr. Ian Dryg, Dr. Blake Bluestein, Dr. Denise Chac, Dr. Andrew Johnson, Melissa Kordahi, Dr. Marion Avril, Dr. Arushi Verma, Dr. Surya Kotha, Dr. Joseph Powers, Carrie Glenney, Dr. Sonia Singhal, Hannah Jordt, Katrina Van Raay, Dr. Peter Conlin, Professor Luis Zaman, Dr. Trevor Harrison, Dr. Renae Cruz, David Sparkman-Yager, Lee Organick, Dr. Bichlien Nguyen, David Ward, Nandana Rao, C. Andew Mencel, Alex Harris, Theodore Gatchell, Amber Agnis, Jaclyn Mendenhall, and anyone else I may have forgotten.

My wonderful and supportive family: my parents – William and Annette Brettner, my brother and sister-in-law – Eric and Lisa Brettner, my niece – Evelyn Brettner, my aunt and uncle and godparents – Dr. Anita and Ken Kuhnle, my aunt and uncle – Karen and Carmen Renda, and my cousins – Brian and Pam Renda, Tamara, Koby, Benjamin, and Grayson Owen.

Your support has been priceless, thank you with all of my being.

CONTENTS

CHAPTER 1

Exploring heterogeneous phenotypic states in bacteria

1.1 INTRODUCTION

There are an estimated 10^{30} microbes on the planet—that is 10^9 more than the number of stars in the known universe[1]. The diversity among single-celled organisms is vast and largely uncharted. Perhaps even more astronomical is the fact that these organisms are not living isolated lives, ambiguous to those around them, but instead form relationships with their neighbors that are proving to be every bit as complex as creatures on the macroscopic scale. Communities of microbes can cooperate to access resources, divide laborious tasks between factions, build shelters from the elements and for offspring production, collectively forage for nutrients, and perform acts of altruism to fend off predators, just to name a few[2]. They also form intricate rapports with the eukaryotic hosts that serve as contained, and often transient, habitats for some species[3]. Bacteria have even been discovered in places previously thought strictly sterile, including the placenta and the brain[4,5].

Powerful research is being done in the realm of natural and synthetic microbial ecosystems[6]. In fact, there is a push in biology to completely redefine organismal

hierarchies based on the species and activities present in corresponding microbial consortia[7]. Paradoxical phenomena such as the persistence of cooperative traits suggest that natural selection may act at levels superior to the individual, but are dependent on the collective behavior of individuals[8]. Populations of both homotypic (kin) and heterotypic groups phenotypically diversify and perform functions that increase their productivity as a whole while not necessarily promoting single organisms[9]. These complex social interactions can drive the evolution of increasingly complicated behaviors that we are only beginning to understand.

In the subsequent chapters, we illustrate our characterization of the behavior of individual cells and their contribution to the communal phenotype in two systems. The first system (Chapter 2), is a synthetic population of bacteria engineered to behave cooperatively to promote the welfare of the group. In the second system (Chapter 3), we worked to further single-celled transcriptomics sequencing technology to enable expression profiling in bacteria at the individual level which has broad applications in the field of microbiology. The accomplishments of this exam are described in the following summaries.

1.2 CHAPTER SUMMARIES

Chapter 2. Self-destructive altruism in a synthetic developmental program enables complex feedstock utilization[10]. We genetically engineered a strain of *Escherichia coli* to differentiate between two expression states. Individual cells make stochastic decisions to preside in either a germ cell like growth state or a suicidal state that delivers cellulose digestive enzymes to the extracellular environment. The structure of the genetic circuit is

such that the likelihood that a cell exists in either state is tunable. This allows populations of cell to subsist on a novel carbon source with varying degrees of success dependent on their engineered genotype. We took advantage of the genetic circuit's modular design to experimentally determine the independent dynamics of the system components on population behavior including differentiation rate, lysis rate, and cellulase enzyme kinetics.

Chapter 3. Microbial single-cell RNA sequencing by split-pool barcoding[11]. We successfully modified a novel single-cell RNA sequencing platform, called SPLiT-seq[12], to work on prokaryotic samples. Reverse transcription and barcode sequence ligation occur *in situ*, using the cell structure as a container to keep associated RNAs together. The method uses random split-pool barcoding to allow for unique cell labelling reactions to take place in a single pot experiment. Our major obstacles were modifying the existing protocol to contend with the cell wall diversity and low mRNA copy numbers of single-celled organisms. Now that we have successfully developed the method to work in common laboratory strains of bacteria, the technology can be used to investigate the expression diversity of heterologous bacterial populations, such as the microbiome.

1.3 REFERENCES

1. Curtis, T. P. & Sloan, W. T. Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Curr. Opin. Microbiol.* **7,** 221–226 (2004).
2. Crespi, B. J. The evolution of social behavior in microorganisms. *Trends Ecol. Evol.* **16,** 178–183 (2001).
3. Fabbiano, S., Suárez-Zamorano, N. & Trajkovski, M. Host–Microbiota Mutualism in Metabolic Diseases. *Front. Endocrinol.* **8,** (2017).
4. Aagaard, K. *et al.* The Placenta Harbors a Unique Microbiome. *Sci. Transl. Med.* **6,** 237ra65-237ra65 (2014).
5. Branton, W. G. *et al.* Brain microbiota disruption within inflammatory demyelinating lesions in multiple sclerosis. *Sci. Rep.* **6,** 37344 (2016).
6. Xavier, J. B. Social interaction in synthetic and natural microbial communities. *Mol. Syst. Biol.* **7,** 483 (2011).
7. Zilber-Rosenberg, I. & Rosenberg, E. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32,** 723–735 (2008).
8. Godfrey-Smith, P. & Kerr, B. Selection in ephemeral networks. *Am. Nat.* **174,** 906–911 (2009).
9. Boles, B. R., Thoendel, M. & Singh, P. K. Self-generated diversity produces 'insurance effects' in biofilm communities. *Proc. Natl. Acad. Sci.* **101,** 16630–16635 (2004).
10. Egbert, R., Brettner, L., Zong, D. & Klavins, E. Self-destructive altruism encoded by a synthetic developmental program enables complex feedstock utilization. *ELife* **(in review)**
11. Kuchina, A. and Brettner, L. *et al.* Microbial single-cell RNA sequencing by split-pool barcoding. *bioRxiv* (2019) doi:10.1101/869248.
12. Rosenberg, A. B. *et al.* Scaling single cell transcriptomics through split pool barcoding. *bioRxiv* 105163 (2017). doi:10.1101/105163

CHAPTER 2

Self-destructive altruism in a synthetic developmental program enables complex feedstock utilization

2.1 ABSTRACT

Stochastic differentiation and programmed cell death are common developmental processes in microbes that balance cellular fitness against the utility of sacrifice. The genetic basis for many of these behaviors is poorly understood. Here, we engineered a synthetic developmental gene network that implements a two-member division of labor to utilize cellulose for growth. Progenitor consumers differentiate to export an otherwise sequestered enzyme payload as altruists. We used an experimentally parameterized model to set key parameters and achieve net population growth. An inevitable consequence of engineering altruistic developmental behaviors is the emergence of cheaters that undermine cooperation. We observed cheater phenotypes for each cell state and identified genetic modes of escape. This work introduces the altruistic developmental program as a tool for synthetic biology, demonstrates the utility of population dynamics models to engineer complex phenotypes and provides a testbed for probing the evolutionary biology of self-destructive altruism.

2.2 INTRODUCTION

Compartmentalization of function across differentiated cell types was essential to the emergence of complexity in biological systems. Organogenesis in plants and animals[1], schizogamy in polychaete worms[2] and germ-soma differentiation in *Volvox* algae[3] are clear examples of these divisions of labor. Many microbial developmental programs utilize stochastic differentiation and programmed cell death as vital components of population fitness[4]. Selection for programmed cell death has been proposed to drive complex behaviors that delay commitment to costly cell fate decisions[5], enable adaptation to environmental fluctuations[6], eliminate competitor species[7], reinforce biofilm structure[8] and promote colonization of hostile environments[9]. These behaviors represent divisions of labor between subpopulations of progenitor cells that propagate the species and sacrificial cells that provide a public good, analogous to germ and somatic cell lines in multicellular organisms. Many aspects of the emergence of multi-cellular cooperation and the genetic circuits that control its complexity remain unclear. Limited understanding of the network architectures and stimuli that control developmental gene networks constrains efforts to repurpose them for engineered cell behaviors.

Current engineering paradigms of DNA-encoded cellular logic and feedback control circuits fail to encompass the full suite of behaviors desired to advance the fields of bioprocessing, bioremediation and cell-based therapeutics. Synthetic microbial consortia have been demonstrated to improve bioprocessing efficiency[10] or to explore other complex behaviors[11]. A major challenge to engineering microbial consortia is the control of community distributions for complex traits. While syntrophic interactions in

defined communities may address some of these needs, they may not be sustainable in environments with fluctuating nutrient or microbial constituents. Further, efficient delivery of protein or chemical payloads to the environment is constrained by the cell membrane, often requiring the expression of payload-specific pumps or secretion signals. Autolysis triggered by chemical[12,13] or autoinducer[14] signals allows release of protein payloads, but prevents applications that may require continuous delivery. Synthetic developmental programs could address these challenges, enabling ways to seed and regenerate microbial communities founded by individual cells to cooperatively perform complex tasks.

2.3 A SYNTHETIC ALTRUISTIC DEVELOPMENTAL PROGRAM

Here we present a synthetic developmental program that implements a germ-soma division of labor to cooperatively digest cellulose. The program links a synthetic differentiation controller with autolysis-mediated enzymatic payload delivery, balancing the rates of stochastic differentiation and programmed cell death to drive overall population growth. We developed a genetic circuit to create a synthetic community of cellobiose consumers that differentiate at a controlled rate to create a sub-population of self-destructive altruists. Together the community utilizes cellulose as a sole carbon source for growth through the extracellular release of a cellulase payload (Figure 2.1a).

We refer to the system as SDAc, short for self destructive altruism with a cellulase payload. We implemented SDAc in *Escherichia coli* by engineering a native operon to efficiently utilize cellobiose, introducing a genetic toggle switch tuned to function as a differentiation controller, and constructing a cellulase-lysis payload module

to execute the altruist behavior (**Figure 2.1b**). We analyzed dynamical systems models to identify parameter sets critical to achieving overall growth and demonstrated control over the circuit behavior by fine-tuning each parameter.

Using multiplexed mutagenesis and selection, we isolated a strain with a growth rate in cellobiose that is 63% its growth rate in glucose. Though *E. coli* does not natively digest cellobiose, we modified the *chb* operon in a recombinogenic MG1655 derivative[15] by replacing the native chitobiose-regulated promoter with a strong constitutive promoter[16]. We further improved growth on cellobiose by subjecting the constitutive *chb* expression variant to multiple cycles of multiplexed recombineering targeting the *chb* genes and selected for cellobiose utilization in minimal cellobiose media (Figure 2.1 - figure supplement 1). We identified the variant with the highest growth rate, DL069, as a

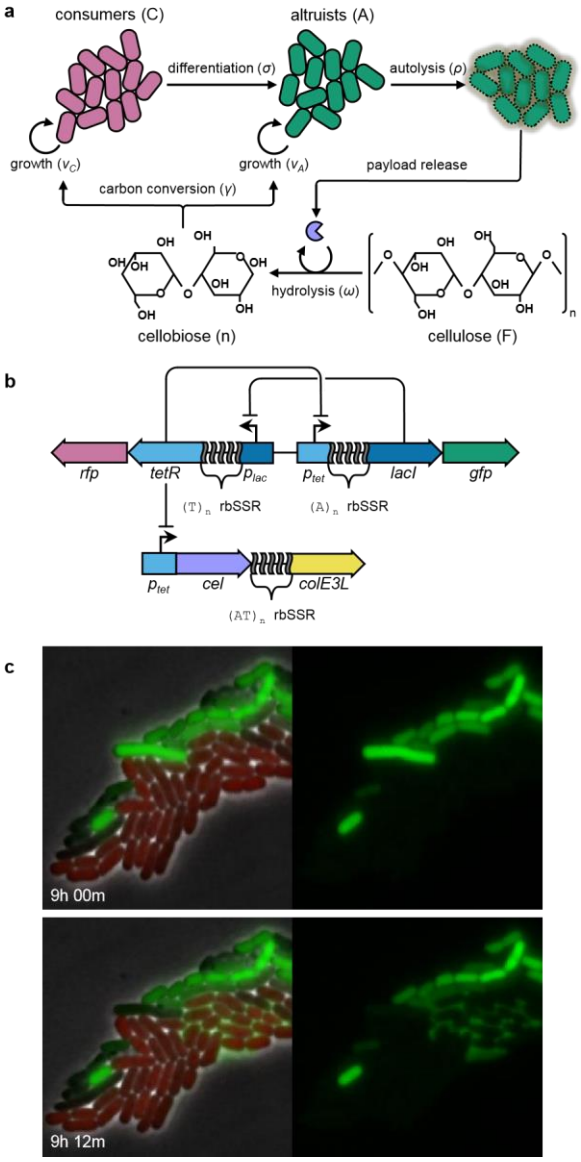


Figure 2.1: A synthetic developmental program for cooperative cellulose digestion. (a) Cellobiose consumers stochastically transition to self-destructive altruists. Altruists, in turn, produce and release cellulase payloads via autolysis to support the consumer population. (b) The genetic implementation of the SDAc developmental program includes a differentiation control plasmid (above) and a payload delivery plasmid (below). Cell states are mediated by a mutual inhibition toggle switch using transcriptional repressors LacI and TetR. TetR-dominant cells express RFP as consumers; LacI-dominant cells co-express cellulase and colE3 lysis (colE3L) proteins with GFP as altruists. Differentiation and lysis rates are fine-tuned with rbSSR sequences for tetR and lacI (differentiation) and colE3L (lysis). (c)

chbR deletion mutant.

Demonstration of differentiation and autolysis within a microcolony seeded by a single cell. A large altruist cell (arrow) accumulates its GFP payload (upper panel) until it undergoes autolysis (lower panel), enabling payload diffusion to surrounding cells (circle).

To control differentiation rate, we constructed and sampled from a library of mutual inhibition toggle switch variants that exhibit regular stochastic state transitions. While genetic toggle switches are often designed to function as bistable memory devices[17], a quasi-steady state can be achieved by properly balancing expression levels of the repressor proteins[18]. Simple sequence repeats embedded in the ribosome binding site (rbSSR) allow predictable modulation of translation initiation rate to tune the balance between transcriptional repressors[19].

We engineered altruist payload delivery by constructing a cellulase and lysis gene cassette. The cassette was designed to maximize production of the cellulase payload with an efficient ribosome binding site and a poly-(AT) rbSSR to fine-tune expression of the lysis gene. Optimal altruist efficiency would require maximizing payload accumulation before lysis. Colicin gene networks share this trait, using stochastic gene expression of colicin and lysis genes within subpopulations to kill ecological competitors[7,20]. We found that coupling the lysis gene from colicin E3 to the differentiation controller enabled stochastic state transitions and delayed lysis at the microcolony level, evidenced by accumulation of a GFP payload followed by cell death and payload release (Figure 2.1c).

2.4 PARAMETER ESTIMATES THROUGH MODULAR SYSTEM DECOMPOSITION

Analysis of a population dynamics model of SDAc behavior suggested optimal parameter regimes for cellulose utilization and guided implementation of the developmental circuit. Though we observed the requisite behaviors of differentiation,

payload accumulation and autolytic payload release at the microcolony level, it was not clear what combination of expression levels for circuit components would enable cooperative growth on cellulose. To reason about the functional parameter space for SDAc behavior we developed a population dynamics model using a system of ordinary differential equations that maps system parameters to experimentally tunable features of the genetic circuit. The model species are consumers, altruists, cellulose feedstock, and cellulose-derived nutrients. These species and the associated kinetic parameters are described in Box 1.

Box 1. Systems of equations for modelling self destructive altruism

i.
$$\dot{C} = \frac{n}{k_C + n} v_C C - \frac{n}{k_\sigma + n} \sigma C \quad (1)$$
$$\dot{A} = \frac{n}{k_A + n} v_A A + \frac{n}{k_\sigma + n} \sigma C - \rho A \quad (2)$$
$$\dot{F} = -\omega \rho A F \quad (3)$$
$$\dot{n} = \omega \rho A F - \frac{n}{\gamma}\left(\frac{v_C}{k_C + n} C + \frac{v_A}{k_A + n} A\right) \quad (4)$$

ii.
$$\dot{C} = (v_C - \sigma) C \quad (5)$$
$$\dot{A} = v_A A + \sigma C \quad (6)$$

iii.
$$\dot{C} = (v_C - \sigma) C \quad (5)$$
$$\dot{A} = (v_A - \rho) A + \sigma C \quad (7)$$

iv.
$$\dot{F} = -\omega \rho A F \quad (3)$$
$$CFU(n) = \gamma n + b \quad (8)$$

v.
$$\dot{C} = (v_C - \sigma - \chi_C) C \quad (9)$$
$$\dot{S} = v_C S + \chi_C C \quad (10)$$
$$\dot{A} = (v_A - \rho - \chi_A) A + \sigma C \quad (11)$$
$$\dot{P} = v_A P + \chi_A A \quad (12)$$

i. We constructed a population scale model composed of first order ordinary differential equations. The model contains four relevant species: consumers (C), altruists (A), cellulose (F, for feedstock), and cellobiose (n, for nutrients) with corresponding units of colony forming units per mL for cells and grams per mL for molecules (1-4). Consumer and altruist cells each grow in the presence of nutrients at rates v_C and v_A, respectively. Individual consumer cells differentiate to altruists at rate σ, and altruists lyse at rate ρ. We assume here that the catalytic cellulase enzyme activity occurs much faster than enzyme degradation, and thus consolidate parameters to remove cellulase dynamics as a variable. We represent enzyme levels in the system as a product of the population of altruists and the lysis rate. The equations account for non-growing cells by including nutrient-saturating terms for growth and differentiation. Commitment to lysis is assumed to be nutrient-independent.

Experimental Parameter Measurements

The SDAc developmental program is highly modular at the behavioral level. The system state can be set by chemical inducers, the dynamics of the other variables are measured using knockout strains and cellulase enzymes can be interchanged and combined as payload cocktails. These properties allow for the direct measurement of model parameters through the design of experiments that investigate isolated circuit components and activities. Thus, we built simplified sub-models to ascertain the key parameters empirically. Experimental and modelling approaches to parameter estimates are described in detail in the Supplementary Information.

ii. Module I, differentiation rate (σ)
To measure the rate at which cells switch between states, we isolated the differentiation module by knocking out the cellulase and autolysis genes, inducing the culture to the consumer state, and growing in pure cellobiose media (Figure 2.2a). Given that the growth rates (v_C, v_A) are known, this allowed us to model the system and fit σ as simple, linear DEs with analytical solutions (5 6).

iii. Module II, autolysis rate (ρ)
By reintroducing the functional cellulase and autolysis cassette, we expanded the sub-model for Module I to include the altruistic, self-sacrifice behavior (5,7, Figure 2.2B). Again, since growth rates and differentiation rates have been previously measured, we can fit for the autolysis rate.

iv. Module III, cellulose hydrolysis (ω)
Each cellulase enzyme has a different efficiency at which it converts cellulose to shorter polymers like cellobiose. We aimed to measure these hydrolysis rates by isolating the feedstock equation from the full model. To isolate ω, we forced A and ρ to constants by growing autolysis deficient strains induced to the altruist state, measuring cell density, and then lysing the entire culture using sonication to release intracellular cellulase. We then inoculated a 0.4% PASC solution with a fixed volume of filtered lysate and measured the degradation of cellulose. The cellulose hydrolysis parameter was fit to the simple, exponential solution to the feedstock equation (3).

v. Cheater dynamics
Given the strong negative selection pressure of autolysis, faster differentiating strains are prone to long-term instability. Defector cells can fail to perform the cooperative function, or cheat, in theoretically two ways: failure to switch and/or failure to lyse. We model the mutation rates of switch and lysis cheaters by modifying the model in Module II by adding the defector states (9-12, Figure 2.3c).

The modularity of implementing the developmental program as a synthetic gene circuit allowed us to decompose the full system model and its experimental components to predict cellulose utilization. We developed the modules described in Box 1 to measure growth rate, differentiation rate and lysis rate in cellobiose as well as cellulase activity, drawing from a small parts library for each module to sample a range of parameters.

We experimentally tuned differentiation rate over an order of magnitude with a collection of SDAc strains lacking cellulase and autolysis genes. Specifically, we used multiple poly-(T) rbSSR variants controlling expression of the consumer-dominant regulator TetR to modulate the differentiation rate from consumer to LacI-dominant altruists (Figure 2.2a,b). We measured the population fraction of differentiated cells as a function of time using flow cytometry (Figure 2.2 - figure supplement 1) and fit a two-state, continuous growth model to the data for consumers transitioning to altruists at rate

σ (Module II of Box 1). We found the repeat length to be inversely proportional to differentiation rate, supporting previous results for a switch encoded on a higher copy number plasmid[19] and resulting in σ estimates ranging from $2.7\times10^{-2}\pm9.0\times10^{-3}$ h^{-1} for $(T)_{12}$ to $2.11\times10^{-1}\pm1.5\times10^{-2}$ h^{-1} for $(T)_{18}$ (**Figure 2.2c, Table 2.4**). An onset delay for differentiation was inversely proportional to differentiation rate (Figure 2.2 - figure supplement 2).

Incorporating autolysis into the differentiation circuit resulted in predictive control of steady-state altruist levels. Using the intermediate rate differentiator $(T)_{16}$, we tested a set of poly-(AT) rbSSR variants to modulate expression of the colicin E3 lysis gene (colE3L) (Figure 2.2d,e). We used the same population fraction assay used to quantify differentiation to fit a consumer growth and differentiation model that includes altruist lysis parameter ρ (**Box 1, Module III**). We measured lysis rates from 7.2×10^{-2} $\pm1.8\times10^{-2}$ h^{-1} for $(AT)_{10}$ to $1.9\times10^{-1}\pm2.5\times10^{-2}$ h^{-1} for $(AT)_{8}$ (Figure 2.2f, Table 2.5). Model-derived predictions of altruist fractions that incorporate independently measured rates for differentiation and lysis are consistent with the observed behaviors (Figure 2.2 - figure supplement 3). As expected, the population fraction of differentiated cells for each autolytic SDAc variant is lower than for the equivalent colE3L-deficient strain (compare Figure 2.2c to Figure 2.2 - figure supplement 3a), due to loss of altruists to cell death. Interestingly, in contrast to rbSSRs used to control expression of differentiation rate, we found that the lysis rate did not correlate with rbSSR length (Figure 2.2 - figure supplement To estimate cellulase activity we quantified cellulose degradation from cell lysates of autolysis-deficient SDAc strains producing one or two cellulases, observing hydrolysis rates over a four-fold range. We measured cellulose degradation and digestible

of M9 minimal pnutrient release for three endoglucanases from two glycoside hydrolase (GH) families: CelD04[21] and BsCel5[22] from GH5; and CpCel9 from GH9[23] (Figure 2.2g,h). We also measured the activity of multi-enzyme cocktails using each GH5 enzyme with CpCel9, combinations with reported synergistic activities[24]. We used Congo Red staining phosphoric acid swollen cellulose (PASC) media spiked with cell lysate to observe cellulose degradation up to 23% (Supplementary Note 6) in and quantified cell growth on the resulting supernatant to estimate nutrient release of up to 14% of cellobiose equivalents (see Figure 2.2 - figure supplement 5, Supplementary Note 6). We used these cellulase activity measurements to fit a value for ω to the feedstock differential equation (Module IV of Box 1). Cellulase activity estimates range from 3.6×10^{-13} CFU^{-1} mL for CpCel9 to 1.6×10^{-12} CFU^{-1} mL for CelD04 (Figure 2.2i, Table 2.7).

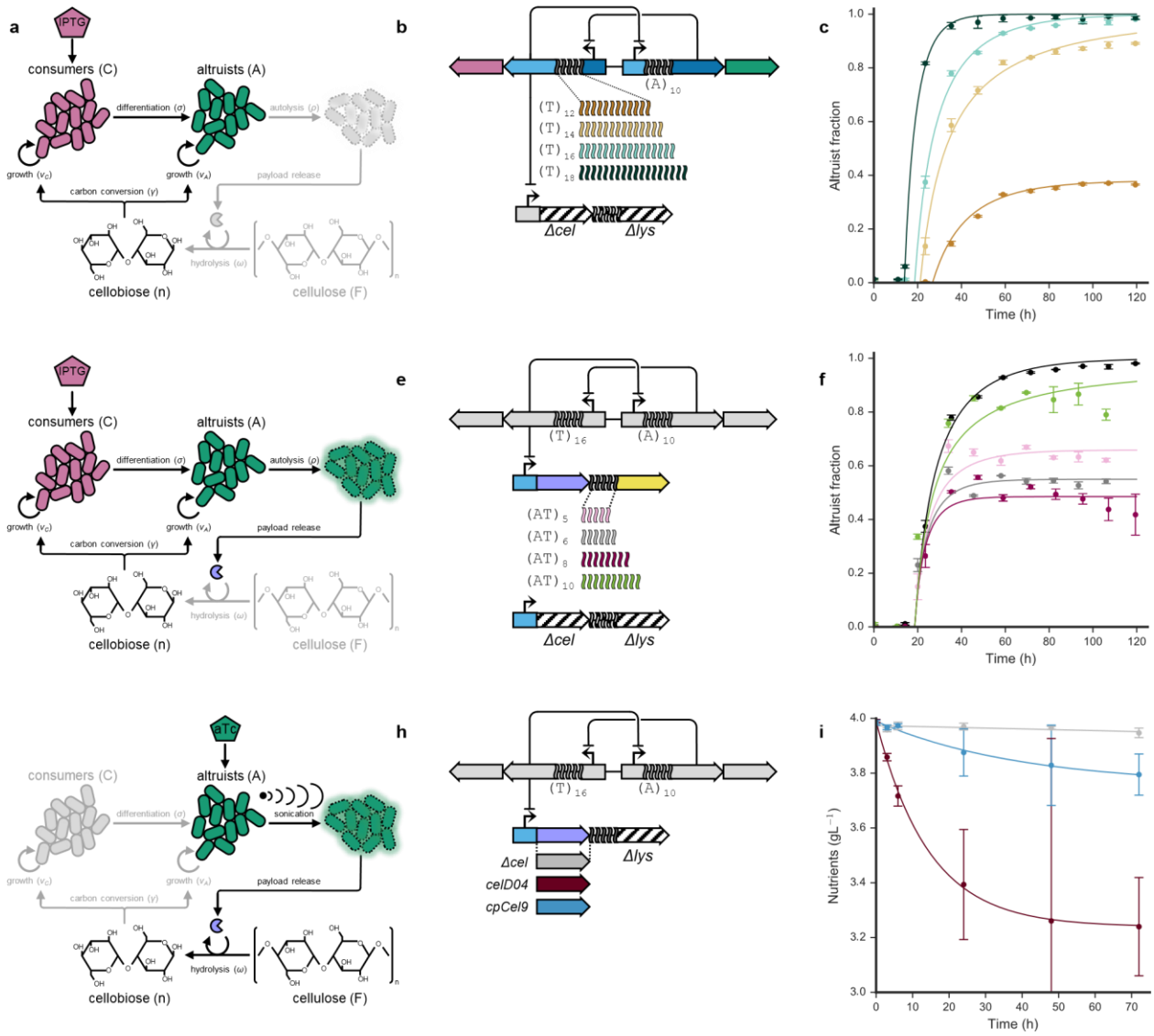


Figure 2.2. Experimental parameter sweeps for differentiation, autolysis and cellulase activity. (a) For the differentiation assay, cultures were initialized to the consumer state using IPTG and grown continuously in cellobiose to differentiate into cellulase- and lysis-deficient altruists. Cartoon components in grey correspond to unutilized states for the assay. (b) Genetic variants tested for differentiation vary poly-(T) rbSSR length to modulate TetR expression. Cellulase and lysis genes were removed from the payload plasmid (Δcel, Δlys). (c) Differentiation data and model fits to estimate $\boldsymbol{\sigma}$ (see Supplementary Note 4). Plot colors correspond to constructs depicted in (b). (d) For the lysis assay, cultures were initialized as in (a) for consumers to differentiate into autolytic altruists. (e) Genetic variants tested for lysis used intermediate rate differentiator $(T)_{16}$, varying (AT)-rbSSR repeats that control lysis gene expression or using a control plasmid with no cellulase or lysis genes (Δcel, Δlys). (f) Lysis data and model fits to estimate ρ (see Supplementary Note 5). Control from (c) shown in black for comparison. (g) For the cellulase activity assay, lysis-deficient strains were induced to the altruist state using aTc, grown to saturation and sonicated to generate crude cell extracts. (h) Genetic variants tests for cellulase activity by expressing different cellulases or maintaining a control plasmid (Δcel). (i) Cellulase activity data and model fits to estimate ω (see Supplementary Note 7). Data points represent the mean of 3 biological replicates (N = 3), and the error bars one standard deviation from the mean.

2.5 COOPERATIVE UTILIZATION OF CELLULOSE FOR GROWTH

To quantify the effects of differentiation and autolysis dynamics on feedstock degradation and population growth we measured cellulase activity from SDAc strains with the full genetic circuit. Cellulose hydrolysis by individual colonies was measured by Congo Red clearing assays from agar plates supplemented with carboxymethylcellulose (Figure 2.3). We found that the clearing diameter for switch variants increased as a function of differentiation rate (Figure 2.3a,b and Figure 2.3 – figure supplement 1a). We observed no clearings for a control lacking cellulase. We also tested the effect of rbSSR lysis variants combined with cellulase CelD04 (Figure 2.3c,d and Figure 2.3 – figure supplement 1b) as well as for individual cellulases (Figure 2.3g,h) using intermediate rate differentiator $(T)_{16}$. We found the GH5 cellulases generated larger clearings than CpCel9, consistent with the in vitro cellulase activity results.

Fine-tuning the differentiation, lysis and cellulase activity parameters is critical to realizing robust SDAc growth on cellulose as the sole carbon source. To determine population fitness on cellulose and validate the full dynamics model (Module I of Box 1), we measured viable cell counts in PASC for SDAc variants that span a range of values for each core parameter. We observed the highest population fitness at intermediate differentiation rates (Figure 2.3c), with high lysis rates (Figure 2.3f) and with high cellulase activity (Figure 2.3i), trends that are consistent with the naive model predictions from Figure 2.2. Model fits of growth dynamics using parameter estimates from individual modules match observations for most variants (Figure 2.3 - figure supplement 2), though the model predicted more growth for differentiation variant $(T)_{18}$ and failed to capture growth lag dynamics for a BsCel5-CpCel9 cellulase cocktail.

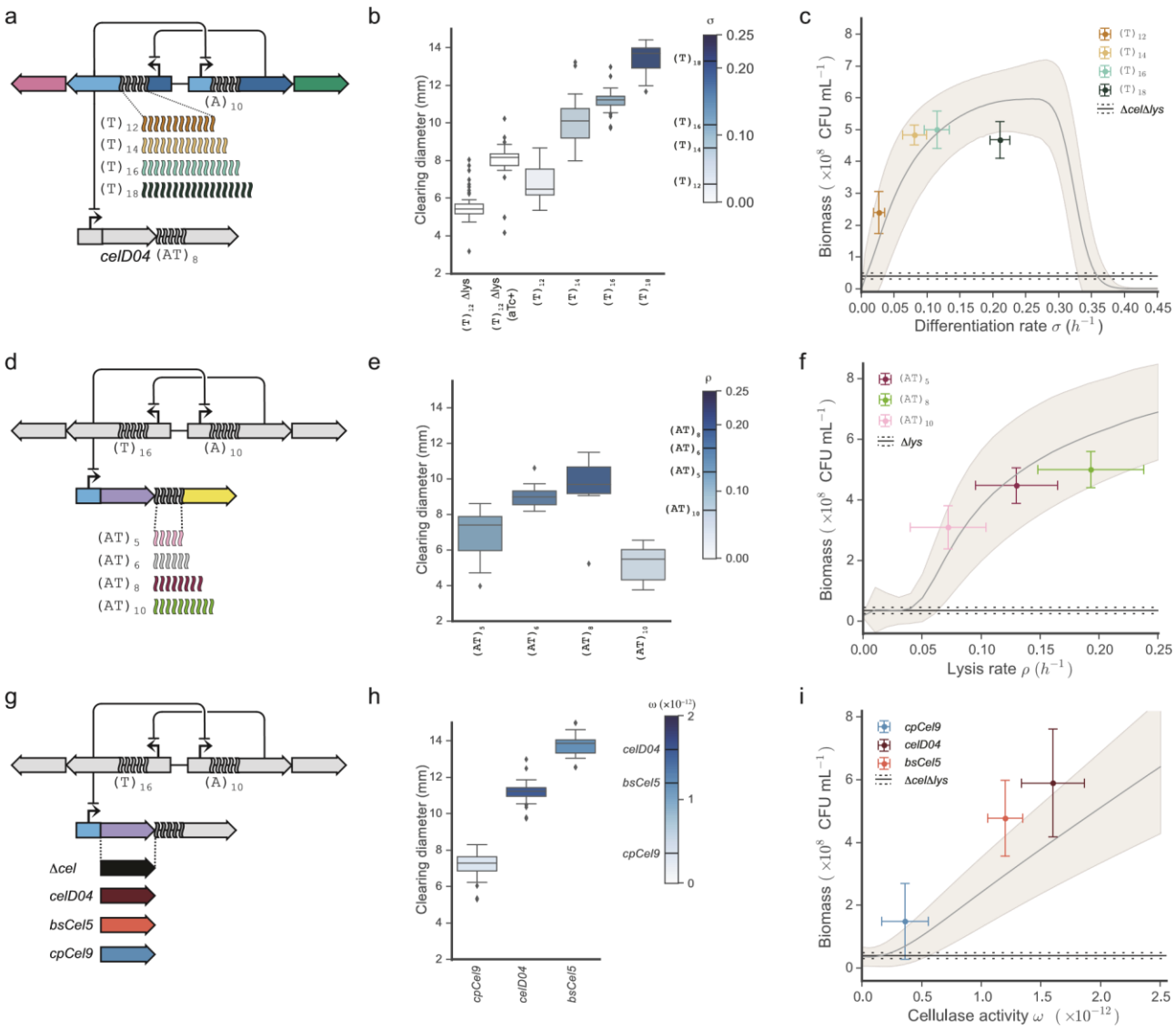


Figure 2.3. Characterization of cellulose hydrolysis and utilization for growth reveals SDAc model prediction accuracy. (a) Differentiation rate variants expressing cellulase CelD04 with lysis rbSSR $(AT)_8$. (b) Clearing size distributions for individual colonies from differentiation variants in (a) (N = 31, 21, 17, 15, 6). (c) Growth titer in M9 minimal 0.4% PASC media after 72 hours for differentiation variants in (a). Error bars on the x-axis represent standard deviation of the parameter estimate and error bars on the y-axis represent standard error for CFU counts from at least four replicates. Growth for control strain DL146 shown as black line. The shaded region represents one standard deviation of model uncertainty for cell growth. (d) Lysis rate variants for intermediate differentiator $(T)_{16}$ expressing cellulase CelD04. (e) Clearing size distributions for individual colonies from lysis variants in (d) (N = 20, 14, 11, 15). (f) Growth titer as in (c) for some lysis variants in (d). Growth for control strain DL147 shown as black line. (g) Cellulase variants for intermediate rate differentiator $(T)_{16}$. (h) Clearing size distributions for individual colonies from cellulase variants in (g) (N = 20, 15, 16). (i) Growth titer as in (c) for cellulase variants in (g). Error bars for the x-axis here represent the interquartile range for each ω estimate. Growth for control strain DL146 shown as black line. Whiskers shown for box plots in (b,e,h) extend one interquartile range.

2.6 EXCESSIVE DIFFERENTIATION REVEALS A TRAGEDY OF THE COMMONS

The full model of SDAc growth dynamics on cellulose predicts system collapse at high differentiation rates (Figure 2.3a), but does not account for mutational dynamics that could generate non-cooperative cheaters. Indeed, we observed functional instability for hyper differentiator switch variant $(T)_{18}$. The instability was manifest in continuous cellobiose culture as a temporally unstable altruist fraction (Figure 2.2 – figure supplement 3). Growth rate dynamics consistent with enrichment for mutants that overcome the population growth defects imposed by differentiation or lysis (see Figure 2.4e) support this hypothesis. We also observed mutant colony phenotypes for the same strain after extended growth in PASC media (Figure 2.4a), further suggesting functional instability at extreme differentiation rates.

Analytical solutions to candidate dynamic models that account for non-cooperative mutants suggest that two cheater subpopulations – one deficient in differentiation and the other deficient in lysis – are required to realize the observed dynamics. To estimate SDAc mutation rates we developed a model that introduces new species for switch-deficient cheaters (S) and for lysis-deficient pseudo-altruist cheaters (P), including their respective escape rates, χ_C and χ_A (Module V of Box 1). Four candidate models were investigated to account for the cheater dynamics: $\chi_C = \chi_A = 0$ (no cheaters), $\chi_A = 0$, $\chi_C > 0$ (only differentiation cheaters), $\chi_C = 0$, $\chi_A > 0$ (only lysis cheaters) or $\chi_C > 0$, $\chi_A > 0$ (both cheater types) (Figure 2.4b). We derived analytical solutions for each model, finding that the only model that supports the observed dynamics includes both cheater types (Supplementary Note 11). Cheaters may emerge from discrete mutational events during growth or be contained of the inoculum, rising in population fraction once

a large fraction of cells differentiate and lyse. Our analytical solutions do not distinguish between these options.

Fits of proposed mutagenesis model to measured differentiation and growth dynamics validate the dual cheater model and estimate mutation rates for each cheater type. We fit each model to the differentiation and lysis data for growth in cellobiose using hyper differentiator $(T)_{18}$. Fits for differentiation and growth dynamics are shown in Figure 2.4 – figure supplement 1 and rate estimates are shown in Table 2.9. The lysis cheater model provided no improvement to the null case and is not shown. The dual cheater model fit estimates the emergence of differentiation cheaters at a rate of 1.8×10^{-6} $\pm6.3\times10^{-7}$ h^{-1} and the emergence of altruist cheaters at a rate of $1.9\times10^{-5}\pm8.2\times10^{-6}$ h^{-1}. Using the mutagenesis parameters fit from hyper differentiator $(T)_{18}$, intermediate differentiator $(T)_{16}$ is also predicted to accumulate cheaters within the measurement interval (Figure 2.4c), which is consistent with the trend of the cellobiose switching data. When applying escape rate estimates to a model of the overall population growth dynamics for differentiation rate variants, we found the model predicted the variable growth rate dynamics observed for variants with high differentiation rates (Figure 2.4d).

DNA sequencing of cheater isolates confirms the genetic basis for both differentiation and lysis cheaters (Figure 2.4e). We observed large colonies with strong red fluorescence or, alternately, strong green fluorescence phenotypes – putative differentiation and lysis cheaters, respectively – in addition to small, mixed-color wild-type colonies on agar plates after extended growth in PASC media. Sequence analysis of the differentiation controller plasmid from red escape colonies isolated from six replicate cultures revealed mutations to two hypermutable loci with predictable effects (Table 2.10):

expansion or contraction of the tandem $(CTGG)_3$ mutational hotspot observed in four of six replicates should prevent altruist emergence through inactive, truncated Lac repressor[25]; and deletions within the $(T)_{18}$ rbSSR controlling TetR expression (one of six replicates) should abolish differentiation by reducing σ. A transposition event of insertion sequence IS2[26] internal to *lacI* (one of six replicates) should also prevent differentiation. Sequence analysis of the payload delivery plasmid revealed a transposition event of IS1[27] between the cellulase and lysis genes in one of six sequenced replicates, likely disrupting operon expression (Table 2.11). The majority of the altruist cheater colonies we sequenced revealed no mutations in the differentiation controller sequence or payload delivery transcription operon, suggesting lysis evasion via mutations on the genome or elsewhere on the plasmids. Given that the lysis gene is sourced from a colicin plasmid found in natural *E. coli* populations, it is possible our strain's genome encodes high-rate evolutionary paths to lysis immunity.

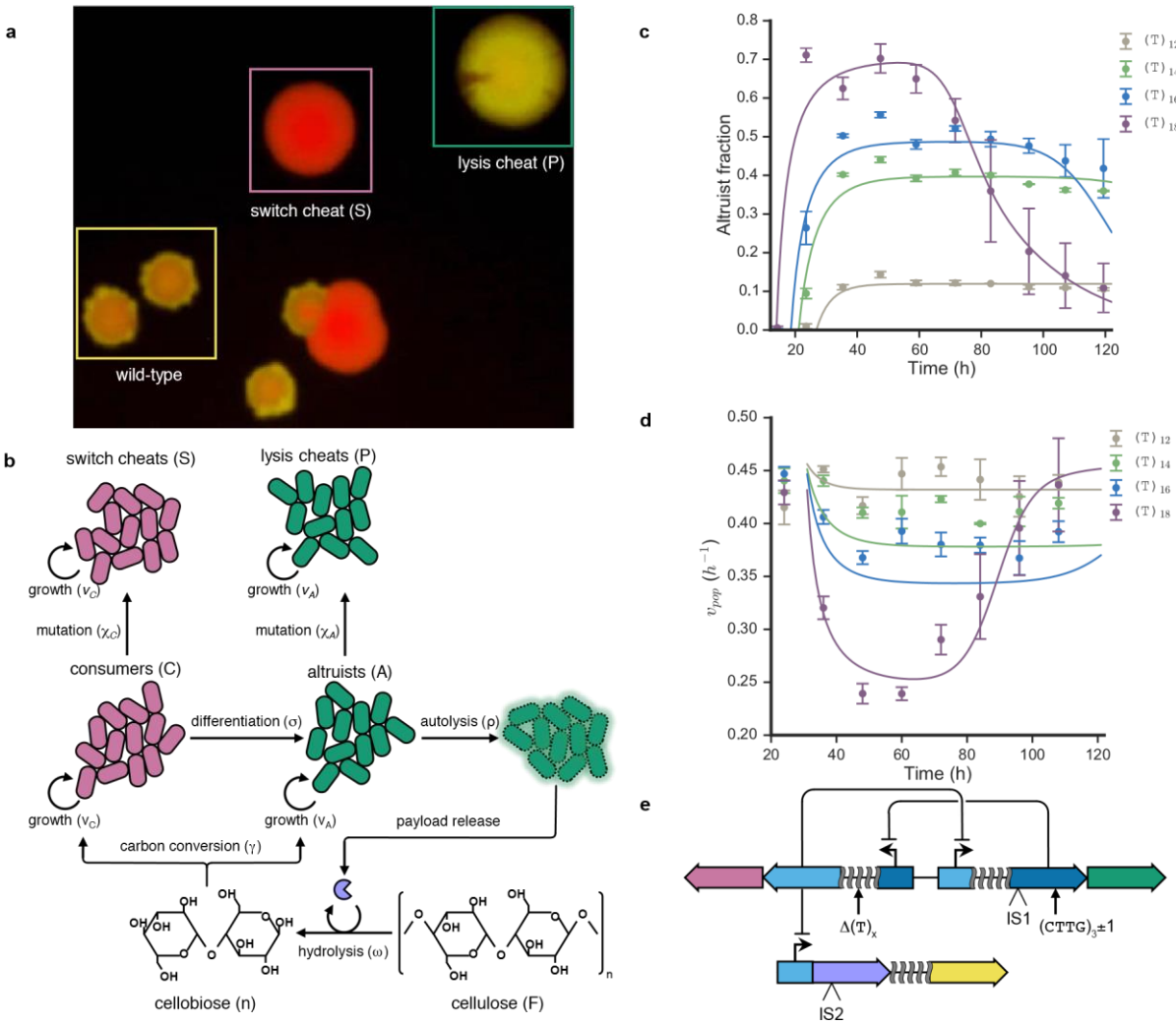


Figure 2.4. Mechanisms and rates of escape for SDAc cheaters. (a) Fluorescence image of wild-type and cheater colonies isolated from PASC cultures of hyper differentiator $(T)_{18}$ that no longer differentiate (S) or no longer lyse (P). (b) Representation of alternate SDAc model that incorporates mutational dynamics by including sink states for switch cheaters (S) and pseudo-altruist cheaters (P) with associated mutagenesis rates (see Module V of Box 1). (c) Dual cheater model fit to altruist fraction measurements for differentiation (compare to Figure 2.2 – figure supplement 3). Escape rate estimates for hyper differentiator $(T)_{18}$ are used to fit altruist fractions for the other differentiation rates. (d) Continuous growth model fit for overall population growth rate measurements of differentiation variants ($v_{!"!}$), using dual cheater rate estimates as in (c). (e) Summary of observed mutations that produce differentiation or lysis cheaters. Differentiation mutants included an expansion or contraction of a native simple sequence repeat element in the *lacI* coding sequence, repeat unit truncations of rbSSR $(T)_{18}$ that controls differentiation rate and insertion sequence disruption of LacI expression. Insertion sequence disruption of the CelD04 gene (green arrows) was observed for altruist cheaters. See Tables 10 and 11 for additional details on observed mutations. Data points represent the mean of 3 biological replicates (N = 3), and the error bars one standard deviation from the mean.

2.7 DISCUSSION

We have demonstrated a first-principles approach to construct a developmental gene circuit and have implemented a two-member developmental system to cooperatively utilize the complex feedstock cellulose. In-depth system deconstruction and characterization enables model-guided optimization of growth on cellulose. At extreme differentiation rates, genetic instability drives the emergence of cheaters that fail to differentiate or fail to lyse. This foundation will enable development of more robust and complex developmental divisions of labor to advance sustainable bioprocessing and cell-based therapeutics. Similar systems may also prove to be effective tools to advance the study of the evolution of cooperation.

Due to the observed functional instability for some variants, the SDAc program likely suffers from a tragedy 324 of the commons[28]. In well-mixed cellulose media, emergent cheaters fully benefit from the public good provided by the altruists. Further, due to the costs of switching and lysis, the cheaters can out-compete cooperators and sweep the population. In the absence of altruists, cellulase release ceases, driving population collapse. Previous work has shown that when the environment is spatially organized a communal benefit applies only to nearby, closely related cells who are likely fellow cooperators[29]. Indeed, research suggests the cellulosome evolved to localize the benefits of cellulase expression, as in sucrose utilization in yeast[30]. Thus cheaters are stranded with limited or no access to the shared resource. This phenomenon, attributed to kin selection, could preserve cooperative behavior for many more generations, potentially avoiding the functional instability we observed. Future work could elucidate the role of structured environments in this synthetic system to reduce the impact of

cheaters or to evolve more stable cooperator phenotypes.

While we only observed a high fraction of SDAc cheaters from hyper differentiation variant $(T)_{18}$, engineering developmental circuits for deployment in bioreactors or other complex environments would require long-term evolutionary stability to minimize cheaters and maintain engineered function. Interestingly, previous studies have shown that *lacI* tandem repeat mutations occur at a rate $>10^{-!!}$ events per generation[25] and transposon insertion elements jump at rates of $10^{-!!}$-$10^{-!!}$ insertions per generation[31]. These rates are consistent with our experimental estimates for mutagenesis, suggesting relatively simple modifications may considerably boost SDAc circuit longevity. Analysis of the cheater model suggests that a reduction of cheater rates by 100-fold and 1000-fold for intermediate differentiator $(T)_{16}$ would increase circuit stability by 56% and 85%, respectively, boosting the functional period in continuous culture from 2.8 days to 5.1 days. Genetic strategies to boost evolutionary stability include recoding the repeat region of *lacI*, introducing stabilizing degeneracy into rbSSR sequences and porting the system to a low mutation rate strain deficient for insertion elements[32]. Further gains in system performance could be achieved by chromosomal integration of the SDAc network to prevent the fixation of mutant plasmids in the population[33] or plasmid loss. Finally, incorporating more efficient cellulase cocktails will reduce evolutionary pressure for cheating by decreasing the optimal altruist load.

The division of labor system outlined here is a template for the construction of other developmental programs to perform complex tasks in engineered microbial communities. This work can be extended in many ways. For SDAc, the developmental program could be triggered in response to nutrient depletion when the supply of simple

sugars is depleted. Alternative protein and small molecule payloads from a general autolytic delivery system could be designed to mediate microbial interactions, aid in bioprocessing or bioremediation or as a cellular therapeutic. Further, stochastic strategies could be employed with or without self-destructive altruism to seed multicellular developmental programs for distributed metabolic engineering[34], evolutionary engineering[35] or to control distributions of multiple cell types in microbial communities[36,37]. Tunable developmental programs could also be applied to better understand the emergence and persistence of well studied developmental programs, substituting complex regulatory networks with tunable differentiation dynamics.

2.8 REFERENCES

1. Weissman, I. L. Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100,** 157–168 (2000).
2. Franke, H. D. Reproduction of the Syllidae (Annelida: Polychaeta). *Hydrobiologia* **402,** 39–55 (1999).
3. Prochnik, S. E. *et al.* Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* **329,** 223–226 (2010).
4. Rice, K. C. & Bayles, K. W. Molecular control of bacterial death and lysis. *Microbiol Mol Biol Rev* **72,** 85–109, table of contents (2008).
5. González-Pastor, J. E., Hobbs, E. C. & Losick, R. Cannibalism by Sporulating Bacteria. *Science* **301,** 510–513 (2003).
6. Claverys, J. P., Martin, B. & Håvarstein, L. S. Competence-induced fratricide in streptococci. *Mol Microbiol* **64,** 1423–1433 (2007).
7. Riley, M. A. & Gordon, D. M. The ecological role of bacteriocins in bacterial competition. *Trends Microbiol* **7,** 129–133 (1999).
8. Perry, J. A., Cvitkovitch, D. G. & Lévesque, C. M. Cell death in *Streptococcus mutans* biofilms: a link between CSP and extracellular DNA. *FEMS Microbiol Lett* **299,** 261–266 (2009).
9. Ackermann, M. *et al.* Self-destructive cooperation mediated by phenotypic noise. *Nature* **454,** 987– 990 (2008).
10. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat Biotech* **33,** 377–383 (2015).
11. Chen, Y., Kim, J. K., Hirning, A. J., Josic, K. & Bennett, M. R. Emergent genetic oscillations in a synthetic microbial consortium. *Science* **349,** 986–989 (2015).

12. Tanji, Y., Asami, K., Xing, X.-H. & Unno, H. Controlled expression of lysis genes encoded in T4 phage for the gentle disruption of *Escherichia coli* cells. *Journal of Fermentation and Bioengineering* **85,** 74–78 (1998).
13. Huh, J. H., Kittleson, J. T., Arkin, A. P. & Anderson, J. C. Modular design of a synthetic payload delivery device. *ACS Synth Biol* **2,** 418–424 (2013).
14. Din, M. O. *et al.* Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536,** 81–85 (2016).
15. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
16. Kachroo, A. H., Kancherla, A. K., Singh, N. S., Varshney, U. & Mahadevan, S. Mutations that alter the regulation of the chb operon of *Escherichia coli* allow utilization of cellobiose. *Mol Microbiol* **66,** 1382– 1395 (2007).
17. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403,** 339–342 (2000).
18. Wu, M. *et al.* Engineering of regulated stochastic cell fate determination. *Proc Natl Acad Sci USA* **110,** 10610–10615 (2013).
19. Egbert, R. G. & Klavins, E. Fine-tuning gene networks using simple sequence repeats. *Proc Natl Acad Sci USA* **109,** 16817–16822 (2012).
20. Cascales, E. *et al.* Colicin biology. *Microbiol Mol Biol Rev* **71,** 158–229 (2007).
21. Bokinsky, G. *et al.* Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc Natl Acad Sci USA* **108,** 19949–19954 (2011).
22. Zhang, X. Z., Sathitsuksanoh, N., Zhu, Z. & Percival Zhang, Y. H. One-step production of lactate from cellulose as the sole carbon source without any other organic nutrient by recombinant cellulolytic *Bacillus subtilis*. *Metab Eng* **13,** 364–372 (2011).
23. Zhang, X. Z., Sathitsuksanoh, N. & Zhang, Y. H. Glycoside hydrolase family 9 processive endoglucanase from *Clostridium phytofermentans*: heterologous expression, characterization, and synergy with family 48 cellobiohydrolase. *Bioresour Technol* **101,** 5534–5538 (2010).
24. Liao, H., Zhang, X. Z., Rollin, J. A. & Zhang, Y. H. P. A minimal set of bacterial cellulases for consolidated bioprocessing of lignocellulose. *Biotechnol J* **6,** 1409–1418 (2011).
25. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J Mol Biol* **126,** 847–857 (1978).
26. Ghosal, D., Sommer, H. & Saedler, H. Nucleotide sequence of the transposable DNA-element IS2. *Nucleic Acids Res* **6,** 1111–1122 (1979).
27. Johnsrud, L. DNA sequence of the transposable element IS1. *Mol. Gen. Genet.* **169,** 213–218 (1979).
28. Hardin, G. The tragedy of the commons. *Science* **162,** 1243–1248 (1968).
29. Kerr, B., Neuhauser, C., Bohannan, B. J. M. & Dean, A. M. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature* **442,** 75–78 (2006).
30. Gore, J., Youk, H. & van Oudenaarden, A. Snowdrift game dynamics and facultative

cheating in yeast. *Nature* **459,** 253–256 (2009).
31. Sousa, A., Bourgard, C., Wahl, L. M. & Gordo, I. Rates of transposition in *Escherichia coli*. *Biology Letters* **9,** 20130838–20130838 (2013).
32. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R. & Pósfai, G. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microb Cell Fact* **11,** 11 (2012).
33. Tyo, K. E. J., Ajikumar, P. K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nat Biotech* **27,** 760–765 (2009).
34. Babson, D. M., Held, M. & Schmidt Dannert, C. Designer microbial ecosystems – toward biosynthesis with engineered microbial consortia. *Natural Products: Discourse, Diversity, and Design*
35. 23–38 (John Wiley & Sons, Inc., 2014). doi:10.1002/9781118794623.ch2
36. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472,** 499–503 (2011).
37. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* **26,** 483–489 (2008).
38. Bernstein, H. C. & Carlson, R. P. Microbial consortia engineering for cellular factories: *in vitro* to *in silico* systems. *Comput Struct Biotechnol J* **3,** e201210017 (2012).

CHAPTER 3

Microbial single-cell RNA sequencing by split-pool barcoding

3.1 ABSTRACT

Single-cell RNA-sequencing (scRNA-seq) has become an essential tool for characterizing multi-celled eukaryotic systems but current methods are not compatible with bacteria. Here, we introduce microSPLiT, a low cost and high-throughput scRNA-seq method that works for gram-negative and gram-positive bacteria and can resolve transcriptional states that remain hidden at a population level. We applied microSPLiT to >25,000 *Bacillus subtilis* cells sampled from different growth stages, creating a detailed atlas of changes in metabolism and lifestyle. We not only retrieve detailed gene expression profiles associated with known but rare states such as competence and PBSX prophage induction, but also identify novel and unexpected gene expression states including heterogeneous activation of a niche metabolic pathway in a subpopulation of cells. microSPLiT empowers high-throughput analysis of gene expression in complex bacterial communities.

3.2 INTRODUCTION

Gene expression in bacteria is highly heterogeneous even in isogenic populations grown under the same lab conditions. Using bet-hedging strategies, bacteria differentiate into subpopulations that assume different roles for the survival of the community[1,2]. For example, gene expression programs governing developmental and stress-response states such as competence or antibiotic resistance may switch on stochastically in a small number of single cells[3–5]. Population level gene expression measurements are insufficient to resolve such rare states which, to date, have been characterized only in tractable model systems and through single-cell methods such as fluorescence microscopy that can only measure a limited set of reporter genes at a time[6].

Single-cell RNA-seq (scRNA-seq) methods developed for use with eukaryotic cells can provide comprehensive gene expression profiles for tens of thousands of cells[7–11]. However, although the need for microbial scRNA-seq has been recognized, early attempts have been limited to relatively small cell numbers because of the technical challenges associated with adapting scRNA-seq technology to microbes (Figure 3.1a)[12–14]. Specifically, bacteria have very low mRNA content, typically about two orders of magnitude less than human cells[13]. Separation of mRNA from rRNA is challenging because bacterial mRNA is not polyadenylated. Bacteria have a wide diversity of cell walls and membranes which interfere with the lysis or permeabilization steps required for scRNA-seq and their small size can hinder microfluidic single-cell isolation. Here, we present microSPLiT (Microbial Split-Pool Ligation Transcriptomics), a scRNA-seq platform that can overcome these challenges and become a transformative technology for microbiology research.

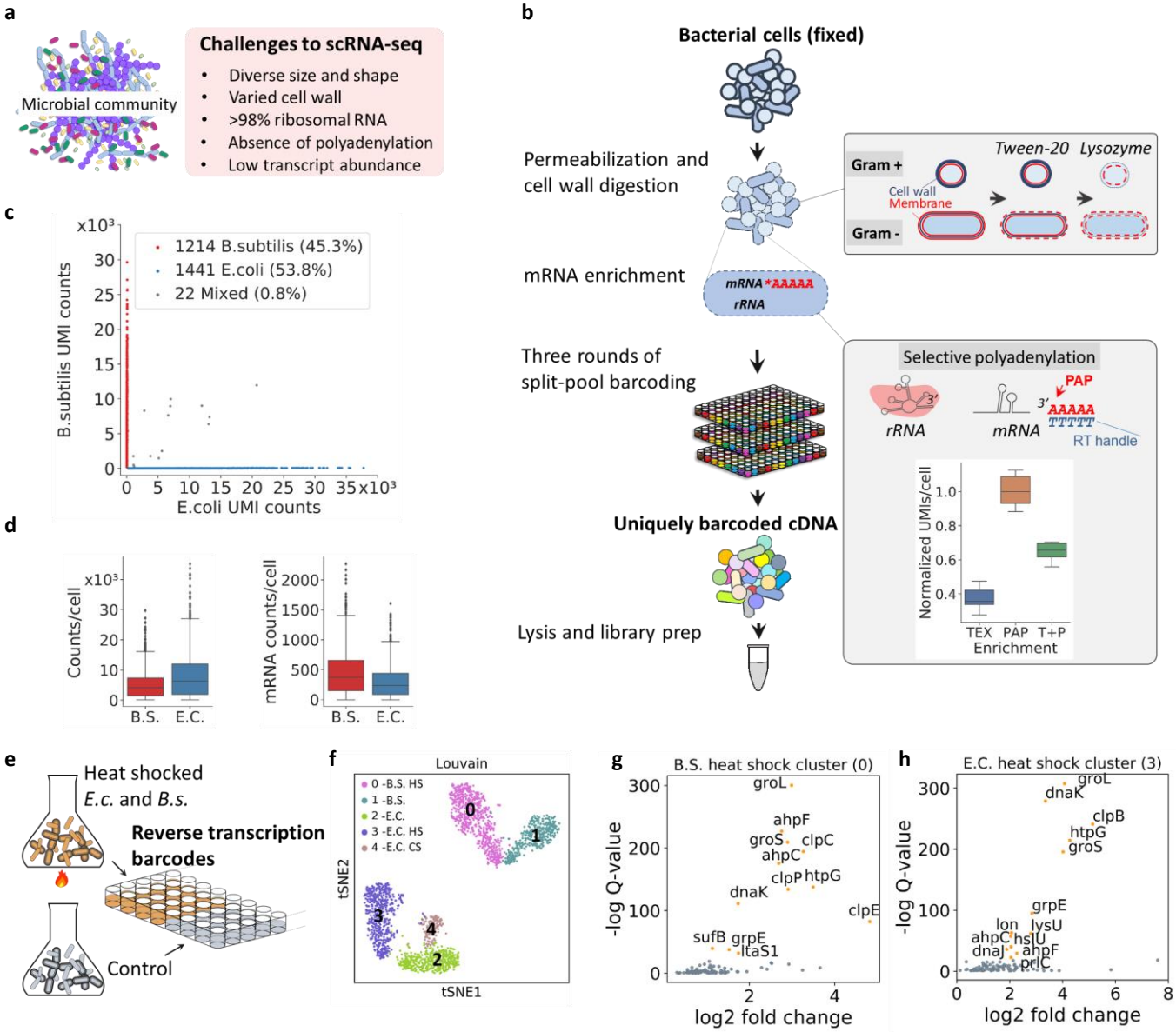


Figure 3.1. microSPLiT development and validation. (**a**) Challenges associated with single-cell RNA sequencing of bacteria. (**b**) microSPLiT method summary. Fixed gram-positive and gram-negative bacterial cells are permeabilized with a mild detergent Tween-20 and their cell walls are degraded by lysozyme. The mRNA is then polyadenylated in-cell with *E.coli* Poly(A) Polymerase I (PAP). Inset: comparison of mRNA read counts (normalized) obtained from PAP-treated cells compared with cells treated with Terminator 5' phosphate-dependent exonuclease (TEX) and both methods consecutively (T+P). The cellular RNA then undergoes three rounds of combinatorial barcoding including in-cell reverse transcription (RT) and two in-cell ligation reactions, followed by lysis and library preparation. (**c**) Barnyard plot for the *E. coli* and *B. subtilis* species-mixing experiment. Each dot corresponds to a putative single-cell transcriptome. UMI – unique molecular identifier. (**d**) Total (in thousands) and mRNA UMI counts per cell for both species. (**e**) Summary of the heat shock experiment showing the first barcode as sample identifier. (**f**) t-stochastic neighbor embedding (t-SNE) of the data from heat shock experiment showing distinct clusters. "HS" – heat shocked. (**g**) and (**h**) Genes enriched in the B.S and E.C. heat shock clusters, respectively.

3.3 DEVELOPING microSPLiT

SPLiT-seq, which labels the cellular origin of RNA through combinatorial barcoding, provides a starting point for bacterial single cell transcriptomics[7]. In SPLiT-seq, cells are fixed, permeabilized and mRNA is converted to cDNA through in-cell reverse transcription (RT) with barcoded poly-T and random hexamer primers in a multi-well format. Cells are then pooled, randomly split into a new 96-well plate, and a well-specific barcode is appended to the cDNA through ligation (Figure 3.1 – figure supplement 1**)**. This split-ligation-pool cycle is repeated and a fourth, optional barcode is added during sequencing library preparation to ensure that each cell acquires a unique barcode combination with very high likelihood (Figure 3.1b).

Because SPLiT-seq does not require cell isolation, it is compatible with a wide range of cell shapes and sizes. Moreover, because SPLiT-seq already uses random hexamer primers in addition to poly-T primers for RT, we reasoned that it might be suitable for detecting bacterial mRNA. However, a direct application of the mammalian SPLiT-seq protocol to bacteria, not surprisingly, resulted in low total UMI counts (<100 max UMIs/cell, median 0 mRNA reads/cell) and a bias toward gram-negative over gram-positive bacteria (Figure 3.1 – figure supplement 2a).

Next, we set out to optimize sample processing steps for bacteria. We took advantage of SPLiT-seq's multiplexing capabilities to test a wide range of approaches to cell wall removal and membrane permeabilization. We settled on a combination of a mild detergent, Tween-20, and lysozyme, as that treatment protocol demonstrated the best capture for both the gram-positive and gram-negative bacteria tested. (Figure 3.1b, inset, and Table 3.1). Then, we compared different methods that would be

compatible for in-cell mRNA enrichment. We tested polyadenylation with *E.coli* Poly(A) Polymerase I (PAP) which was previously used to preferentially polyadenylate mRNA [15], 5'-phosphate-dependent exonuclease ("Terminator", Epicentre) treatment and reverse transcription with ribosomal RNA-specific probes followed by RNaseH-mediated degradation (Figure 3.1 — figure supplement 2c,d and Table 3.1). We found that the treatment of fixed and permeabilized cells with PAP resulted in the highest (about 2.5-fold, or approximately 7% of total RNA) enrichment of mRNA reads (Figure 3.1, inset, and figure supplement 2c,d). We also optimized the fixation protocol as well as the downstream enzymatic reaction conditions (Table 3.1 and Materials and Methods**).** Notably, we found that RT resulted in cell clumping and that mild sonication after this step was necessary to reliably obtain single cell suspensions (Figure 3.1 — figure supplement 2b).

3.4 microSPLiT GENERATES HIGH-QUALITY SINGLE-CELL RNA seq DATA

In order to validate microSPLiT performance on a mixture of gram-positive and gram-negative organisms, we grew *Escherichia coli* MW1255 and *Bacillus subtilis* PY79 cells to OD0.5 and subjected half of each culture to a brief 47C heat-shock. We performed a microSPLiT experiment on both samples, using the first barcode as a sample identifier. We prepared and sequenced a cDNA library from 2677 total bacteria from heat-shocked and control treatments and aligned the reads to a combined *B. subtilis-E. coli* genome. 99.2% of the putative single cell transcriptomes were unambiguously assigned to a single species (Figure 3.1c). We sampled a median of 237 unique mRNA transcripts per cell for *E. coli* and 376 for *B. subtilis*, or approximately 5-10% of the

estimated total mRNA [16]. In total, we detected 3717 genes for *E. coli* and 3476 genes for *B. subtilis* (Figure 3.1d).

Next, we tested whether microSPLiT could detect transcriptional responses to heat shock (Figure 3.1e). Unsupervised clustering identified five distinct clusters which were visualized by t-distributed stochastic neighbor embedding (t-SNE) (Figure 3.1f). The first barcode identified two pairs of clusters corresponding to the heat treated and control cultures, and gene expression analysis within each pair further labeled them as corresponding to *B. subtilis* and *E. coli* cells (Figure 3.1 – figure supplement 3a). Analysis of genes enriched within each cluster revealed classical heat shock genes differentially expressed in each of the E.C. and *B. subtilis* heat treated clusters. We detected induction of genes for abundant class I-IV and VI *B. subtilis* heat shock genes such as *groEL, dnaK, clpC* and *htpG* operons (Figures 3.1g, figure supplement 3) as well as both of the major chaperone systems of E.C. (*groEL* and *dnaK* operons, Figures 3.1h, figure supplement 3b-d)[17,18].

Unexpectedly, we found an additional small cluster, representing *E. coli* cells from both control and heat-treated samples that expressed a dramatically different signature of DEAD-box helicase *deaD* induction as well as cold shock genes *cspA-G* consistent with a transcriptional response to cold (Figures 3.1f and figure supplement 3b)[19]. This subpopulation of *E. coli* might be displaying a very rapid response to cold from a brief cold centrifugation step performed as the first step in sample preparation before formaldehyde fixation.

3.5 TRANSCRIPTIONAL PATTERNS DURING *B. subtilis* GROWTH IN RICH MEDIA

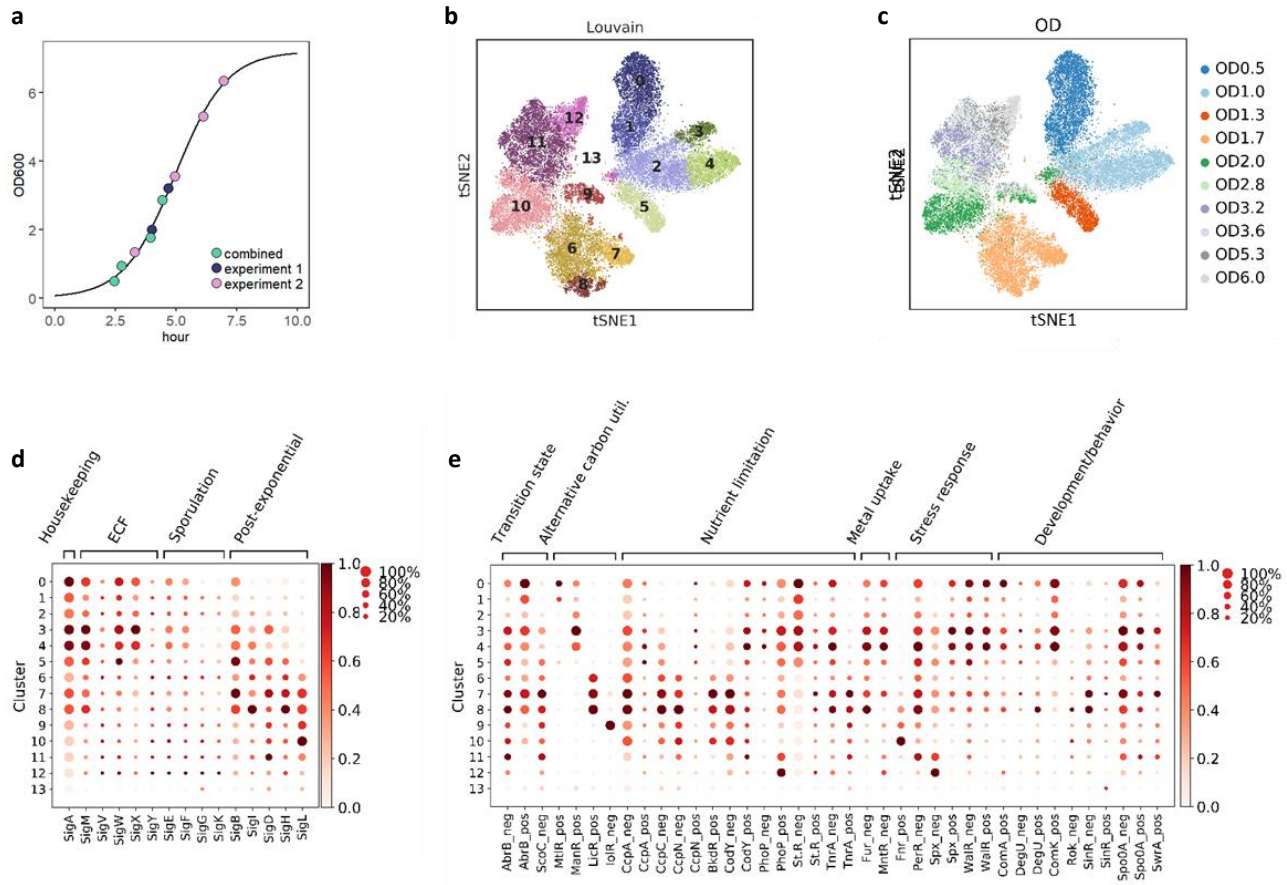


Figure 3.2. microSPLiT detects global transcriptional states during *B. subtilis* growth. **(a)** Optical density (OD) points sampled in two experiments overlayed on the growth curve from experiment 2 (see Supplementary Text). (**b**) t-SNE embedding of the combined growth curve data colored by cluster. A total of 14 clusters were found using Louvain clustering. (**c**) Inferred normalized sigma factor activity for each cluster. The size of each dot indicates the proportion of cells in the cluster in which the sigma factor is active, while the color indicates the average activity. (**d**) Inferred normalized activity of select transcriptional regulators per cluster. "Neg" indicates that activity was calculated for the genes known to be negatively regulated by this TR, and "pos" indicates the activity was calculated for the genes positively regulated by the given TR.

Next, we applied microSPLiT to capture transcriptional states across the *B. subtilis* growth curve in a rich medium (LB). In total, we sampled ten optical density (OD) points along the growth curve of the laboratory strain PY79 ranging from OD 0.5 (early exponential phase) to 6.0 (early stationary phase). Four of the OD points were sampled in both replicates of microSPLiT while the rest were sampled only in one of the experiments (Figure 3.2a). We retained all optical density measurements, however, the time data from the first experiment was not recorded. We fit the growth curve in Figure 2 to the time-stamped OD samples from the second experiment using the formula for a sigmoidal curve: $L/(1+e^{-k(t-t_0)})$ To plot the OD points from the first experiment, we set the fit to the recorded OD curve and solved for time. In both experiments, the first barcode was used to record sample identity (i.e. OD). The data from two different experiments are consistent and produced a combined dataset of 25,214 cells (Figures 3.2b, figure supplements 1,2). Unsupervised clustering of the combined datasets revealed 14 clusters, most of which contained cells predominantly from a single OD (Figure 3.2b). The most notable exceptions are two smaller clusters that contain cells from multiple ODs: a cluster with cells from OD2-3.2 that differentially express myo-inositol metabolism pathway genes, and a very small cluster containing only 36 cells from 5 different OD points uniquely expressing genes associated with the defective PBSX prophage (Figures 3.2b, figure supplement 5).

Alternative sigma factors are the primary regulators of the prokaryotic RNA polymerase specificity and thus directly shape transcriptional changes in response to environmental conditions. To understand whether microSPLiT could capture variation in sigma factor utilization across different growth stages, we averaged activity of genes

regulated by each sigma factor, recording, for each cluster, both the percentage of cells expressing at least one gene regulated by a given sigma factor and the average intensity of gene expression (Figure 3.2c). We note that while averaged intensity could also be obtained from bulk RNA-seq, only single-cell methods can reveal information about cell-to-cell variation in regulon activity.

Consistent with expectations, we observe that the housekeeping σ^A activity is highest at OD0.5, while the activity of general stress response sigma factor σ^B rises as cells begin to exit from exponential phase (OD1.3-1.7) and then declines as cells approach stationary. Sporulation sigma factors σ^F, σ^J and σ^K were induced at later ODs but in a small fraction of cells, consistent with the previously reported heterogeneous initiation of sporulation[3,20]. The extracellular function (ECF) sigma factors σ^M, σ^W and σ^X implicated in maintaining cell envelope function reached maximal activity at OD 1.0 in a large proportion of cells, consistent with reports of their basal activity in logarithmic phase in non-stressed cells[21]. Meanwhile, the remaining ECF sigma factors σ^V, raising defenses against lytic endoglycosidases, and σ^Y increased in activity towards later OD points in a small subpopulation of cells, similar to the sporulation sigma factors. σ^I and σ^H activities, regulating heat response and post-exponential behavior respectively, peak in cluster 8 which represents a subgroup of cells at OD 1.7. In contrast, σ^B and σ^D regulating general stress response and motility are most active in cluster 7, a second distinct subgroup of OD 1.7 cells. Finally, σ^L implicated in utilization of arginine, acetoin and fructose as well as regulation of the cold shock response peaks in cells at OD 2-2.8 represented in cluster 10, most likely due to the highly enriched acetoin utilization genes in this cluster. Additionally, we found that correlations between the sigma factor

regulons largely agreed with the concept of molecular time sharing, i.e. the idea that sigma factors compete for RNA polymerases (Figure 3.2 – figure supplement 3)[22].

To obtain an even finer-grained picture of the transcriptional programs during exponential growth and entry to stationary phase, we inferred the activity profiles of select transcriptional regulators (TR) from expression of the genes in their respective regulons (Figure 3.2d, figure supplement 4)[23]. This analysis revealed pronounced changes in regulation of carbon utilization, stress responses, metal uptake, developmental decisions and more. The main transition state regulator AbrB becomes inactive after OD1.0, indicating that the preferred carbon sources such as glucose start getting depleted. In response, cells begin to activate transcriptional programs to utilize different alternative carbon sources. Toward the intermediate growth stages, the cells start to sense carbon, nitrogen and phosphate limitation. Clusters 7 and 8 (OD 1.7) display a striking change in carbon metabolism indicated by the strong expression of genes repressed by the carbon catabolite control proteins CcpA, CcpC and CcpN. Similarly, the regulator of nitrogen assimilation TnrA becomes activated at OD1.7, while PhoP, regulating the phosphate metabolism, becomes active at three different growth stages: OD1.0, OD1.7 and later on at OD6.0. In addition, cells respond to metal deficiency, switching off the negative regulators of iron and manganese uptake Fur and MntR after OD1.0. Finally, we can observe cellular response to a variety of intrinsic and cell-envelope stresses, as well as temporal activation patterns of a battery of developmental regulators including ComA, SinI, DegU, Rok, Spo0A and others. Surprisingly, we also observe an upregulation of ComK-regulated genes in a high proportion of cells in the early ODs which is not consistent with the primary role of this transcription factor in a rare developmental state

of competence. It could be explained by the fact that a large cohort of ComK-induced genes are involved in metabolism and DNA repair and can be activated by other regulators. These data show that microSPLiT not only captures known regulatory programs but also reveals intriguing heterogeneity in a wide range of these pathways.

Central carbon metabolism changes and differential expression of tricarboxylic acid (TCA) cycle enzymes

Given the dramatic changes in regulation of carbon utilization observed in our TR analysis, we turned to a more comprehensive examination of carbon metabolism genes enriched in each cluster (Figure 3.2 — figure supplement 5, Figure 3.3). When glucose and other preferred sugars (discussed below) are present, they are converted to pyruvate during glycolysis, which is the primary metabolic route under excess of these sugars. In these conditions promoting rapid growth, pyruvate is then converted to lactate, acetate, acetoin, and other by-products through fermentation. Upon depletion of preferred sugars, cells redirect the fermentation by-products to be metabolized in the TCA cycle generating additional adenosine triphosphate (ATP) and carbon dioxide.

In the exponential phase (Clusters 0 and 1, OD0.5) we find high expression of *ptsG,* a glucose permease which transports and phosphorylates glucose. The enzymes in the *gapA* operon constituting the metabolic pathway from glyceraldehyde-3P to phosphoenolpyruvate (PEP): *gapA, pgk, pgm* and *eno*[24] were upregulated in this cluster (Figure 3.3a,b).

Next, in clusters 3 and 4 (OD1.0) we observe transcriptional patterns suggesting an increase in flux from pyruvate either being converted to lactate by *ldh*[25] which is then exported via a malate antiporter *mleN*[26] or converted to acetate via intermediates by

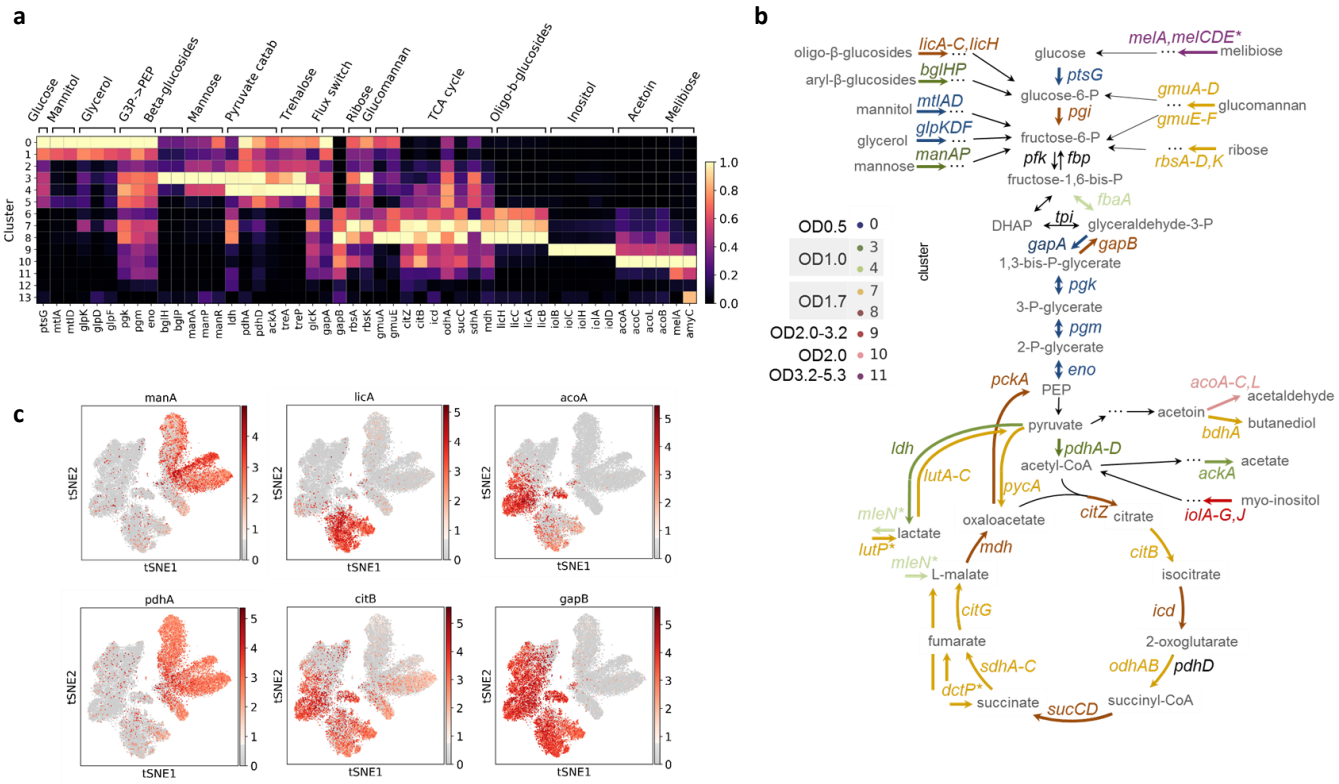


Figure 3.3. Central carbon metabolism changes and alternative carbon sources utilization during *B. subtilis* growth. **(a)** Normalized expression of genes from select metabolic pathways and central carbon metabolism shown per cluster. Gene expression clearly shows distinct carbon utilization programs associated with different clusters and growth states. (**b**) Expression of select genes from (A) overlaid on the t-SNE plot to illustrate the differential patterns of activation. (**c**) Schematic of the central carbon metabolism pathway showing alternative carbon sources, metabolic products and genes in the pathway. The genes are color-coded according to the cluster they are highest expressed in.

pdhAD[27] and *ackA*[28] (Figures 3.3a,b). These observations are consistent with previous reports of transient medium acidification via acetate production during rapid fermentative *B. subtilis* growth[29].

At OD1.7 cells appear to undergo a dramatic transition from glycolysis to gluconeogenesis with multiple genes from the gluconeogenetic pathway activated in clusters 7 and 8. There are two glyceraldehyde-3-phosphate dehydrogenases in *B. subtilis*: GapA and GapB, mediating the flux of carbon either from glucose to the TCA cycle or vice versa[30]. The glucose and intermediates generated by the gluconeogenetic pathway under conditions of glucose limitation are then used for synthesis of necessary structural constituents. We observe the switch from GapA to GapB expression in clusters 7 and 8 along with an upregulation of most of the TCA cycle enzymes. We also find a different pattern of pyruvate production and utilization, together with catabolism of acetoin, another fermentation product, and additional nutrient fluxes into the TCA cycle (Figures 3.3a-c, Figure 3.2 – figure supplement 5). In cluster 7, in contrast to clusters 3 and 4, instead of excretion we now observe uptake of lactate via *lutP* and conversion to pyruvate by *lutAC*[31]. The conversion of pyruvate to oxaloacetate is also upregulated by increased expression of *pycA*[32]. In addition, acetoin, a product of acetate metabolism, in this cluster is actively converted into butanediol by action of *bdhA*[33]. Cells in this cluster also express more *dctP* implicated in direct import of TCA cycle intermediates succinate, fumarate, malate and oxaloacetate[34]. In cluster 8, we additionally find enrichment of *pckA* converting oxaloacetate to PEP (Figures 3.3c) [35].

Interestingly, the TCA cycle enzymes which are collectively enriched in the entire OD1.7 sample display differential transcriptional abundances between clusters 7 and 8

(Figure 3.3a,b). For instance, cells in cluster 7 express more aconitase (*citB*), 2-oxoglutarate dehydrogenase (*odhAB*), and succinate dehydrogenase (*sdhABC*), while cells from cluster 8 display higher expression of genes in the *citZ-icd-mdh* operon and *sucCD* (succinyl-CoA synthetase). CitB was reported to destabilize the *citZ-icd-mdh* mRNA upon citrate accumulation or iron limitation[36]. Thus, differential transcript abundance of the TCA cycle enzymes could reflect the regulatory interactions between the respective genes.

Transient activation of alternative carbon utilization pathways in distinct subpopulations

Complementing changes in the core carbon metabolism, we observe expression of pathways responsible for uptake and utilization of a variety of different carbon sources. As the preferred sources of carbon are depleted, the major repressor of alternative carbon utilization pathways CcpA becomes inactive, permitting the cells to catabolize alternative carbon sources including a variety of carbohydrates[37] (Figure 3.3). We find that the activation and suppression of these pathways happen in varying proportions of the cells in each OD sample and appear to follow a temporal order (Figures 3.3, figure supplement 2). The preferred carbon sources for *B. subtilis* are glucose along with other Group A sugars (sucrose, fructose, and mannitol) and malate[38,39]. While the latter is directly metabolized in the TCA cycle, the former sugars are converted to one of the glucose metabolic intermediates. At OD0.5, in addition to glucose-specific *ptsG*, we observe increased expression of *mtlA-D* and *glpK,D,F* responsible for utilization of mannitol and glycerol, respectively. Cells in clusters 3 and 4 (OD1.0) activate catabolism of mannose and aryl-β-glucosides (*manA,P* and *bglH,P*). In cluster 7 (OD1.7) we observe

the upregulation of genes for utilization of glucomannan (*gmuA-F*) and the ribose transporter (*rbsA-D*), while the gene for utilization of ribose *rbsK*, curiously, is upregulated much earlier in cluster 3. Finally, at even later ODs three additional alternative carbon source utilization programs switch on. Cluster 9 comprising cells from ODs 1.7, 2.0, 2.8, and 3.2 is defined by the expression of genes implicated in the most common stereoisomer of inositol, myo-inositol catabolism (*iolABCDEFGHIJ,* further "*iolAJ*"), while cluster 10 (OD2.0) is enriched for genes responsible for utilization of acetoin (*acoABCL*). Finally, cluster 11, representing a range of ODs from 3.2 to 5.3, differentially expresses genes for melibiose utilization (*melA*, *melCDE*).

Heterogeneous activation of myo-inositol catabolism pathway at intermediate growth stages

Inositol is an abundant resource in soil, and *B. subtilis* is able to subsist on inositol as its sole carbon source[40]. While LB medium is not typically expected to contain myo-inositol (further "inositol"), heterogeneous inositol utilization pathway activation is observed in a small (3-15%) subpopulation in both of our independent LB growth experiments (cluster 9, see Figure 3.3a, Figure 3.4a,b and Figure 3.4 — figure supplement 1). It has been shown that the inositol catabolism intermediate, 2-deoxy-5-keto-D-gluconic acid 6-phosphate (DKGP), is responsible for the pathway induction[40]. We hypothesize that the trace amounts of inositol may be present in the LB medium, potentially from the yeast extract since yeast is capable of inositol production as a precursor to the essential membrane component, phosphatidylinositol.

There are three operons involved in inositol utilization, *iolT* (main transporter), *iolRS* (the first gene is a repressor and the second is a likely dehydrogenase), and *iolAJ*

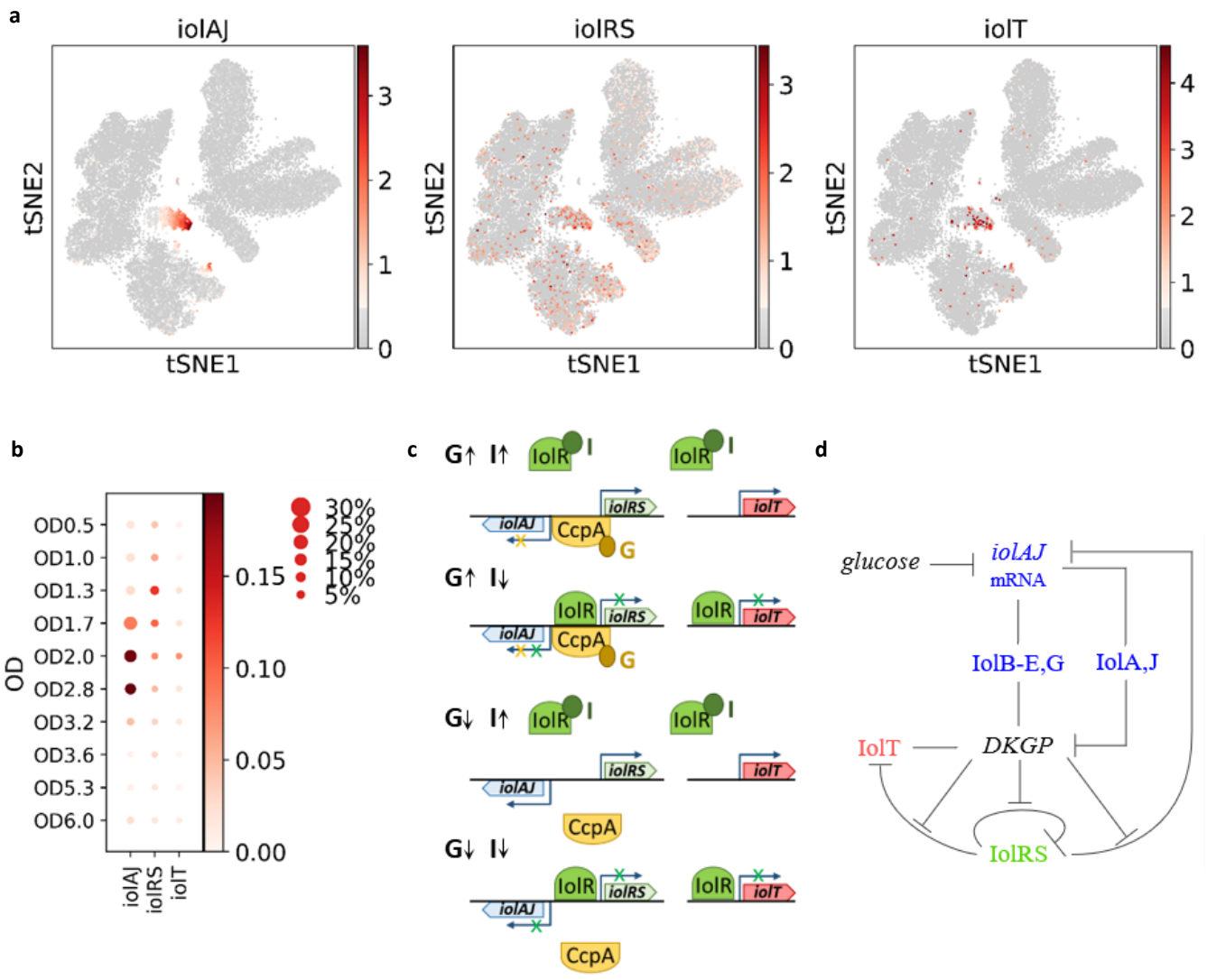


Figure 3.4. Myo-inositol utilization pathway genes are expressed heterogeneously at intermediate growth stages. **(a)** Expression of each of the three inositol utilization operons, averaged across all genes in a given operon, and overlaid on the t-SNE plot. (**b**) Activities of the three inositol utilization operons across ODs. The size of each dot indicates the proportion of cells in each OD sample expressing any of the genes in the selected operon, while the color shows the average expression of the genes in a given operon. (**c**) Schematic of the inositol utilization operons structure along with the main regulators in different metabolic states. When both glucose (G) and inositol (I) are present in the cell, the IolR repression of all three operons is relieved, but CcpA still represses the *iolAJ* operon preventing it from metabolizing inositol (we show the signaling intermediates bound to the repressors also as G and I for simplicity). As glucose levels drop, CcpA no longer inhibits transcription of *iolAJ*, and if inositol is present, it relieves the repression of *iolRS* and *iolT*, inducing the expression of the entire metabolic network. Meanwhile, in the absence of inositol, all of the operons remain repressed regardless of the glucose levels. (**d**) A simplified network diagram representing the interactions and regulations within the inositol utilization pathway.

(metabolic enzymes), with *iolC* producing and *iolJ* cleaving the pathway-activating DKGP intermediate. *iolRS* and *iolAJ* are normally transcribed by σ^A through divergent transcription[41]. In the absence of the inducer (I), IolR supresses transcription of all three operons[42]. In addition, CcpA represses the *iolAJ* operon in the presence of glucose[43] (Figure 3.4c). Interestingly, the pathway suppressor *iolR* gene is more broadly expressed both outside and inside of cluster 9 (Figure 3.4a,b and figure supplement 1).

Next, we asked whether our data could be explained by the underlying gene regulatory network architecture (Figure 3.4d). Within the complex system of interactions, we observe two topological features capable of amplifying small molecular variations. An *iolAJ*-mediated incoherent feed-forward loop can generate a pulse of DKGP, while IolT can generate a positive-feedback loop by transporting inositol into the cell. These features of the network can explain the difference in clusters, as probed with a qualitative model (Figure 3.4 – figure supplement 2). Due to the small proportion of cells expressing the main catabolic operon *iolAJ*, this particular metabolic behavior could only be reliably detected on a single-cell level.

Motility, antimicrobials production, stress response, and metal ion import

B. subtilis is known to exhibit a variety of behaviors to enhance survival in adverse conditions including, but not limited to, production of degradative enzymes and antimicrobials, secretion and uptake of siderophores, three types of motility (swimming, swarming, and sliding), natural competence, and sporulation[44]. Many if not most of these behaviors are not displayed by all cells at the same time [45], thus providing good targets for single-cell interrogation.

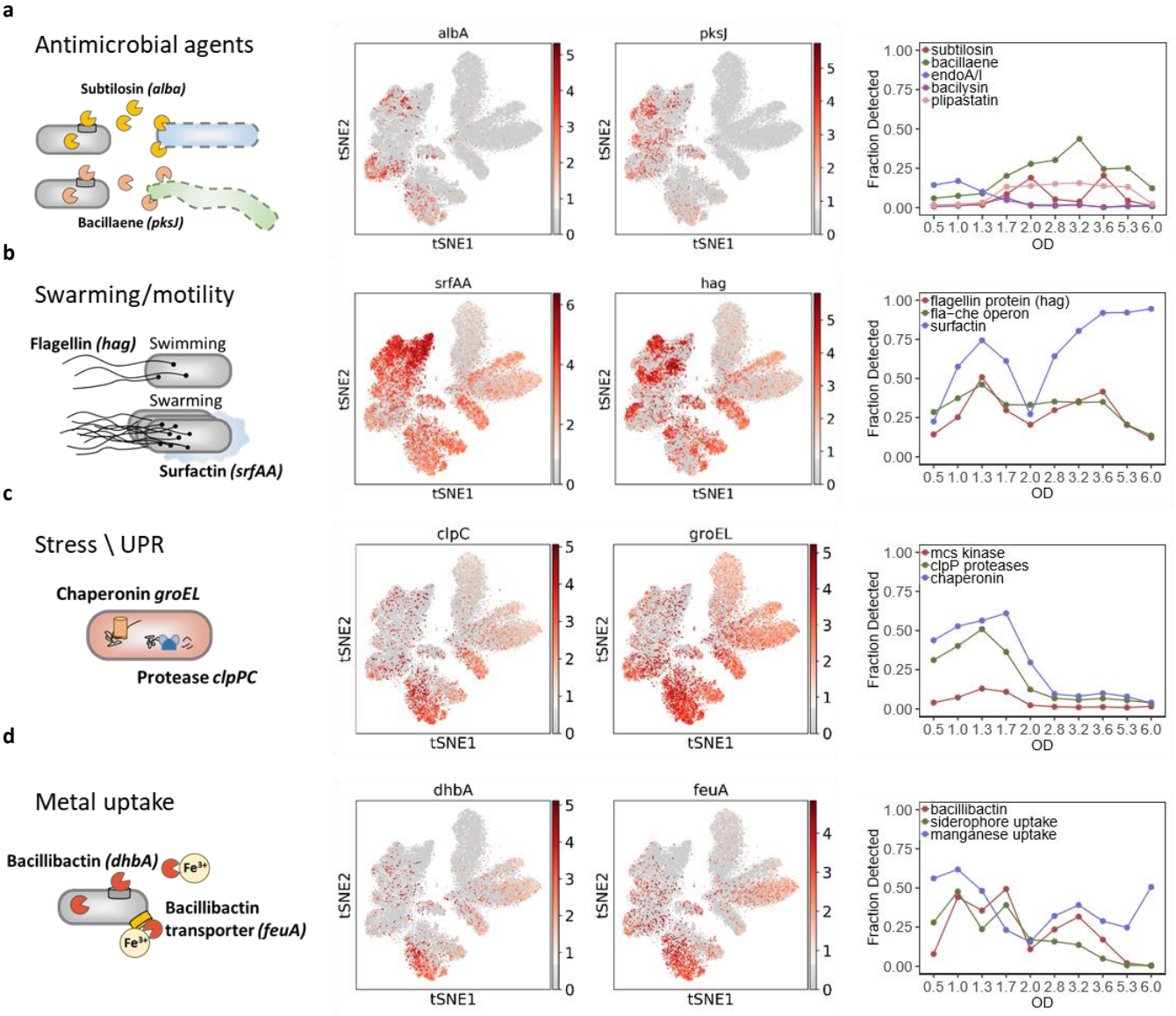


Figure 3.5. Intrinsic stress responses and developmental gene expression. **(a)** Antimicrobial agents and endoA toxin-antitoxin secretion. Subtilosin and bacillaene action diagram. Overlay of representative subtilosin (*albA*) and bacillaene (*pksJ*) production pathway gene expression on the t-SNE. Fraction of cells expressing the indicated antimicrobials and endotoxin-antitoxin as a function of OD. **(b)** Swarming and motility genes expression. Flagellin and surfactin action diagram. Overlay of surfactin (*srfAA*) and flagellin (*hag*) gene expression on the t-SNE. Fraction of cells expressing the indicated motility operons as a function of OD. **(c)** Intrinsic stress and unfolded protein response (UPR) genes expression. GroEL chaperonin and ClpCP protease action diagram. Overlay of *clpC* and *groEL* gene expression on the t-SNE. Fraction of cells expressing the indicated genes as a function of OD. **(d)** Metal (iron and manganese) uptake genes expression. Bacillibactin siderophore and transporter action diagram. Overlay of bacillibactin (*dhbA*) and siderophore transporter (*feuA*) production pathway gene expression on the t-SNE. Fraction of cells expressing the indicated metal uptake genes as a function of OD. The genes used for fractional plots are listed in Table 3.2.

Bacteria universally produce peptide and small molecule antimicrobials that are meant to target both closely and distantly related organisms (Figure 3.5a)[46]. We observe the expression of endoA toxin/antitoxin (*ndoA/I*) and bacilysin (*bacA-G*) more broadly prior to OD2.0 (Figure 3.5a**)**. Subtilosin (*sboA,X-albA-F*), bacillaene (*pksC-R*), and plipastatin (*ppsA-E*) production, on the other hand, are detected in a greater proportion of cells after OD2.0. All three of these peptides are broad spectrum antimicrobials[47–49] which suggests that under nutrient limitation, *B. subtilis* becomes competitive to preserve its access to the available resources, even in the absence of other species. Additionally, we also see a rise in spore killing factor (SKF) and spore delay protein (SDP) in the last three ODs (Figure 3.5 – figure supplement 1).

B. subtilis has two morphological phenotypes during active growth: filamentous, sessile cells and smaller motile cells[50,51]. In liquid media, the bacteria primarily swim which requires the expression of the flagellin protein (*hag*) and the *fla-che* operon (Figures 3.5b). We detected expression of these genes in 25 to 50% of cells in every cluster from early to intermediate growth stages before the fraction noticeably declines at late ODs, consistent with reports of heterogeneity of the motility operon expression at mid-exponential phase (*42*). The *B. subtilis* population is expected to be similarly differentiated into surfactin-producing and extracellular matrix-producing bacteria as cell density increases[52]. We see the fraction of cells producing *srfA-D* genes gradually increase, slightly dipping at OD 1.7 and 2.0, and finally reaching almost 100% detection at OD 6.0, consistent with the PY79 strain having defective matrix production genes that cannot negatively regulate *srfA-D* expression[53].

Cellular stress response pathways exemplified by GroEL chaperonin and ClpC

proteases peak at OD1.7, the same time as the cells switch from glycolysis to gluconeogenesis (Figure 3.5c). The ClpP associated proteases (*clpP,C,X,E*), McsA and McsB kinases (*mcsA,B*), and chaperonins (*groEL,ES*) are all involved in the unfolded protein response[54]. A transient increase in the regulatory sigma factor, σ^B, inducing expression of these genes, has been reported during normal exponential growth and attributed to intrinsic cellular stresses[55]. *clpP* mutants have been linked to slower glucose consumption and overproduction of TCA cycle metabolites during rapid fermentative growth, which suggests their participation in glycolysis and may explain their highest expression during the early growth stages[56].

B. subtilis needs both manganese and iron to grow[57], and the two ions often antagonize each other's regulators[58]. Upon iron limitation, *B. subtilis* produces and secretes siderophores including bacillibactin (*dhbA*) together with the ABC transporter (*feuA*) (Figure 3.5d)[59,60]. Both genes show increased expression at OD1.7, similar to the stress response proteins above. Manganese, on the other hand, is transported into the cells via a separate ABC transporter (*mntA-D*) and a proton symporter (*mntH*)[58]. It is required for the successful transition into stress related states such as biofilm formation and sporulation[61], which likely explains the increase in manganese transport related genes detected at later ODs.

microSPLiT quantifies a rare stress response

Cluster 13 (36 cells, or 0.142% of total cells, representing ODs between 0.5 and 2.8) contains a rare subpopulation of cells expressing PBSX prophage genes (Figure 3.6a). The PBSX element is a defective prophage that is non-infectious but upon

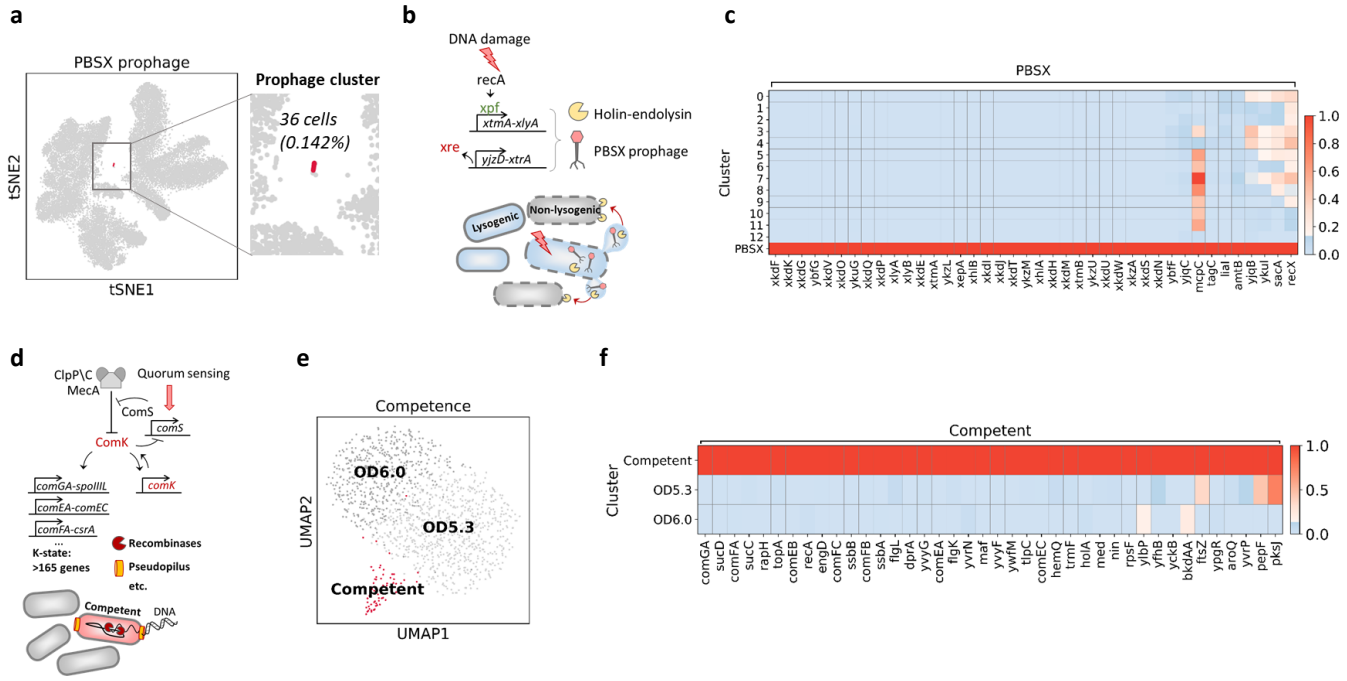


Figure 3.6. Rare developmental states induced by cellular stress. (a) PBSX prophage induction and killing of the non-lysogenic cells. Upon DNA damage, two large prophage operons are induced and non-infectious phage particles along with endolysins are released. The latter degrade cell walls of other cells containing heterologous prophage elements. (**b**) PBSX prophage cluster (36 cells) shown on the t-SNE plot. (**c**) Genes enriched in the PBSX prophage cluster, including both prophage and host genes. (**d**) Overview of competence development. Activated by stress and quorum-sensing, ComS binds to and redirects MecA complex from degrading master competence regulator ComK. Amplified via positive autoregulation, ComK induces >165 genes involved in extracellular DNA uptake and cellular physiology. (**e**) UMAP embedding of the subclustered OD5.3 and 6.0 samples, showing the competence cluster (62 cells). (**f**) Genes enriched in the discovered competence cluster relative to the rest of the cells in OD5.3 and 6.0 samples.

induction causes the release of phage-like particles[62] containing 13 kb of random fragmented chromosomal DNA[63,64]. Prophage gene expression is induced by DNA damage mediated by the SOS response[62,65] (Figure 3.6b). We identified eleven host genes with known or putative functions expressed in the PBSX prophage cluster (Figure 3.6c). Five of these genes have previously been shown to be induced only in PBSX-harboring strains of B. S. after DNA damage[63]. The rest, including a chemoreceptor (*mcpC*), an ATP-binding cassette transporter (*liaL*), a cell wall binding protein (*ykuG*), an ammonium transporter (*amtB*), a sucrose-6-phospate hydrolase (*sacA*), and a regulatory protein of homologous recombination (*recX*), have not previously been linked to prophage induction. Two of the host genes enriched in cluster 13, *mcpC* and *amtB*, are linked to the GlnR regulon[66,67] which responds to excess nitrogen for nitrogen assimilation. Given that the cells in this cluster were largely in stationary phase and therefore in a nutrient limited environment, it is unlikely that there was excess nitrogen to induce this regulon. It is possible that induction of PBSX might lead to expression of the GlnR host regulon, but further research is needed. A similar hypothesis can be formulated for the expression of *sacA*, which is involved in carbohydrate uptake[68,69]. The other genes identified, *liaL*, *ykuG* (syn. *fadG*), and *recX*, have been shown to be involved in the LiaRS membrane damage response, fatty acid degradation response, and homologous recombination respectively[70–72]. These genes could have been induced by the damage sustained by the cells which also caused PBSX prophage activation. Alternatively, these genes could have been activated in response to the chromosomal fragmentation and membrane vesicle formation caused by the prophage induction.

microSPLiT detects a rare, stochastically induced developmental state

Under stress or nutrient limitation, a small fraction (2-5%) of *B. subtilis* cells undergoes stochastic transient differentiation into a state of natural competence, characterized by the ability to uptake extracellular DNA and integrate it into the chromosome (Figure 3.6d)[73]. The master transcriptional regulator of competence ComK is activated via a positive feedback loop, inducing expression of a suite of >165 genes involved in a variety of cellular processes in addition to DNA uptake[74]. Competence is expected to naturally occur under nutrient limitation. We thus separately subclustered the last two OD points (OD5.3 and 6.0). UMAP embedding revealed a small cluster (62 cells, or 4.6% of cells at OD 5.3 and 6) expressing a distinct transcriptional signature of the competent state, or K-state (Figure 3.6e,f). The most enriched gene was *comGA*, as expected from prior transcriptomic data, followed by the succinyl-CoA synthetase (*sucCD*) operon which has been shown to be induced in competent cells[29,74,75]. We also see strong enrichment of other genes encoding the DNA uptake machinery: *comF* and *comE* operons, as well as the response regulator RapH which represses sporulation development in competent cells by dephosphorylating the phosphorelay component Spo0F[76]. As the cells in the competent state arrest replication to integrate the incoming DNA into the chromosome, we observe the upregulation of genes necessary for processing of internalized ssDNA such as DNA recombination and repair protein RecA along with single-strand DNA binding proteins SsbA and SsbB consistent with prior reports[74]. We also observe enrichment of several genes related to DNA processing such as *topA* encoding topoisomerase A and *holA*, delta subunit of DNA polymerase III which is a part of the replisome. These genes are not in the annotated ComK regulon but were

tentatively identified in the microarray data comparing gene expression between *mecA* strain in which essentially all cells express ComK and a double mutant *mecA comK* strain[74]. This coordinated upregulation of *topA* and *holA* is consistent with the model of RecA binding to the SsbA\SsbB coated ssDNA and forming a complex with the replisome during competence[77]. In addition, we found four genes not previously linked to the competent state: *ywfM* (unknown), *hemQ* (coproheme decarboxylase), *tlpC* (an orphan membrane-bound chemotaxis receptor), and *trmF,* a folate- and FAD-dependent tRNA methyltransferase.

Notably, we detected enrichment of *ywfM* and *hemQ* in the competence cluster. *ywfM* whose product is unknown is located very close but in an opposite orientation to *hemQ* encoding a coproheme decarboxylase implicated in heme biosynthesis. It was prior hypothesized that ComK could be activating divergent transcription by binding to the palindromic site between *yfmK* and *hemQ*, however, *yfmK* and *yfmM* were not previously reported to be co-transcribed. In a similar vein, we detected enrichment of *nin* and *tlpC* with the latter previously not reported to be ComK-induced, however, it is chromosomally located next to the *hxlA/hxlR* gene pair on the opposite strands of DNA divergently transcribed by ComK binding between these genes and activating transcription in both directions. *tlpC* encodes an orphan membrane-bound chemotaxis receptor[78]. ComK is known to induce expression of chemotaxis proteins, encoded by the genes *flgL,K,M* in the comFA operon which we also detected in our competence cluster. However, competence phenotype is considered suppressive to motility via action of anti-sigma factor *flgM*. Currently the exact ligand for *tlpC* is not known except *tlpC* being functionally connected to *A. thaliana* root colonization[79] and it is therefore hypothesized

to be involved in other cellular process involving a two-component signal transduction system[78]. *hemQ* product, coproheme decarboxylase, on the other hand, functions in heme biosynthesis. Various intermediary metabolism genes are known to be upregulated by ComK but they were not previously known to include the heme biosynthetic pathway. Finally, another gene not hitherto connected to K-state but enriched in the competence cluster is *trmF*, a folate- and FAD-dependent tRNA methyltransferase involved in tRNA maturation. This finding is in line with another methyltransferase *gidB* acting on 16S rRNA which has been found enriched in competent cells previously[74].

3.6 DISCUSSION

We applied microSPLiT to *B. subtilis* cells growing in liquid rich medium, conditions which are not usually associated with abundant cellular heterogeneity. Nevertheless, we found a large variety of subpopulations displaying differential gene expression of select metabolic, stress response or developmental pathways. In particular, we identified a metabolic pathway, myo-inositol catabolism, which was strikingly activated only in a fraction of cells from several later OD points in a distinct temporal fashion. We expect microSPLiT to have even more utility in identifying heterogeneous cell states in an environment more conducive to bacterial differentiation, such as in multi-species biofilms and natural microbiota.

We were able to detect subpopulations of cells as rare as 0.142%, pointing to microSPLiT's potential to uncover physiologically relevant rare cell states, such as persistence, that are hard to study by bulk or low-throughput methods. For many such states, the regulators are not well known and consequently, the reporters or mutants

producing the desired state at a higher frequency cannot be engineered. Even for states that are better understood and can be artificially induced, such as prophage induction by UV irradiation, microSPLiT is useful since the transcriptional signatures it produces are state-specific and free of artifacts introduced by the perturbation.

In order to use microSPLiT on complex natural communities, the protocol will likely need to be further optimized, particularly the permeabilization and mRNA enrichment steps. Since cell wall and membrane composition vary significantly among bacteria, any permeabilization treatment is likely to work better for some species than others. However, alternate treatment for different subsamples may still provide optimal results. In addition, we experienced lower mRNA counts from bacteria in stationary phase as opposed to logarithmic growth phase, consistent with slower growth rate and smaller cell size at this stage. Although the resulting data were still sufficient to reliably identify rare cell states such as the K-state, further improvement of the protocol will be instrumental to increase sensitivity for applications to slower dividing bacteria or challenging environmental conditions. Still, we expect microSPLiT to provide an exciting new dimension to studies of bacterial gene expression heterogeneity and community behavior facilitated by the method's potential scalability to millions of bacterial cells and single-cell resolution without need for constructing reporters.

3.7 REFERENCES

1. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).
2. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).
3. Kuchina, A. *et al.* Temporal competition between differentiation programs determines cell fate choice. *Molecular Systems Biology* **7**, 557–557 (2014).
4. Elowitz, M. B. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186 (2002).
5. Russell, J. R., Cabeen, M. T., Wiggins, P. A., Paulsson, J. & Losick, R. Noise in a phosphorelay drives stochastic entry into sporulation in Bacillus subtilis. *EMBO J.* **36**, 2856–2869 (2017).
6. Locke, J. C. W. & Elowitz, M. B. Using movies to analyse gene circuit dynamics in single cells. *Nat. Rev. Microbiol.* **7**, 383–392 (2009).
7. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* eaam8999 (2018) doi:10.1126/science.aam8999.
8. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
9. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187–1201 (2015).
10. Fan, H. C., Fu, G. K. & Fodor, S. P. A. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367–1258367 (2015).
11. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
12. Wang, J., Chen, L., Chen, Z. & Zhang, W. RNA-seq based transcriptomic analysis of single bacterial cells. *Integr Biol (Camb)* **7**, 1466–1476 (2015).
13. Kang, Y. *et al.* Transcript amplification from single bacterium for transcriptome analysis. *Genome Res.* **21**, 925–935 (2011).
14. Liu, Y. *et al.* Bacterial Single Cell Whole Transcriptome Amplification in Microfluidic Platform Shows Putative Gene Expression Heterogeneity. *Anal. Chem.* **91**, 8036–8044 (2019).
15. Wendisch, V. F. *et al.* Isolation of Escherichia coli mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. *Anal. Biochem.* **290**, 205–213 (2001).
16. Bartholomäus, A. *et al.* Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos Trans A Math Phys Eng Sci* **374**, (2016).
17. Arsène, F., Tomoyasu, T. & Bukau, B. The heat shock response of Escherichia coli. *International Journal of Food Microbiology* **55**, 3–9 (2000).
18. Schumann, W. The Bacillus subtilis heat shock stimulon. *Cell Stress Chaperones* **8**, 207–217 (2003).
19. Thieringer, H. A., Jones, P. G. & Inouye, M. Cold shock and adaptation. *BioEssays* **20**, 49–57 (1998).
20. Chastanet, A. *et al.* Broadly heterogeneous activation of the master regulator for sporulation in Bacillus subtilis. *Proceedings of the National Academy of Sciences*

107, 8486–8491 (2010).
21. Luo, Y., Asai, K., Sadaie, Y. & Helmann, J. D. Transcriptomic and Phenotypic Characterization of a Bacillus subtilis Strain without Extracytoplasmic Function Factors. *Journal of Bacteriology* **192**, 5736–5745 (2010).
22. Park, J. *et al.* Molecular Time Sharing through Dynamic Pulsing in Single Cells. *Cell Syst* **6**, 216-229.e15 (2018).
23. Zhu, B. & Stülke, J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res* **46**, D743–D748 (2018).
24. Ludwig, H. *et al.* Transcription of glycolytic genes and operons in Bacillus subtilis: evidence for the presence of multiple levels of control of the gapA operon: Regulation of glycolysis in Bacillus subtilis. *Molecular Microbiology* **41**, 409–422 (2001).
25. Larsson, J. T. Coordinated patterns of cytochrome bd and lactate dehydrogenase expression in Bacillus subtilis. *Microbiology* **151**, 3323–3335 (2005).
26. Wei, Y., Guffanti, A. A., Ito, M. & Krulwich, T. A. *Bacillus subtilis* YqkI Is a Novel Malic/Na $^{+}$ -Lactate Antiporter That Enhances Growth on Malate at Low Protonmotive Force. *Journal of Biological Chemistry* **275**, 30287–30292 (2000).
27. Lowe, P. N., Hodgson, J. A. & Perham, R. N. Dual role of a single multienzyme complex in the oxidative decarboxylation of pyruvate and branched-chain 2-oxo acids in Bacillus subtilis. *Biochemical Journal* **215**, 133–140 (1983).
28. Grundy, F. J., Waters, D. A., Allen, S. H. & Henkin, T. M. Regulation of the Bacillus subtilis acetate kinase gene by CcpA. *Journal of Bacteriology* **175**, 7348–7355 (1993).
29. Rosenthal, A. Z. *et al.* Metabolic interactions between dynamic bacterial subpopulations. *Elife* **7**, (2018).
30. Fillinger, S. *et al.* Two Glyceraldehyde-3-phosphate Dehydrogenases with Opposite Physiological Roles in a Nonphotosynthetic Bacterium. *Journal of Biological Chemistry* **275**, 14031–14037 (2000).
31. Chai, Y., Kolter, R. & Losick, R. A Widely Conserved Gene Cluster Required for actate Utilization in Bacillus subtilis and Its Involvement in Biofilm Formation. *Journal of Bacteriology* **191**, 2423–2430 (2009).
32. 32. Tojo, S., Kumamoto, K., Hirooka, K. & Fujita, Y. Heavy Involvement of Stringent Transcription Control Depending on the Adenine or Guanine Species of the Transcription Initiation Site in Glucose and Pyruvate Metabolism in Bacillus subtilis. *Journal of Bacteriology* **192**, 1573–1585 (2010).
33. Nicholson, W. L. The Bacillus subtilis ydjL (bdhA) Gene Encodes Acetoin Reductase/2,3-Butanediol Dehydrogenase. *Applied and Environmental Microbiology* **74**, 6832–6838 (2008).
34. Baik, S.-H., Ogasawara, N., Asai, K., Kasahara, Y. & Moriya, S. Regulation of the transport system for C4-dicarboxylic acids in Bacillus subtilis. *Microbiology* **146**, 263–271 (2000).
35. Blencke, H.-M. *et al.* Transcriptional profiling of gene expression in response to glucose in Bacillus subtilis: regulation of the central metabolic pathways. *Metabolic Engineering* **5**, 133–149 (2003).

36. Pechter, K. B., Meyer, F. M., Serio, A. W., Stulke, J. & Sonenshein, A. L. Two Roles for Aconitase in the Regulation of Tricarboxylic Acid Branch Gene Expression in Bacillus subtilis. *Journal of Bacteriology* **195**, 1525–1537 (2013).
37. Fujita, Y. Carbon Catabolite Control of the Metabolic Network in *Bacillus subtilis*. *Bioscience, Biotechnology, and Biochemistry* **73**, 245–259 (2009).
38. Kleijn, R. J. *et al.* Metabolic Fluxes during Strong Carbon Catabolite Repression by Malate in *Bacillus subtilis*. *Journal of Biological Chemistry* **285**, 1587–1596 (2010).
39. Singh, K. D., Schmalisch, M. H., Stulke, J. & Gorke, B. Carbon Catabolite Repression in Bacillus subtilis: Quantitative Analysis of Repression Exerted by Different Carbon Sources. *Journal of Bacteriology* **190**, 7275–7284 (2008).
40. Yoshida, K. *et al.* myo-Inositol Catabolism in Bacillus subtilis. *J. Biol. Chem.* **283**, 10415–10424 (2008).
41. Yoshida, K. I., Aoyama, D., Ishio, I., Shibayama, T. & Fujita, Y. Organization and transcription of the myo-inositol operon, iol, of Bacillus subtilis. *Journal of Bacteriology* **179**, 4591–4598 (1997).
42. Yoshida, K.-I., Shibayama, T., Aoyama, D. & Fujita, Y. Interaction of a Repressor and its Binding Sites for Regulation of the Bacillussubtilis iol Divergon. *Journal of Molecular Biology* **285**, 917–929 (1999).
43. Marciniak, B. C. *et al.* High- and low-affinity cre boxes for CcpA binding in Bacillus subtilis revealed by genome-wide analysis. *BMC Genomics* **13**, 401 (2012).
44. Lopez, D., Vlamakis, H. & Kolter, R. Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiology Reviews* **33**, 152–163 (2009).
45. Ben-Jacob, E. & Schultz, D. Bacteria determine fate by playing dice with controlled odds. *Proceedings of the National Academy of Sciences* **107**, 13197–13198 (2010).
46. Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The Evolution and Ecology of Bacterial Warfare. *Current Biology* **29**, R521–R537 (2019).
47. Shelburne, C. E. *et al.* The spectrum of antimicrobial activity of the bacteriocin subtilosin A. *J Antimicrob Chemother* **59**, 297–300 (2007).
48. Patel, P. S. *et al.* Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by Bacillus subtilis: production, taxonomy, isolation, physico-chemical characterization and biological activity. *J. Antibiot.* **48**, 997–1003 (1995).
49. Gao, L. *et al.* Plipastatin and surfactin coproduction by Bacillus subtilis pB2-L and their effects on microorganisms. *Antonie van Leeuwenhoek* **110**, 1007–1018 (2017).
50. van Gestel, J., Ackermann, M. & Wagner, A. Microbial life cycles link global modularity in regulation to mosaic evolution. *Nature Ecology & Evolution* **3**, 1184–1196 (2019).
51. Kearns, D. B. & Losick, R. Cell population heterogeneity during growth of Bacillus subtilis. *Genes Dev.* **19**, 3083–3094 (2005).
52. López, D., Vlamakis, H., Losick, R. & Kolter, R. Paracrine signaling in a bacterium. *Genes Dev.* **23**, 1631–1638 (2009).
53. Zeigler, D. R. *et al.* The Origins of 168, W23, and Other Bacillus subtilis Legacy Strains. *J Bacteriol* **190**, 6983–6995 (2008).
54. Krüger, E., Witt, E., Ohlmeier, S., Hanschke, R. & Hecker, M. The clp proteases of Bacillus subtilis are directly involved in degradation of misfolded proteins. *J. Bacteriol.* **182**, 3259–3265 (2000).

55. Blom, E.-J., Ridder, A. N. J. A., Lulko, A. T., Roerdink, J. B. T. M. & Kuipers, O. P. Time-Resolved Transcriptomics and Bioinformatic Analyses Reveal Intrinsic Stress Responses during Batch Culture of Bacillus subtilis. *PLOS ONE* **6**, e27160 (2011).
56. Schultz, D., Schlüter, R., Gerth, U. & Lalk, M. Metabolic Perturbations in a Bacillus subtilis clpP Mutant during Glucose Starvation. *Metabolites* **7**, 63 (2017).
57. Helmann, J. D. Specificity of Metal Sensing: Iron and Manganese Homeostasis in Bacillus subtilis. *J. Biol. Chem.* **289**, 28112–28120 (2014).
58. Guedon, E. *et al.* The global transcriptional response of Bacillus subtilis to manganese involves the MntR, Fur, TnrA and sigmaB regulons. *Mol. Microbiol.* **49**, 1477–1491 (2003).
59. Baichoo, N., Wang, T., Ye, R. & Helmann, J. D. Global analysis of the Bacillus subtilis Fur regulon and the iron starvation stimulon: Bacillus subtilis Fur regulon. *Molecular Microbiology* **45**, 1613–1629 (2002).
60. Pi, H. & Helmann, J. D. Sequential induction of Fur-regulated genes in response to iron limitation in Bacillus subtilis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12785–12790 (2017).
61. Mhatre, E. *et al.* The impact of manganese on biofilm development of Bacillus subtilis. *Microbiology (Reading, Engl.)* **162**, 1468–1478 (2016).
62. Wood, H. E., Dawson, M. T., Devine, K. M. & McConnell, D. J. Characterization of PBSX, a defective prophage of Bacillus subtilis. *J. Bacteriol.* **172**, 2667–2674 (1990).
63. Goranov, A. I., Kuester-Schoeck, E., Wang, J. D. & Grossman, A. D. Characterization of the Global Transcriptional Responses to Different Types of DNA Damage and Disruption of Replication in Bacillus subtilis. *Journal of Bacteriology* **188**, 5595–5605 (2006).
64. Anderson, L. M. & Bott, K. F. DNA packaging by the Bacillus subtilis defective bacteriophage PBSX. *J. Virol.* **54**, 773–780 (1985).
65. Shingaki, R., Kasahara, Y., Inoue, T., Kokeguchi, S. & Fukui, K. Chromosome DNA fragmentation and excretion caused by defective prophage gene expression in the early-exponential-phase culture of *Bacillus subtilis*. *Canadian Journal of Microbiology* **49**, 313–325 (2003).
66. Kormelink, T. *et al.* Comparative genome analysis of central nitrogen metabolism and its control by GlnR in the class Bacilli. *BMC Genomics* **13**, 191 (2012).
67. Randazzo, P., Aucouturier, A., Delumeau, O. & Auger, S. Revisiting the in vivo GlnR-binding sites at the genome scale in Bacillus subtilis. *BMC Research Notes* **10**, (2017).
68. Morabbi Heravi, K. & Altenbuchner, J. Cross Talk among Transporters of the Phosphoenolpyruvate-Dependent Phosphotransferase System in *Bacillus subtilis*. *Journal of Bacteriology* **200**, (2018).
69. Garrity, L. F. *et al.* Unique regulation of carbohydrate chemotaxis in Bacillus subtilis by the phosphoenolpyruvate-dependent phosphotransferase system and the methyl-accepting chemotaxis protein McpC. *J. Bacteriol.* **180**, 4475–4480 (1998).
70. Jordan, S., Junker, A., Helmann, J. D. & Mascher, T. Regulation of LiaRS-Dependent Gene Expression in Bacillus subtilis: Identification of Inhibitor Proteins, Regulator Binding Sites, and Target Genes of a Conserved Cell Envelope Stress-Sensing Two-Component System. *Journal of Bacteriology* **188**, 5153–5166 (2006).

71. Ye, B.-C. *et al.* Time-Resolved Transcriptome Analysis of Bacillus subtilis Responding to Valine, Glutamate, and Glutamine. *PLoS ONE* **4**, e7073 (2009).
72. Cárdenas, P. P. *et al.* RecX Facilitates Homologous Recombination by Modulating RecA Activities. *PLoS Genetics* **8**, e1003126 (2012).
73. Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M. & Elowitz, M. B. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440**, 545–550 (2006).
74. Berka, R. M. *et al.* Microarray analysis of the Bacillus subtilis K-state: genome-wide expression changes dependent on ComK. *Molecular Microbiology* **43**, 1331–1345 (2002).
75. Ogura, M. *et al.* Whole-genome analysis of genes regulated by the Bacillus subtilis competence transcription factor ComK. *J. Bacteriol.* **184**, 2344–2351 (2002).
76. 76. Smits, W. K. *et al.* Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. *Molecular Microbiology* **65**, 103–120 (2007).
77. Yadav, T. *et al.* Genetic recombination in Bacillus subtilis : a division of labor between two single-strand DNA-binding proteins. *Nucleic Acids Research* **40**, 5546–5559 (2012).
78. Hanlon, D. W., Rosario, M. M. L., Ordal, G. W., Venema, G. & Van Sinderen, D. Identification of TlpC, a novel 62 kDa MCP-like protein from Bacillus subtilis. *Microbiology* **140**, 1847–1854 (1994).
79. Allard-Massicotte, R. *et al.* Bacillus subtilis Early Colonization of Arabidopsis thaliana Roots Involves Multiple Chemotaxis Receptors. *mBio* **7**, (2016).

APPENDIX 1

Chapter 2 – Supplementary materials

A1.1 MATERIALS AND METHODS

Media

Rich media was prepared as LB Miller broth (Cat #244620, Difco), supplemented with 15 g L^{-1} bacto agar (Cat# 214030, Difco) when making solid media and with bacto agar and 0.1% carboxymethylcellulose (C5678, Sigma) to quantify cellulase clearings. M9 minimal media was prepared with M9 salts (Cat# 248510, Difco), 1 mM $MgSO_4$, 100 µM $CaCl_2$, supplemented with 1 mg mL^{-1} biotin and a carbon source (glucose, cellobiose or PASC) at concentrations specified in the text. M9 minimal 0.4% cellobiose plates for isolating cellobiose utilizer strains were prepared as above supplemented with 15 g L^{-1} bacto agar. Ampicillin and kanamycin antibiotics were supplemented when required at 20 µg mL^{-1} and 20 µg mL^{-1}, respectively, unless indicated.

PASC preparation

Phosphoric acid swollen cellulose (PASC) was prepared following a reported

protocol[1]. Briefly, a cellulose slurry was created from sterile water and cellulose powder (Sigmacell 20, Sigma), combined with ice-cold 85% phosphoric acid (Sigma) and incubated at 4°C for 2 hours. Cellulose was precipitated in sterile water and washed repeatedly to remove the acid and bring the slurry to neutral pH.

Strains and plasmid construction

All strains used for assays in this study were derived from a variant of EcNR1[2] with genomic deletions of the *fim* operon, *ampR* and *lacIZ* (CT009). Cellobiose utilizer DL069 was used as parent strain for all SDAc variants. Control strains for cellobiose utilization were derived from progenitor strain CT009. All strains are listed in Supplementary Table 2.2.

Plasmid construction was carried out using standard Gibson assembly protocols (Gibson et al. 2009). PCRs were performed using Phusion PCR Master Mix (NEB, M0531L) with a T-100 thermal cycler (Bio-Rad). Synthetic oligonucleotides for cloning were synthesized by Integrated DNA Technologies. Gibson assemblies were transformed into cloning strain DH5alpha. Transformations were performed by electroporation at 1250 V using an Eppendorf 2510 electroporator. Transformants were cultured on LB Miller agar plates with appropriate antibiotic or LB liquid media (Difco) at 37°C. Plasmids were extracted using the QIAprep Spin Miniprep Kit. Plasmids were sequenced by Genewiz (Seattle, WA). All plasmid assemblies were performed with vector pSC101 vector pGA4A5 or p15A vector pGA3K3 [3].

Cellulase genes were synthesized following standard polymerase cycling assembly (PCA) protocols[4] using synthetic oligonucleotides provided by OligoCo, Inc.

PCA was performed with Phusion polymerase and cloned into plasmids for sequence verification using Gibson Assembly.

Genome engineering

Cellobiose utilizer strain RGE531 was engineered using multiple cycles of multiplex genome engineering following Wang *et al.*[2]. *E. coli* strain CT009 was grown in LB Miller broth to OD 0.4-0.6 at 30°C, then heat shocked in a water bath for 15 minutes at 42°C. The culture was chilled via ice slurry and centrifuged for 3 minutes at 5000 rpm in a centrifuge (Legend XR1, Thermo Scientific) pre-chilled to 4°C. The supernatant was decanted and cells were resuspended in 3 mL of ice-cold sterile water. The resuspension was centrifuged again for 3 minutes at 5000 rpm and the supernatant decanted. The cell pellet was resuspended in 800 µL of ice-cold sterile water and transferred to a 1.5 mL microcentrifuge tube. Cells were centrifuged at 4000 x g for 1 minute (Accuspin Micro 17R, Fisher Scientific) pre-chilled to 4°C. Supernatants were aspirated by pipette and cell pellets were resuspended in 40 mL ice-cold sterile water.

Competent cells were transformed with an oligonucleotide cocktail targeting ribosome binding site variation of *chbB/A/C/F/G* or deletion of *chbR* (synthesized by OligoCo, Inc., see Supplementary Table 2.10) by electroporation at 1800V using an Eppendorf electroporator 2510. Transformants were recovered in 10 mL of M9 0.4% cellobiose media in a shaker flask and grown for 48-72 hours at 30°C with shaking. This process was repeated for two additional cycles, using the saturated recovery culture from the previous as the inoculum for the next cycle of mutagenesis. Following each outgrowth in cellobiose, individual colonies were screened for rapid growth on M9 0.4%

cellobiose agar plates, and several variants were isolated for *chb* operon sequencing and further use.

Microscopy

Fluorescence microscopy of SDAc microcolonies was carried out using equipment and procedures described previously[3]. *E. coli* strain 2.320 was transformed with a differentiation plasmid (*tetR*/*lacI* switch with $(T)_{10}/(A)_{12}$ repeats, p15A origin) and a payload delivery plasmid (colicin E3 lysis gene with $(AC)_{11}$ repeat, colE1 origin). Individual colonies were grown to saturation in minimal M9CA media (M8010, Teknova), 50 μg mL^{-1} kanamycin and 20 μg mL^{-1} ampicillin. Cultures were diluted 100:1, grown for two hours with shaking at 32°C and vortexed to break up cell aggregates. 1 μL of the culture was added to a glass-bottomed microscopy dish (GWSB-3512, WillCo Wells BV) and cells were immobilized on the dish surface using a nutrient agar slab prepared with M9CA media, as described above, and 1% bacto agar (Difco). Microscopy dishes were transferred to a Nikon Ti-S inverted microscope (Nikon Instruments) equipped with a CoolsnapHQ camera (Roper Scientific) and an environmental chamber (In Vivo Scientific). Microcolony growth and fluorescent states were recorded every 15 min by capturing phase contrast, GFP and RFP images while the dishes were maintained at 32°C for 10 to 12 hours.

Measurements of growth, differentiation and lysis rates

Individual strains were initialized to the consumer state or altruist state (lysis-deficient control only) by inoculating individual colonies from a restreaked LB agar

plate. The colonies were forced to the consumer state via IPTG induction (1 mM) or the altruist state via aTc induction (100 ng/mL) and grown at 37°C for 16 hours on LB agar media supplemented with ampicillin and kanamycin. Cultures were washed to remove inducer, inoculated 2000:1 into a well of a deep-well plate with 500 µL M9 minimal 0.4% cellobiose media supplemented with biotin and antibiotics as above, and grown at 37°C. To maintain exponential growth and low cell density, each culture was sampled every 12 hours and transferred to the cytometer (C6 with CSampler, Accuri) to measure culture density and the fluorescence distribution to determine the fraction of altruists. Growth rates between measurements were measured as a simple exponential fit using the initial cell density, the final cell density, and the number of hours between measurements. Each culture was diluted into fresh broth with a dilution factor defined by the culture density and the estimated growth rate. The dilution factors ranged from 5:1 to 50:1 every 12 hours. Growth rates and population fractions of consumers and altruists were calculated as described in Supplementary Notes 3 and 4, respectively.

Cell growth in PASC media

Cultures grown in PASC were initially inoculated into LB liquid culture with ampicillin and kanamycin from individual colonies restreaked on LB agar plates with the same antibiotics and IPTG to force the consumer state. Saturated cultures were transferred at 100:1 dilutions into M9 minimal PASC media and viable cell counts were periodically quantified by serial dilution and plating.

Measurements of carbon to biomass conversion efficiency

Carbon to biomass yield for cellobiose and PASC were measured using strains grown in M9 minimal media supplemented with increasing levels of each carbon source, up to 4 g L^{-1}. Cellobiose growth was quantified after 36 hours with cellulase and lysis deficient DL146 by serial dilution, plating and counting colonies. PASC growth was quantified after 72 hours with DL110 by the same method. All cultures were inoculated according to the PASC growth assay protocol and colony forming units were quantified for at least three replicates on LB plates supplemented with ampicillin and kanamycin. The estimate for cellobiose was fit only to the linear, carbon-limited regime observed at concentrations less than 1g L^{-1} (see Supplementary Figure 8 and Supplementary Note 6)

Measurements of cellulose degradation

Cellulase production strains for individual cellulases and cellulase cocktails were initialized by restreaking on LB agar plates supplemented with kanamycin (50 µg/mL), ampicillin (20 µg/mL) and aTc (100 ng/mL). Single colonies were transferred to 3 mL LB supplemented with kanamycin and ampicillin and grown to saturation overnight. Turbid 1 mL cultures were sonicated on ice (Ultrasonic Cell Disruptor, SharperTek) to release intracellular cellulase. The sonicator settings were 15 minutes with 95% power, and 75% duty cycle of 20 seconds on and 10 seconds off. Viable cell counts were measured before and after sonication using serial dilution and CFU plating to measure cell lysis efficiency (ρ). Cell lysates were filtered (0.4 micron filter, Company) to generate sterile supernatant. 10 µL cellulase lysate was transferred to 1 mL of 0.4% M9 minimal PASC + biotin media and incubated at 37°C for 72 hours, periodically measuring cellulose degradation by Congo Red absorbance or nutrient release. Uniprot

identifiers for cellulases are Q45430 (CelD04), P10475 (BsCel5) and A9KT90 (CpCel9).

Congo Red absorbance measurements were taken by sampling 100 uL of the PASC growth media with cellulase at 0, 6, 12, 24, 48 and 72 hours post-inoculation, adding 100uL 0.15% (w/v) aqueous Congo red. After staining for 30 minutes, samples were centrifuged at 5000 rpm (Legend XR1) and 100 µL of PASC supernatant was transferred to a clear, flat-bottom microwell plate (Corning 2595). Absorbance readings were measured by plate reader (Biotek) at a wavelength of 480 nm.

Cellulolytic release of digestible nutrients was estimated by measuring DL146 growth on PASC supernatant after lysate incubation. 10 µL aliquots of sterile cell lysates, prepared as above, were inoculated in 1 mL M9 minimal 0.4% PASC media, prepared in parallel for each time measurement. At each time point, 1 mL PASC samples were centrifuged in triplicate to pellet the PASC and collect the supernatant. DL146 cultures were grown for 36 hours in the supernatant media and culture densities were measured by serial dilution and plating.

Congo Red clearing assay

Individual colonies were generated on LB agar containing cellulose and the plates were stained with Congo Red dye to observe clearings. Strains were incubated for 14 hours on LB Miller agar supplemented with kanamycin, ampicillin and 0.1% carboxymethylcellulose. Colonies were removed from plates by washing with 70% ethanol. Plates were stained with 0.1% Congo Red for 20 minutes and destained with 1 M NaCl in two rinse steps, destaining for 5 minutes and 30 minutes, respectively. Clearing diameter for each colony was quantified from bright-field images of the stained

plates by measuring clearing diameter using ImageJ software. Clearing diameter was converted from pixels to millimeters by normalizing to known plate diameter.

Sequencing SDAc switching or lysis cheaters

Bright, monochromatic red or green fluorescent colonies isolated from six replicate cultures of $(T)_{18}/(A)_{10}$ grown in M9 minimal 0.4% PASC media for 72 hours were identified via blue light transillumination. Each colony was replated separately on an LB agar plate and an LB agar plate supplemented with kanamycin and ampicillin to test for plasmid loss. Sequencing was performed on individual colonies by Genewiz, Inc. Differentiation plasmids were sequenced at the dual promoter region (BSSregSeq_2809) and the *lacI* Cterminal sequence (lacIrptSeq_2810). Payload plasmids were sequenced across the payload transcription unit (oligos Vf2 & Vr). Sequencing oligos are listed in Supplementary Table 2.12.

Materials and Methods References

1. Zhang, Y. H. P., Cui, J., Lynd, L. R. & Kuang, L. R. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: evidence from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* **7,** 644–648 (2006).
2. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
3. Egbert, R. G. & Klavins, E. Fine-tuning gene networks using simple sequence repeats. *Proc Natl Acad Sci USA* **109,** 16817–16822 (2012).
4. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* **30,** e43 (2002).

A1.2 SUPPLEMENTARY TABLES

Parameter	Description	Units
v_C, v_A	Consumer, altruist growth rates	h^{-1}
k_C, k_A	Half maximal growth constants	g
σ	Differentiation rate	h^{-1}
k_σ	Half maximal differentiation constant	g
ρ	Lysis rate	h^{-1}
ω	Cellulase activity	CFU^{-1} mL
γ	Nutrient to biomass conversion efficiency	CFU g^{-1}

Table 2.1. SDAc model parameters

Strain ID	Switch variant*	Payload	colE3 lysis repeat	Parent strain	Modifications
CT009	N/A	N/A	N/A	EcNR1	ΔfimAICDFGH ΔampR ΔlacIZ
RGE531	N/A	N/A	N/A	CT009	pChb<>pConst,
DL046	$(T)_{12}/(A)_{10}$	CelD04	$(AT)_8$	DL069	
DL069	N/A	N/A	N/A	CT009	pChb<>pConst, ΔchbR, Δ(λ-Red)
DL108	$(T)_{18}/(A)_{10}$	CelD04	$(AT)_8$	DL069	
DL110	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_8$	DL069	
DL112	$(T)_{14}/(A)_{10}$	CelD04	$(AT)_8$	DL069	
DL146	$(T)_{12}/(A)_{10}$	None	No lysis gene	DL069	
DL147	$(T)_{12}/(A)_{10}$	CelD04	No lysis gene	DL069	
DL180	N/A	CelD04	No lysis gene	CT009	Δ(λ-Red), pCellulose[1] (ampR), mRFP1 (kanR)
DL182	$(T)_{16}/(A)_{10}$	None	No lysis gene	DL069	
DL183	$(T)_{14}/(A)_{10}$	None	No lysis gene	DL069	
DL188	$(T)_{18}/(A)_{10}$	None	No lysis gene	DL069	
DL202	$(T)_{16}/(A)_{10}$	BsCel5	$(AT)_7$	DL069	
DL261	$(T)_{16}/(A)_{10}$	CpCel9	$(AT)_7$	DL069	
DL268	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_5$	DL069	
DL269	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_6$	DL069	
DL270	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_7$	DL069	
DL271	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_9$	DL069	
DL272	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_{10}$	DL069	
DL273	$(T)_{16}/(A)_{10}$	CelD04	$(AT)_{11}$	DL069	
DL292	$(T)_{12}/(A)_{10}$	CpCel9	No lysis gene	DL069	
DL294	$(T)_{12}/(A)_{10}$	BsCel5, CpCel9	No lysis gene	DL069	
DL296	$(T)_{12}/(A)_{10}$	CelD04, CpCel9	No lysis gene	DL069	
DL307	$(T)_{12}/(A)_{10}$	BsCel5	No lysis gene	DL069	
DL322	$(T)_{16}/(A)_{10}$	BsCel5, CpCel9	$(AT)_9$	DL069	
DL329	$(T)_{12}/(A)_{12}$	CelD04	colE3	DL069	
DL327	$(T)_{16}/(A)_{10}$	CelD04, CpCel9	$(AT)_9$	DL069	

* $(T)_x/(A)_y$ indicates *x* poly-T rbSSR repeats driving TetR expression and *y* poly-A repeats driving LacI expression for the differentiation plasmid.

Table 2.2. Strains used

Strain ID	Switch variant	Cellulase	Lysis	Induction state*	Growth rate (hr^{-1})	StDev
DL146	$(T)_{12}/(A)_{10}$	None	None	Consumer	0.463	0.019
DL146	$(T)_{12}/(A)_{10}$	None	None	Altruist	0.391	0.019
DL147	$(T)_{12}/(A)_{10}$	CelD04	None	Consumer	0.459	0.007
DL147	$(T)_{12}/(A)_{10}$	CelD04	None	Altruist	0.413	0.023
DL046	$(T)_{12}/(A)_{10}$	CelD04	colE3	Consumer	0.438	0.018
DL112	$(T)_{14}/(A)_{10}$	CelD04	colE3	Consumer	0.420	0.017
DL183	$(T)_{14}/(A)_{10}$	None	None	Consumer	0.430	0.012
DL110	$(T)_{16}/(A)_{10}$	CelD04	colE3	Consumer	0.389	0.018
DL182	$(T)_{16}/(A)_{10}$	None	None	Consumer	0.423	0.019
DL108	$(T)_{18}/(A)_{10}$	CelD04	colE3	Consumer	0.266	0.041
DL188	$(T)_{18}/(A)_{10}$	None	None	Consumer	0.395	0.003

* Induction to consumer state implies IPTG induction (1 mM). Induction to altruist state implies aTc induction (100 ng/mL).

Table 2.3. Cellobiose growth rates

Strain ID	Switch variant	σ (h^{-1})	StDev	τ (h)	StDev
DL146	$(T)_{12}/(A)_{10}$	0.027	0.009	26.8	4.56
DL183	$(T)_{14}/(A)_{10}$	0.081	0.018	21.1	0.97
DL182	$(T)_{16}/(A)_{10}$	0.116	0.019	18.6	0.78
DL188	$(T)_{18}/(A)_{10}$	0.211	0.015	13.9	0.03

Table 2.4. Differentiation rates for lysis-deficient switch variants

Strain ID	Switch variant	Lysis variant	σ (h^{-1})	ρ (h^{-1})	StDev
DL046	$(T)_{12}/(A)_{10}$	$(AT)_8$	0.027	0.180	0.080
DL112	$(T)_{14}/(A)_{10}$	$(AT)_8$	0.081	0.159	0.052
DL108	$(T)_{18}/(A)_{10}$	$(AT)_8$	0.211	0.256	0.036
DL268	$(T)_{16}/(A)_{10}$	$(AT)_5$	0.116	0.130	0.035
DL269	$(T)_{16}/(A)_{10}$	$(AT)_6$	0.116	0.165	0.040
DL270	$(T)_{16}/(A)_{10}$	$(AT)_7$	0.116	0.134	0.042
DL110	$(T)_{16}/(A)_{10}$	$(AT)_8$	0.116	0.192	0.048
DL271	$(T)_{16}/(A)_{10}$	$(AT)_9$	0.116	0.173	0.080
DL272	$(T)_{16}/(A)_{10}$	$(AT)_{10}$	0.116	0.072	0.032
DL273	$(T)_{16}/(A)_{10}$	$(AT)_{11}$	0.116	0.073	0.032

Table 2.5. Lysis rates for switch and lysis variants

Media	Linear range	γ (CFU g^{-1})	StDev	R2
Cellobiose	0 – 1 g L^{-1}	1.279e+12	3.344e+10	0.9919
PASC	0 – 4 g L^{-1}	1.795e+11	8.475e+09	0.9825

Table 2.6. Carbon to biomass yields for cellobiose and PASC

Strain ID	Cellulase	ω (hydrolysis)	StDev	ω (biomass)	StDev
DL292	CpCel9	4.6e-12	2.4e-13	3.6e-13	3.9e-13
DL147	CelD04	5.7e-12	6.8e-13	1.6e-12	5.3e-13
DL294	CelD04 + CpCel9	6.8e-12	9.5e-13	1.8e-12	5.6e-13
DL307	BsCel5	6.1e-12	9.8e-13	1.2e-12	2.9e-13
DL296	BsCel5 + CpCel9	7.4e-12	4.3e-13	1.9e-12	3.0e-13

Note: units for ω are CFU^{-1} mL

Table 2.7. Cellulase activity measurements for cellulase production strains

Parameter	Fit	StDev
k_C, k_A	0.00112	0.00045
k_σ	0.09999	0.00019

Table 2.8. Half-maximal rate constants for growth and differentiation

Cheater type(s)	χ_C (h^{-1})	Std Error	χ_A (h^{-1})	Std Error
Differentiation	2.473e-07	5.857e-08	n/a	n/a
Differentiation and lysis	1.775e-06	6.279e-07	1.908e-05	8.207e-06

Table 2.9. Cheater rates for expanded model

Plate Replicate	Isolation condition	Promoter mutation	lacI mutation
2	LB	None	None
2	LB kan/amp	None	$(CTGG)_2$
2	LB kan/amp	None	None
3	LB kan/amp	None	IS2 insertion at base 15 of lacI
3	LB kan/amp	None	IS2 insertion at base 15 of lacI
3	LB kan/amp	None	IS2 insertion at base 34 of lacI
4	LB	None	$(CTTG)_4$
4	LB	None	148 bp lacI deletion (A351)
4	LB kan/amp	None	None
4	LB kan/amp	None	$(CTTG)_4$
5	LB kan/amp	None	$(CTTG)_4$
5	LB kan/amp	None	$(CTGG)_2$
5	LB kan/amp	None	$(CTTG)_4$
5	LB kan/amp	None	20 bp lacI deletion (A303)
6	LB kan/amp	$\Delta(T)_5$	None
6	LB kan/amp	None	$(CTGG)_2$
6	LB kan/amp	$\Delta(T)_5$	None
6	LB kan/amp	$\Delta(T)_5$	None

Sequencing results of differentiation plasmid mutations from monochromatic RFP-positive $(T)_{18}/(A)_{10}$ escape colonies (six replicate plates). The first replicate plate had no mutations at the sequenced loci for the differentiation plasmid.

Table 2.10. Genotypes for differentiation cheaters

Plate Replicate	Isolation condition	Payload plasmid mutations
1	LB	None
1	LB Kan/Amp	None
2	LB	None
2	LB	None
2	LB	None
2	LB Kan/Amp	None
2	LB Kan/Amp	None
2	LB Kan/Amp	None
2	LB Kan/Amp	None
3	LB	None
4	LB	IS1 insertion at A514 in celD04
4	LB	IS1 insertion at A514 in celD04
4	LB Kan/Amp	IS1 insertion at A514 in celD04
4	LB Kan/Amp	IS1 insertion at A514 in celD04
5	LB	None
6	LB	None
6	LB	None
6	LB Kan/Amp	None

Sequencing results of payload plasmid mutations from monochromatic GFP-positive $(T)_{18}/(A)_{10}$ escape colonies (six replicate plates).

Table 2.11. Genotypes for lysis cheaters

Oligo name	Sequence
pChb<>pConst	CTTCCATGCTCTGGGTAACTTGCGAAACCAACATGATGAATCTATTATAAAAAAAA AAAAAAAAA AAGTCAATACTCTATCGAACTCAGGCCAAAAAAAACCGGCGCAATGGCCGGTTTC
chbR_del	ggacataaataatccagcaacaggacagatatgtgaattgtcaggtataacgacttactgcatcg actccttatgccttcagtttttcatgaagctcaattaattcagtaatcagttcac
chbA_A15	agctcttcagcttccgtttgcgtatcgggaatgttatcgagatccatcatTTTTTTTTTTTTTTT CCTCCTCttttcttaccggcacgattacccgtaccggcatcgattaaaatttcag
chbC_T15	actgcaaaagggaggagtacctttttcaagcgatgcaataacattactcatAAAAAAAAAAAAAAA CCTCCTaaaaaccgcaatttaaatattgcggtattgatttatgaaataactcttt
chbB_A15	gaggtagacatgcccgcagaacaaaacagataaatgtgtttcttttccatTTTTTTTTTTTTTTT ACCTCCTGATatcgacgattatctgtcagccagacactccgcaagccttaacctg
chbF_A5	gtatagctgctcccgccaccaatagtgacgacttttaatttctggctcatTTTTTACCTCCTagt acagaatactgatatctggcatatctgccccccggacataaata
chbF_A15	gtatagctgctcccgccaccaatagtgacgacttttaatttctggctcatTTTTTTTTTTTTTTT ACCTCCTagtacagaatactgatatctggcatatctgccccccggacataaata
chbG_T15	cctttgcttaagccaaaatcatcggcattaacaatcagtaagcgttccatAAAAAAAAAAAAAAA CCTCCttaatgtgctttttaagctctgcgatgcagtcggcaaagtttggcagcc
BSSregSeq_2809	GGCTGCTCTACACCTAGCTTCTGG
lacIrptSeq_2810	GGTTTGTTGAAAACCGGACATGG
Vf2	TGCCACCTGACGTCTAAGAA
Vr	ATTACCGCCTTTGAGTGAGC

Table 2.12. Oligonucleotides used to modify chb operon and for cheater sequencing

A1.3 SUPPLEMENTARY FIGURES

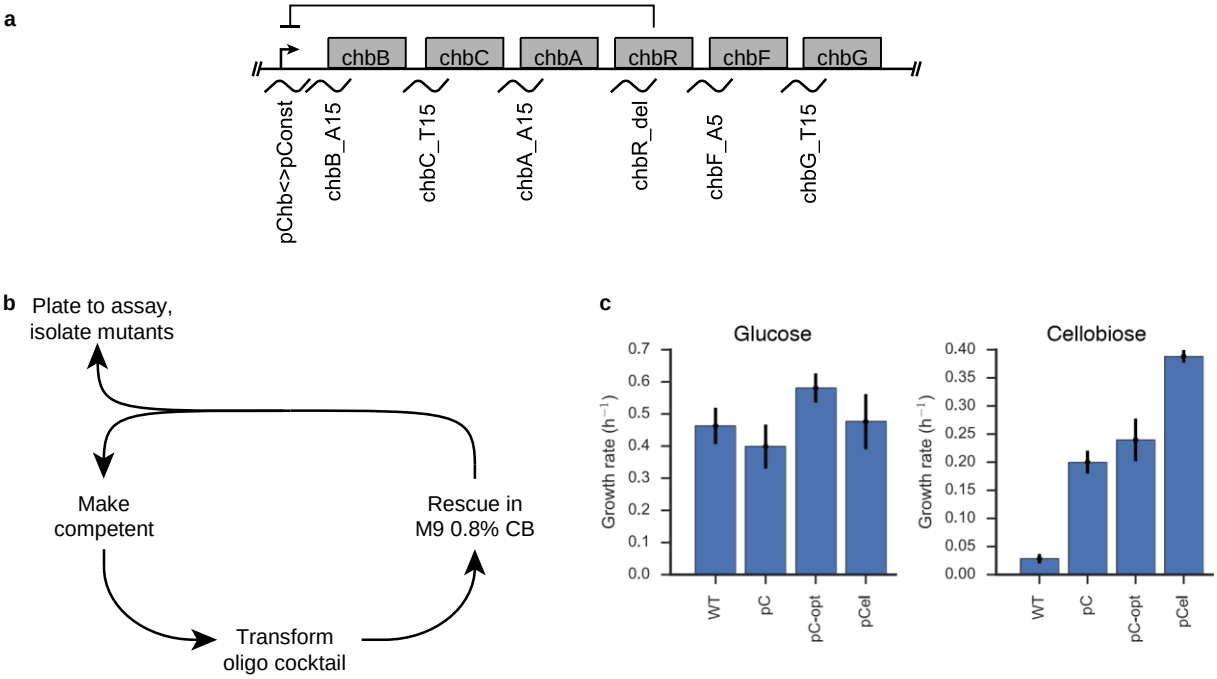


Figure 2.1 figure supplement 1. Engineering cellobiose utilization for SDAc strains. **(a)** Chitobiose (*chb*) operon with targeted sites for promoter replacement (pChb), overexpression (*chbA*, *chbB*, *chbC*, *chbF*, *chbG*) or deletion (*chbR*). See Supplementary Table 2.12 for oligonucleotide sequences. **(b)** Workflow to select for variants that grow faster on cellobiose. **(c)** Growth rates for a selection of variants: wild-type (WT, EcNR1 Δ*fim* Δ*lacIZ*), the native *chb* promoter replaced with a constitutive promoter (pC, RGE531), the fastest growing recombinant mutant (pC-opt, DL069) and a control strain (pCel, DL180) that expresses a membrane-bound beta-glucosidase from *Cellvibrio japonicas*[1]. Growth rates were measured at 32°C in a Biotek Synergy plate reader (N=3) using 200 µL cultures grown in 96well microwell plates.

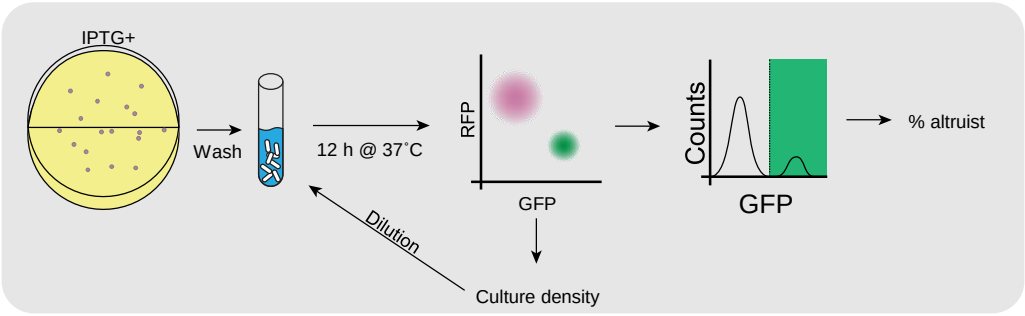


Figure 2.2 – figure supplement 1. Data collection for growth, differentiation and lysis rates. Experimental workflow to estimate growth, differentiation and lysis rates using low-density continuous culture and flow cytometry. Cultures were initialized in the consumer state by growing individual colonies on LB agar supplemented with IPTG. Colonies were washed and diluted into M9 minimal cellobiose media and periodically sampled in the flow cytometer to calculate the population fraction of differentiated cells and the dilution factor for the next growth passage.

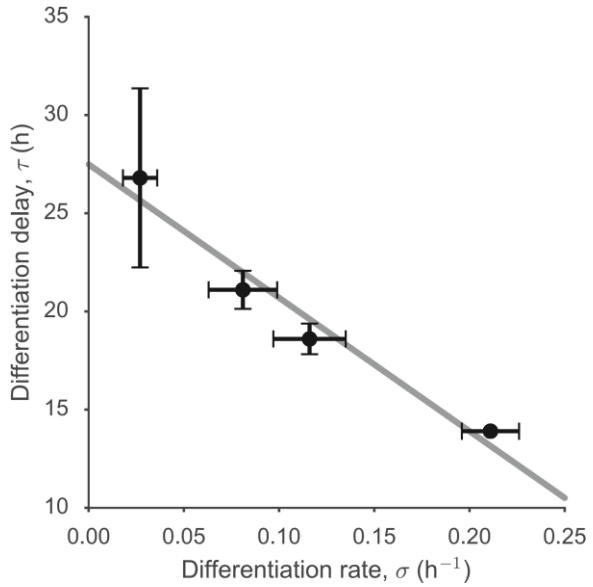


Figure 2.2 – figure supplement 2. Mapping differentiation rate to differentiation delay. Relationship between the differentiation rate and the delay in the initiation of switching following the removal of IPTG inducer for strains DL046, DL112, DL110 and DL108. Linear fit line is y=28.3−72.7x h. The shaded area represents the 95% confidence interval for the fit.

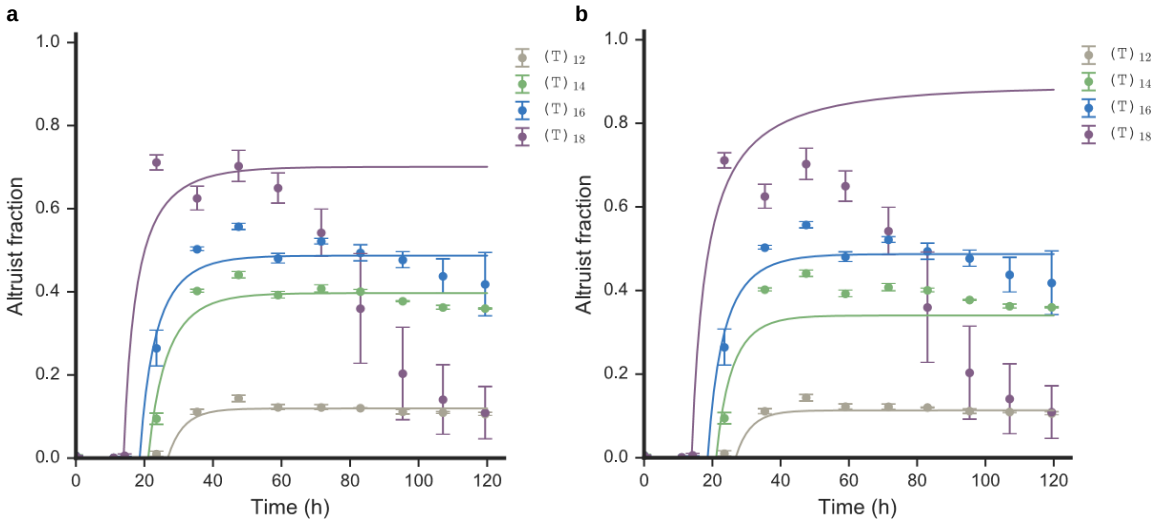


Figure 2.2 – figure supplement 3. Altruist fraction fits combining growth, differentiation and lysis parameters. **(a)** Altruist fractions as a function of time for multiple differentiation variants. Altruist fractions were calculated from flow cytometry measurements of the continuous culture in cellobiose. Curves are derived from model predictions using growth rates of DL147, differentiation rates as estimated for the lysisdeficient switch variants and lysis rate fits for each strain. The lysis rate fit for $(T)_{18}$ was truncated at 65 h. Each strain uses rbSSR $(AT)_8$ for lysis gene expression. Data for labels $(T)_{12}$, $(T)_{14}$, $(T)_{16}$ and $(T)_{18}$ correspond to measurements for strains DL046, DL112, DL110 and DL108, respectively (N=3). **(b)** Fits as in **(a)**, but with a single lysis rate of 0.192 $h^{!!}$ measured from DL110. The error bars represent one standard deviation from the mean.

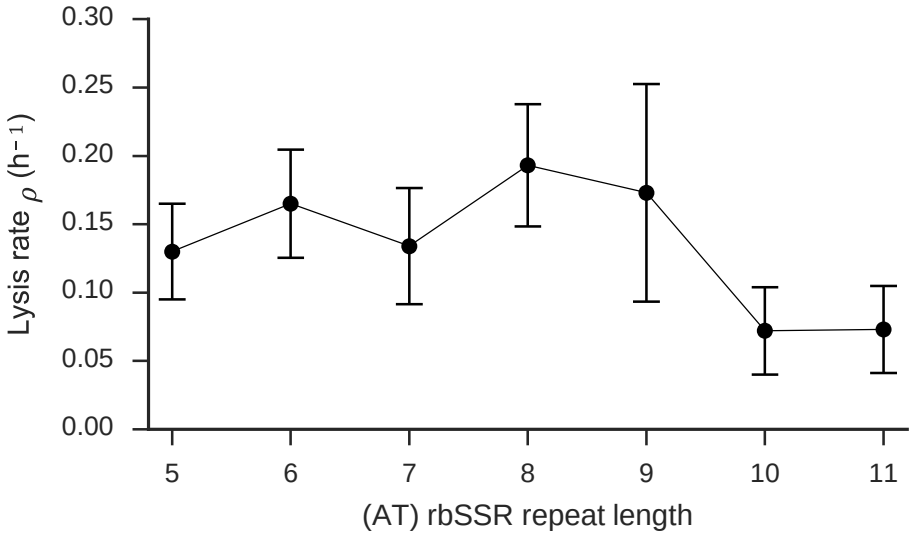


Figure 2.2 figure supplement 4 Mapping colE3L (AT) rbSSR length to lysis rate. Lysis rate fits as a function of (AT) repeat length for multiple strains with switch $(T)_{16}$. Error bars represent parameter estimate uncertainty, inclusive of measurement error and parameter bootstrapping from growth rates and differentiation rates.

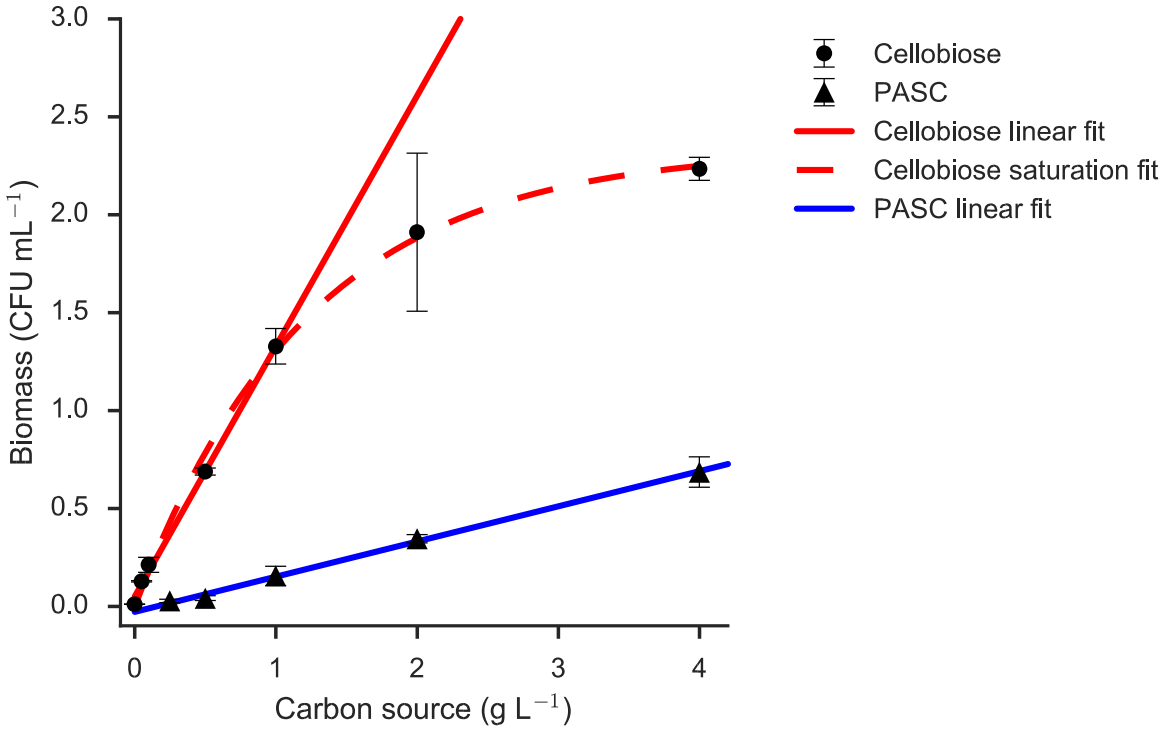


Figure 2.2 – figure supplement 5. Carbon to biomass yield estimates. Carbon utilization measurements and fits for DL146 (Cellobiose, N=3) and DL110 (PASC, N=2). Linear fits are shown for each data set and a saturating fit is included for growth on cellobiose. Error bars represent one standard deviation from the mean.

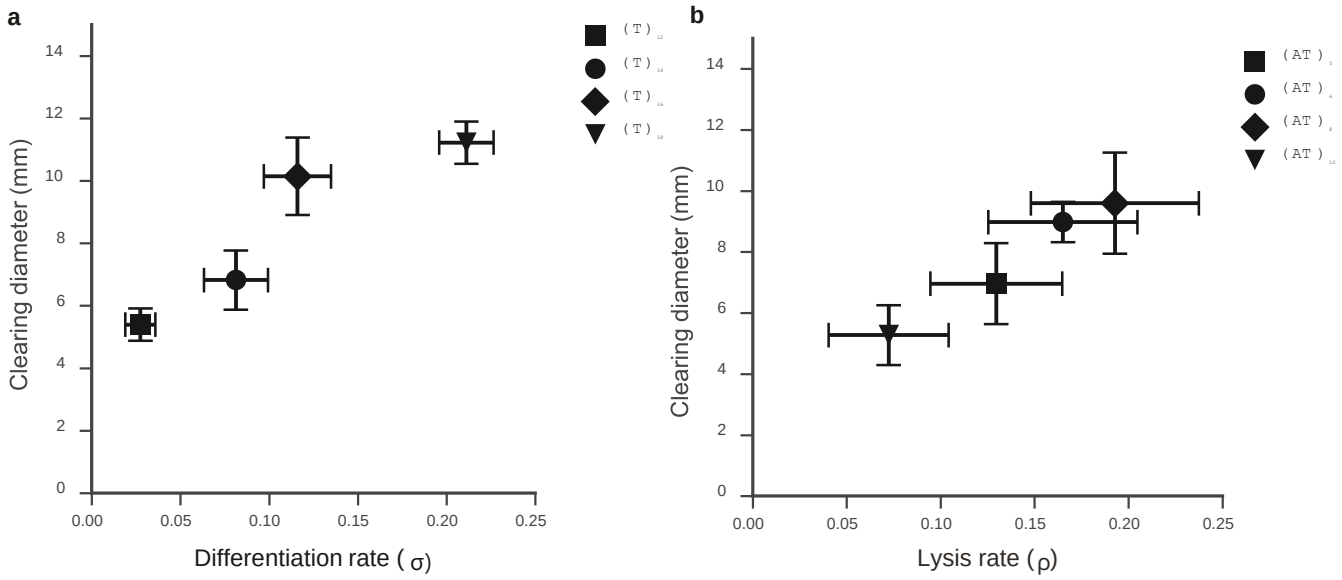


Figure 2.3 – figure supplement 1. Mapping cellulose degradation clearings to differentiation and lysis rates. **(a)** Plot of cellulose degradation clearings as a function of differentiation rate for SDAc variants of the poly-(T) *tetR* rbSSR. **(b)** Plot of cellulose degradation clearings as a function of lysis rate for SDAc variants of the *colE3L* lysis gene rbSSR. For each panel the x-axis error bars represent uncertainty of experimental measurements and parameter bootstrapping and the y-axis error bars represent standard error from at least six replicate clearings of individual colonies (see Figure 2.3).

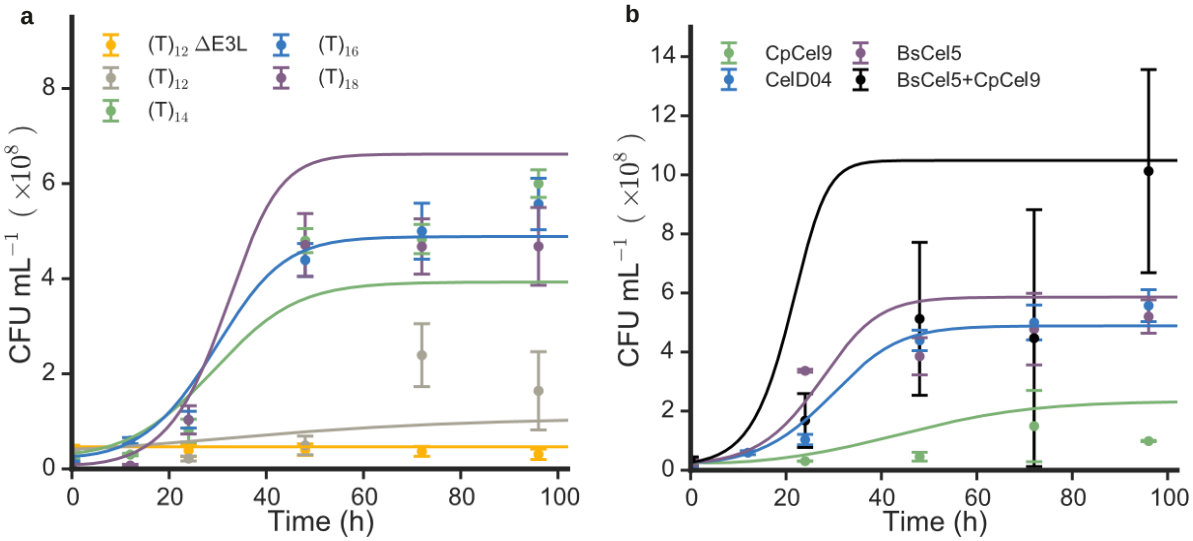


Figure 2.3 – figure supplement 2. Cellulose growth dynamics predictions for SDAc strains. **(a)** Growth dynamics and model predictions for multiple differentiation variants and a lysis control. **(b)** Growth dynamics and model predictions for multiple cellulase variants using intermediate differentiator $(T)_{16}$. Data points represent the means from at least 4 biological replicates ($N \geq 4$), and error bars one standard deviation from the mean. Outliers were excluded from original sample pools of up to 12 replicates using the cov.rob function in the R package, MASS, which removed any data points outside the 95% quantile.

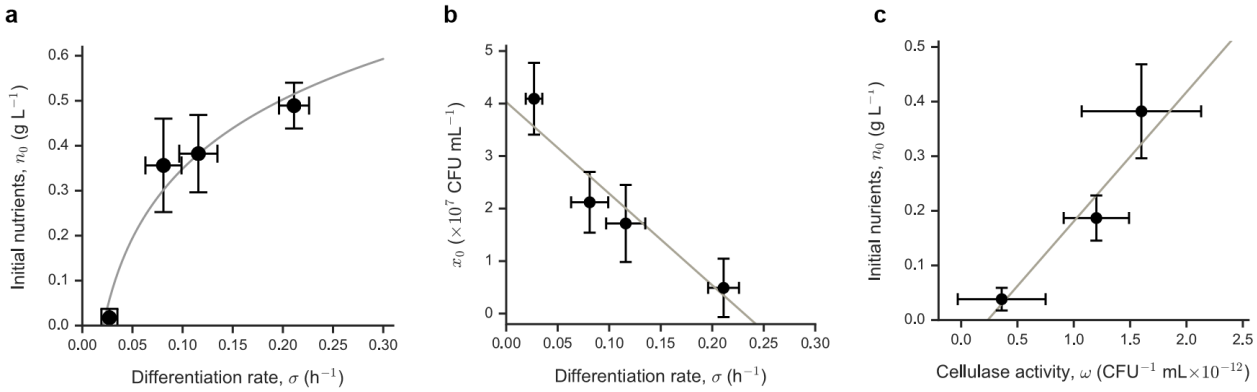


Figure 2.3 – figure supplement 3. Model fit of initial conditions to differentiation rate. **(a)** Logarithmic fit of differentiation rate to initial nutrients generated in PASC media, measured in growth of cellulase and lysis deficient DL146 using supernatant from toggle switch variants. **(b)** Linear fit of differentiation rate to initial viable population size at initiation of PASC growth assay, measured by CFU counts. **(c)** Linear fit of cellulase activity to initial nutrients generated in PASC media, measured as in **(a)**. Differentiation variant $(T)_{16}$ was used for each cellulase. Vertical error bars represent standard deviation of 3 biological replicates from the mean of the value reported on the x-axis (n = 3). Horizontal error bars represent parameter uncertainty as calculated through bootstrapping methods described in Supplementary Note 2.

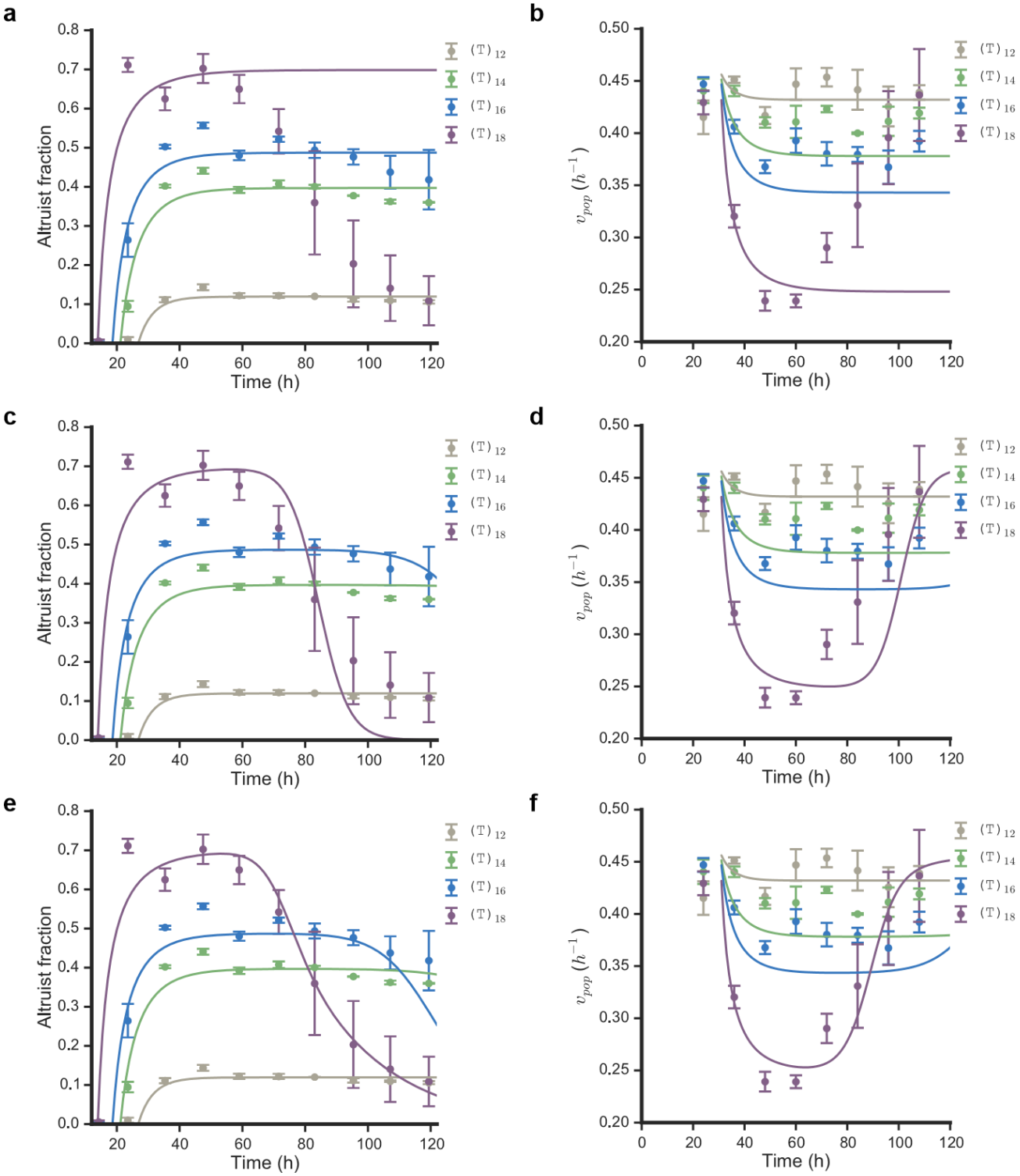


Figure 2.4 figure supplement 1. Model fits to altruist fraction and overall growth for SDAc switch variants. Model fits to altruist fraction dynamics from continuous growth in cellobiose for **(a)** no cheaters, **(c)** differentiation escapees and **(e)** both differentiation and lysis escapees. Model fits to growth rate dynamics for **(b)** no cheaters, **(d)** differentiation escapees and **(f)** both differentiation and lysis escapees. Data points represent the means from 3 biological replicates (N=3), and error bars one standard deviation from the mean.

APPENDIX 2

Chapter 3 – Supplementary materials

A2.1 MATERIALS AND METHODS

EXPERIMENTAL METHODS

Oligonucleotides

The sequences and modifications for all oligonucleotides used can be found in Auxiliary Table S.3[1] at (https://www.biorxiv.org/content/10.1101/869248v1).

Bacterial Culture

Escherichia coli MW1255, a derivative of MG1655 (E.C.), and *Bacillus subtilis* PY79 (B.S.) overnight cultures were inoculated into fresh LB medium at 1:1000 or 1:250, respectively, and grown at 37C with shaking (200rpm). For the heat shock experiment, upon reaching the OD_{600} = 0.5 half of each culture was transferred to the separate 37C incubator where the temperature was increased to 47C, and kept there with shaking (200rpm) for 8 minutes from the time the temperature stabilized. Both control and heat-treated samples then were immediately centrifuged at 4C, 5000 rcf for 5

minutes, and resuspended in cold formaldehyde. For the B.S. growth curve, samples were centrifuged at room temperature, 10000 rpm for 1 minute before fixation.

Fixation and permeabilization

For the steps below, centrifugation was performed at 4C, 5000 rcf for 5 minutes. Following centrifugation, the bacterial pellet was resuspended in 1mL of fresh, cold, 4% formaldehyde solution (in 1x PBS) and incubated at 4C overnight. The next morning, cells were centrifuged and resuspended in 1ml cold 100mM Tris-HCL+RI ('RI' indicates that SUPERase-In RNase Inhibitor, ThermoFisher, was added to a final concentration of 0.1 U/uL). Cells were centrifuged and resuspended in 250ul of 0.04% Tween-20 in 1x PBS, then permeabilized for 3 min on ice. We then added 1ml of cold PBS+RI, centrifuged the cells and resuspended in 200ul lysozyme mix per sample on ice as follows: 0.1M Tris-HCL pH7, 0.05M EDTA, 2.5mg/ml lysozyme, 0.25U/ml SUPERase-In. We incubated the samples at 37C in the thermocycler for 15 minutes as we found that precise timing of lysozyme incubation is critical to maintaining cell integrity at later stages of the protocol. Following the cell wall digestion step, we immediately added 1ml of cold PBS+RI, centrifuged the cells and counted the cells stained with SYTO9 using Accuri C6 flow cytometer. For the species-mixing experiment, we mixed the E.C. and B.S. cells at equal proportions and took 0.6M cells for each of heat shocked and control samples. For the B.S. growth curve experiments, we took 0.25M cells for each OD sample.

In-cell Polyadenylation

In order to enrich for mRNA capture, we performed *in situ* polyadenylation with *E. coli* Poly(A) Polymerase I (PAP) from NEB. For 0.25M cell samples, the reaction proceeded in 50µl volume, and in 100µl volume for the 0.6M cell samples. For each 0.6M cells bacterial pellet, we added 66µl of water, 4µl of SUPERase-In, 10µl of 10x PAP Buffer, 10µl of 10mM ATP, and 10µl of PAP. The reaction mixture was incubated at 37C for 30min, then centrifuged upon addition of 1ml cold PBS+RI. We also added 1µL of 10% Tween-20 in order to make the cells easier to pellet. The cells then were resuspended in 0.5ml of cold PBS+RI.

Reducing aggregate formation

Two steps were crucial to break down cell aggregates and reduce the doublet rate: first, we vortexed and double-filtered the bacterial cells prior to reverse transcription; and second, we performed sonication with double filtration right after the reverse transcription.

Following the polyadenylation step, cells in 500ul of PBS-RI were vortexed for 1 minute on the highest setting, filtered through 10µM pluriStrainer (pluriSelect) by pipetting through the membrane, then filtered again through 1µM pluriStrainer with gentle suction, and finally, right before adding to the reverse transcription wells, vortexed again on the highest setting for 1 minute.

Following the reverse transcription step and after resuspension in 2 mL of cold PBS+RI, cells were vortexed for 1 minute on the highest setting, filtered through 10µM and 1µM pluriStrainer as above, and briefly sonicated at 10% power for 5s on ice for 1 pulse (Sharpertek Ultrasonic Cell Crusher) followed by immediate distribution to the

ligation plate. We found that the sonication step can be replaced by a second vigorous vortexing step, with the roughly 2-fold increase in resulting detected doublet rate (0.7% to 1.3%).

In-cell Reverse Transcription

Like in SPLiT-seq[2], the first round of barcoding occurs through an *in situ* reverse transcription (RT) reaction. Cells are split into up to 48 wells, each containing barcoded well-specific reverse transcription primers. We used both random hexamer and anchored poly(dT)$_{15}$ barcoded RT primers in each well at the ratio of 1:2 (2.5μM ramdom hexamer + 5μM poly(dT)$_{15}$). In addition to primers, each RT well had a mix of 1X RT Buffer, 0.25U/μL RNase Inhibitor (Enzymatics), 0.25U/μL SUPERase-In RNase Inhibitor, 500μM dNTPs each (ThermoFisher), 7.5% of PEG8000, 20U/μL of Maxima H Minus Reverse Transcriptase (ThermoFisher). We pipetted 4 μL of cells at about 1M cells per mL in PBS-RI into every well, in the total resulting RT reaction volume of 20ul. The plate was incubated in a thermocycler for 10 min at 23˚C followed by 50˚C for 50 min. RT reactions were pooled back together and after adding 9.6 μL of 10% Triton X-100, cells were centrifuged for 5 min at 3000g at 4˚C in a swinging bucket rotor centrifuge. The supernatant was removed and cells were resuspended in 2 mL of cold 1X PBS-RI. The cells then underwent two rounds of filtration and sonication as described above.

Preparing Oligonucleotides for Ligations

The oligonucleotide plates for the second and third barcoding round were prepared as previously described[2].

In-cell Ligations

We prepared a 2.04 mL ligation mix containing 727.5 μL of RNase-free water, 500 μL 10X T4 Ligase buffer (NEB), 20 μL T4 DNA Ligase (2000 U/μL, NEB), 30 μL RNase inhibitor (40 U/μL, Enzymatics), 12.5 μL SuperaseIn RNase Inhibitor (20 U/μL), and 750 μL of 50% PEG8000. This ligation mix and the 2 mL of sonicated and filtered cells in 1X PBS were added to a basin and mixed thoroughly to make a total of 4.04 mL.

The ligation steps were performed as in the SPLiT-seq protocol[2], except we did vigorous vortexing combined with the double filtration technique as described above where the protocol called for a filtration step.

Lysis and Sub-library Generation

After the third round of barcoding, we performed a final vigorous vortexing and double filtration step as described above. Then 70 μL of 10% Triton-X100 was added to the cell solution before spinning it down for 5 min at 3000G and 4˚C. We carefully aspirated the supernatant, leaving about 30 μL to avoid removing the pellet. We then resuspended the cells in 4 mL of wash buffer (4 mL of 1X PBS, 40 μL of 10% Triton X-100 and 10 μL of SUPERase-In RNase Inhibitor) and spun down for 5 min at 3000G and 4˚C. We then aspirated the supernatant and resuspended in 50 μL of PBS+RI. After counting cells, we aliquoted them into sublibraries (in 1.7 mL tubes). After adding the desired number of cells to each sublibrary, we brought the volume of each to 50 μL by adding 1x PBS and froze the cells at -80C overnight. Next morning, we flash-thawed the cells and added 50 μL of 2X lysis buffer (20 mM Tris (pH 8.0), 400 mM NaCl, 100 mM EDTA (pH 8.0), and 4.4% SDS) and 10 μL of proteinase K solution (20mg/mL). We

incubated cells at 55˚C for 2 hours with shaking at 200 rpm to lyse the cells and reverse the formaldehyde crosslinks.

Purification of cDNA

cDNA purification and bonding to streptavidin beads was performed according to the SPLiT-seq protocol[2].

Template Switch

Streptavidin beads with bound cDNA molecules were resuspended in a solution containing 99 µl nuclease-free water, 44 µL of 5X Maxima RT buffer (ThermoFisher), 33 µL of 50% PEG8000 solution, 22 µL of 10 mM dNTPs each (ThermoFisher), 5.5 µL of RNase Inhibitor (Enzymatics), 11 µL of Maxima H Minus Reverse Transcriptase (ThermoFisher), and 5.5 µL of 100uM of a template switch primer (BC_0127). The template switch primer contains two ribonucleic guanines followed by a locked nucleic acid guanine at the end of the primer (Exiqon). The beads were incubated at room temperature for 30 minutes and then at 42°C for 90 minutes with gentle shaking.

PCR

The on-beads PCR followed by the qPCR to amplify the product were performed according to the SPLiT-seq protocol[2].

Illumina Sequencing

Libraries were sequenced on MiSeq or NextSeq systems (Illumina) using 150

nucleotide (nt) kits and paired-end sequencing. Read 1 (74 nt) covered the transcript sequences. Read 2 (86 nt) covered the UMI and barcode combinations. The index read (6 nt), serving as the fourth barcode, covered the sublibrary indices introduced after fragmentation.

Fragmentation

Following cDNA amplification, molecules were fragmented with a subsequent adaptor ligation step using a protocol modified from Enzymatics. Briefly, 110 ng of amplified cDNA was placed into a 50 uL reaction containing 5 uL of 10x Fragmentation Buffer (Enzymatics) and 10 uL of 5X WGS Fragmentation Mix (Enzymatics). Samples were placed into a thermocycler with the following steps: 32C for 10 min, 65C for 30 min, 4C hold. Following fragmentation, a double sided SPRI bead size selection with bounds of 0.6x – 0.8x was performed, with the final elution step using a volume of 50 uL. This 50 uL of eluant was placed into a 100 uL reaction containing 20 uL of 5X Rapid Ligation Buffer (Enzymatics), 10 uL of WGS Ligase (Enzymatics), and 2.5 uL of a pre-annealed adaptor duplex consisting of BC_243 and BC_244 at a concentration of 100 uM in 50uM NaCl. This adaptor ligation mix was incubated at 20C for 15 minutes. This was followed by a 0.8x SPRI size selection with final elution in 20 uL. Next, 18.5 uL of eluant was placed into a 50 uL qPCR reaction containing 25 uL of 2X Kapa HiFi Master Mix, 2.5 uL of 20X Evagreen, 2 uL of BC_0027 (10 uM), and 2 uL of BC_0076-BC_0083 (10 uM). This PCR mix was placed in a thermocycler with following conditions: 95C for 3 minutes, beginning cycling of 98C for 20s, 67C for 20s, and 72C for 3 min. Cycles were allowed to continue on a qPCR machine until reaction neared

saturation, as denoted by the exit of exponential phase in amplification. Finally, samples were incubated at 72C for 5min once sufficient cycling occurred. After this reaction, a double-sided SPRI size selection was performed with bounds of 0.5x – 0.7x, where resulting 20 uL eluant was a sequencing-ready library.

COMPUTATIONAL METHODS

Alignment and generation of cell-gene matrices

The data preprocessing and alignment was performed using a modified SPLiT-seq pipeline[2], where the cDNA reads were mapped to either a combined B. S. – E. C genome (ASM49748v1.44 and ASM80076v1.37 from EnsemblBacteria[3]) or the *B. subtilis* 168 genome (ASM904v1.45 from EnsemblBacteria) using STAR with the splicing isoform detection switched off[4]. We chose to use the parent strain *B. subtilis* 168 genome for the growth curve experiment since the respective annotation file had gene names, which the PY79 strain annotation file lacked (using non-standard accession numbers instead). However, we note that using a different genome and annotation increased the total number of detected genes, largely due to the different annotation. In addition, we kept the multimapping reads, assigning a fractional count based on the number of alignments, since bacterial genomes are known to contain overlapping CDSs. We then generated a matrix of gene counts for each cell (N x K matrix, with N cells and K genes).

Processing of data from the heat shock experiment

Clustering and data analysis for the species-mixing experiment with heat shock treatment was performed using Scanpy[5]. We only kept transcriptomes with the number of

total reads higher than 200. Then, we removed the ribosomal and tRNA reads from the data, retaining only reads representing the mRNA counts for both species. We further filtered cells based on the mRNA counts, retaining cells expressing >100 reads and >100 genes, and additionally filtered the genes retaining the genes expressed in >5 cells. We then applied standard Scanpy normalization and scaling, dimensionality reduction, and clustering as described in the Scanpy tutorial[5,6] and below, using 9 PCA components and 45 neighbors for computing the neighborhood graph of cells. For plotting the heat map of the top 6 enriched genes for the heat shock experiment clusters (Figure S4B), we additionally filtered the genes by keeping only the genes expressed in at least 40% of the cluster, with a minimum fold change of 2 and in at most 30% of the rest of the data. For making the half-volcano plot of enriched genes in the *E. coli* heat shock cluster (Figure 1H), for better clarity we omitted the outlier norR gene with a very high fold change but low significance.

Processing of B. S. data from the growth in rich media experiment

Clustering and data analysis for the combined 10 samples of B.S. grown in rich medium was performed using Scanpy[4] and verified with Seurat v3[7] and UNCURL[8]. Experiment 1 sampled OD points 0.5, 1.0, 1.7, 2.0, 2.8, and 3.2, while in experiment 2 we collected OD points 0.5, 1.0, 1.3, 1.7, 2.8, 3.5, 5.3, and 6.0. For the data from both experiments separately, we discarded any transcriptomes with the number of total reads fewer than 200. Then, we filtered the data and retained only reads representing the mRNA counts (excluding the ribosomal and tRNA reads). Finally, we combined the data matrices together. Since the read depth decreased for the higher OD samples (Figure

S15), for selecting the highest quality data, we implemented differential thresholds for each OD in the combined matrix, retaining top 75% of the cells by read counts for each OD sample. This resulted in retention of 25,214 transcriptomes from both experiments. Finally, we performed batch correction through Scanpy, using a python implementation of ComBat[9,10].

Cells that passed the QC were clustered using a pipeline described in previous studies[2,6]. We chose to omit the high variance gene selection since the number of genes we detected (on the order of 3,000) was much lower than what is typically detected for mammalian cells and permitted analysis without confining the gene set. In addition, restriction to high variance genes did not result in a major change in the final clustering output. We followed with standard Scanpy normalization and scaling, dimensionality reduction, and clustering[5,6]. Briefly, we computed the neighborhood graph of cells using the top 16 PCA components of the data matrix (n-neighbors = 8). We then used Louvain graph-clustering method to produce the global clustering of the data. For the two-dimensional embedding of the data, we chose to use the t-distributed Stochastic Neighbor Embedding (t-SNE) using a *scikit-learn* implementation[11] of the Barnes-Hut t-SNE algorithm[12]. Top enriched genes for each cluster were computed by Wilcoxon rank-sum test with Benjamini-Hochberg correction.

Sub-clustering of the late OD samples

For finding the finer grain structure within the late OD point data, we took filtered read matrices from cells belonging to OD 5.3 and 6.0 groups and re-run the processing pipeline as described above. Because of the smaller numbers of cells, we

chose to use the UMAP algorithm[13] to embed the data.

Calculation of Sigma Factors and Transcriptional Regulators Activity

To generate the profiles of activities of sigma factors and transcriptional regulators (TR), we used the *Subti*Wiki resource[14]. For the TR analysis, each regulon was divided into "positive" and "negative" subregulons based on the mode of regulation (we omitted the rare more complex interactions and focused on the more straightforward transcriptional activation or repression), and we then calculated the average expression of genes within each subregulon for each cluster.

Materials and Methods References

1. Kuchina, A. and Brettner, L. *et al.* Microbial single-cell RNA sequencing by split-pool barcoding. *bioRxiv* (2019) doi:10.1101/869248.
2. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* eaam8999 (2018) doi:10.1126/science.aam8999.
3. Howe, K. L. *et al.* Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res* doi:10.1093/nar/gkz890.
4. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
5. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
6. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
7. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
8. Mukherjee, S., Zhang, Y., Fan, J., Seelig, G. & Kannan, S. Scalable preprocessing for sparse scRNA-seq data exploiting prior knowledge. *Bioinformatics* **34**, i124–i132 (2018).
9. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
10. Pederson, Brent. *patsy version of ComBat for removing batch effects.* (GitHub, 2012).
11. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

12. Maaten, L. van der. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 3221–3245 (2014).
13. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* (2018).
14. Zhu, B. & Stülke, J. SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res* **46**, D743–D748 (2018).

A2.2 SUPPLEMENTARY TABLES

Fixation	Permeabilization	rRNA depletion	Reverse transcription
100% Ethanol	0.5%Triton	Ribosomal RT + RNAseH	Maxima RT + Ficoll
50% Ethanol	Bead beater	TEX Buffer A	Superscript IV RT
4% Formalin + 100% Ethanol	Melittin	TEX Buffer B	Smartscribe RT
4% Formalin + 50% Ethanol	Lysozyme (1mg\ml, 30min)	TEX Buffer A 2x	Maxima RT 50C 15min 60C 15min
1.33% Formalin + 100% Ethanol	Lysozyme (5mg\ml, 30min)	Poly A Polymerase (PAP) + TEX Buffer A	Maxima RT 50C for 30 min
1.33% Formalin	Lysozyme (20mg\ml, 15min)	Poly A Polymerase (PAP)	Maxima RT 50C for 1hr
4% Formalin for 10 min	0.5%Triton + Lysozyme (20mg\ml, 15min)		Maxima RT 50C for 4hrs
4% Formalin for 1 hr	0.5%Tween + Lysozyme (5mg\ml, 15min)		Maxima RT Cycling RT (SPLiT-seq)
4% Formalin overnight	0.04%Tween + Lysozyme (2.5mg\ml, 15min)		Maxima RT 23C 10 min 50C 50 min

Table 3.1. Tested fixation, permeabilization, mRNA enrichment, and RT conditions.

Function	Operon	Gene	…	Function	Operon	Gene
antimicrobial	subtilosin	sboA		motility	fla-che operon	fliP
antimicrobial	subtilosin	sboX		motility	fla-che operon	fliQ
antimicrobial	subtilosin	albA		motility	fla-che operon	fliR
antimicrobial	subtilosin	albB		motility	fla-che operon	flhB
antimicrobial	subtilosin	albC		motility	fla-che operon	flhA
antimicrobial	subtilosin	albD		motility	fla-che operon	flhF
antimicrobial	subtilosin	albE		motility	fla-che operon	flhG
antimicrobial	subtilosin	albF		motility	fla-che operon	cheB
antimicrobial	subtilosin	albG		motility	fla-che operon	cheA
antimicrobial	bacillaene	pksC		motility	fla-che operon	cheW
antimicrobial	bacillaene	pksD		motility	fla-che operon	cheC
antimicrobial	bacillaene	pksE		motility	fla-che operon	cheD
antimicrobial	bacillaene	pksF		motility	fla-che operon	sigD
antimicrobial	bacillaene	pksG		motility	fla-che operon	swrB
antimicrobial	bacillaene	pksH		stress response	mcs kinase	mcsA
antimicrobial	bacillaene	pksI		stress response	mcs kinase	mcsB
antimicrobial	bacillaene	pksJ		stress response	clpP proteases	clpC
antimicrobial	bacillaene	pksL		stress response	clpP proteases	clpP
antimicrobial	bacillaene	pksM		stress response	clpP proteases	clpE
antimicrobial	bacillaene	pksN		stress response	clpP proteases	clpX
antimicrobial	bacillaene	pksR		stress response	chaperonin	groEL
antimicrobial	bacillaene	pksS		stress response	chaperonin	groES
antimicrobial	bacilysin	bacB		metal import	manganese uptake	mntA
antimicrobial	bacilysin	bacC		metal import	manganese uptake	mntB
antimicrobial	bacilysin	bacD		metal import	manganese uptake	mntC
antimicrobial	bacilysin	bacE		metal import	manganese uptake	mntD
antimicrobial	bacilysin	bacH		metal import	manganese uptake	mntH
antimicrobial	plipastatin	ppsA		metal import	bacillibactin	dhbA
antimicrobial	plipastatin	ppsB		metal import	bacillibactin	dhbC
antimicrobial	plipastatin	ppsC		metal import	bacillibactin	dhbE
antimicrobial	plipastatin	ppsD		metal import	bacillibactin	dhbB
antimicrobial	plipastatin	ppsE		metal import	bacillibactin	dhbF
antimicrobial	endoA/I	ndoA		metal import	siderophore uptake	feuA
antimicrobial	endoA/I	ndoAI		metal import	siderophore uptake	feuB
antimicrobial	spore killing factor	skfA		metal import	siderophore uptake	feuC
antimicrobial	spore killing factor	skfB		metal import	siderophore uptake	fhuB
antimicrobial	spore killing factor	skfE		metal import	siderophore uptake	fhuC
antimicrobial	spore killing factor	skfF		metal import	siderophore uptake	fhuD
antimicrobial	spore killing factor	skfG		metal import	siderophore uptake	fhuG
antimicrobial	spore delay protein	sdpB		metal import	siderophore uptake	yxeB
antimicrobial	spore delay protein	spbC		carbon metabolism	lichenan	licA
motility	surfactin	srfAA		carbon metabolism	lichenan	licB
motility	surfactin	srfAB		carbon metabolism	lichenan	licC
motility	surfactin	srfAC		carbon metabolism	lichenan	licH

motility	surfactin	srfAD	carbon metabolism	acetoin	acoA
motility	flagellin protein (hag)	hag	carbon metabolism	acetoin	acoB
motility	fla-che operon	flgB	carbon metabolism	acetoin	acoC
motility	fla-che operon	flgC	carbon metabolism	acetoin	acoL
motility	fla-che operon	fliE	carbon metabolism	mannose	manA
motility	fla-che operon	fliF	carbon metabolism	mannose	manP
motility	fla-che operon	fliG	carbon metabolism	gapB	gapB
motility	fla-che operon	fliH	carbon metabolism	citB	citB
motility	fla-che operon	fliI	carbon metabolism	iolAJ	iolA
motility	fla-che operon	fliJ	carbon metabolism	iolAJ	iolB
motility	fla-che operon	ylxF	carbon metabolism	iolAJ	iolC
motility	fla-che operon	fliK	carbon metabolism	iolAJ	iolD
motility	fla-che operon	flgD	carbon metabolism	iolAJ	iolE
motility	fla-che operon	flgE	carbon metabolism	iolAJ	iolF
motility	fla-che operon	fliL	carbon metabolism	iolAJ	iolG
motility	fla-che operon	fliM	carbon metabolism	iolAJ	iolH
motility	fla-che operon	fliY	carbon metabolism	iolAJ	iolI
motility	fla-che operon	cheY	carbon metabolism	iolAJ	iolJ
motility	fla-che operon	fliZ	carbon metabolism	iolRS	iolR
			carbon metabolism	iolRS	iolS
			carbon metabolism	iolT	iolT

Table 3.2. Genes and operons used in select carbon metabolism and lifestyle analyses.

A2.3 SUPPLEMENTARY FIGURES

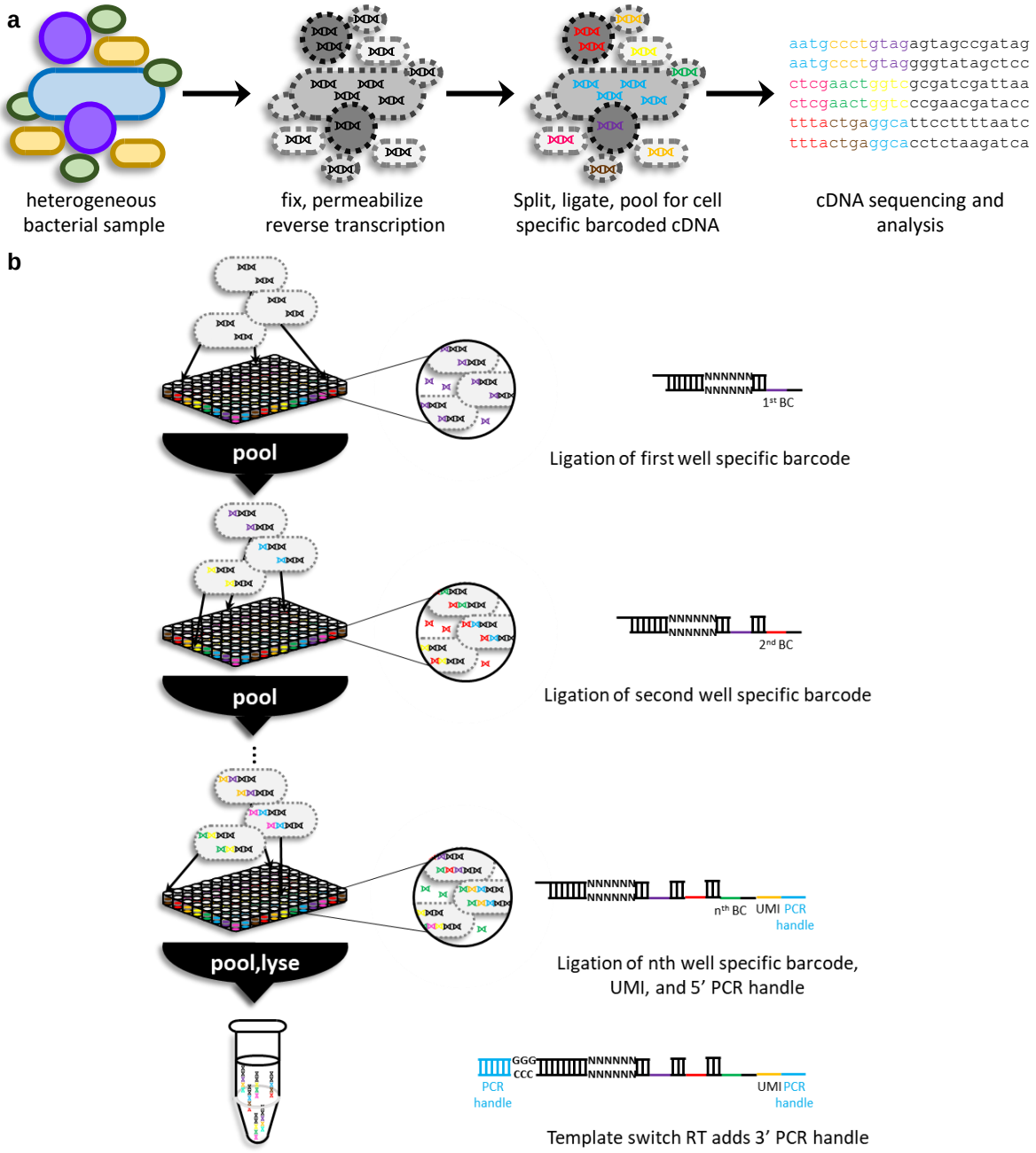


Figure 3.1. – SPLiT seq methodology and mechanisms figure supplement 1. (a) A mixed bacterial sample is fixed using 4% formaldehyde (1.33% formalin). The cells are then permeabilized using ethanol, lysozyme enzyme, a bead beater apparatus, or a combination thereof. *In situ* reverse transcription and combinatorial barcode ligation generates a pool of uniquely tagged cells. The barcoded cDNA is then sequenced using next generation sequencing. **(b)** An intracellular cNDA library is generated from cellular RNA using random hexamer primers and reverse transcriptase. The cells are then split into a multi-welled plate with well specific barcodes. A ligation reaction attaches the nth barcode to the 5' end of the cDNA in each cell aliquoted to a given well. In the final ligation step, a unique molecular identifier (UMI) and 5' PCR handle are also attached. The UMI ensures that each cDNA is counted only once during sequencing, even after amplification. The cells are then lysed, and a template switch reaction attaches a universal 3' PCR handle. Adapted from Rosenberg et al. 2017.

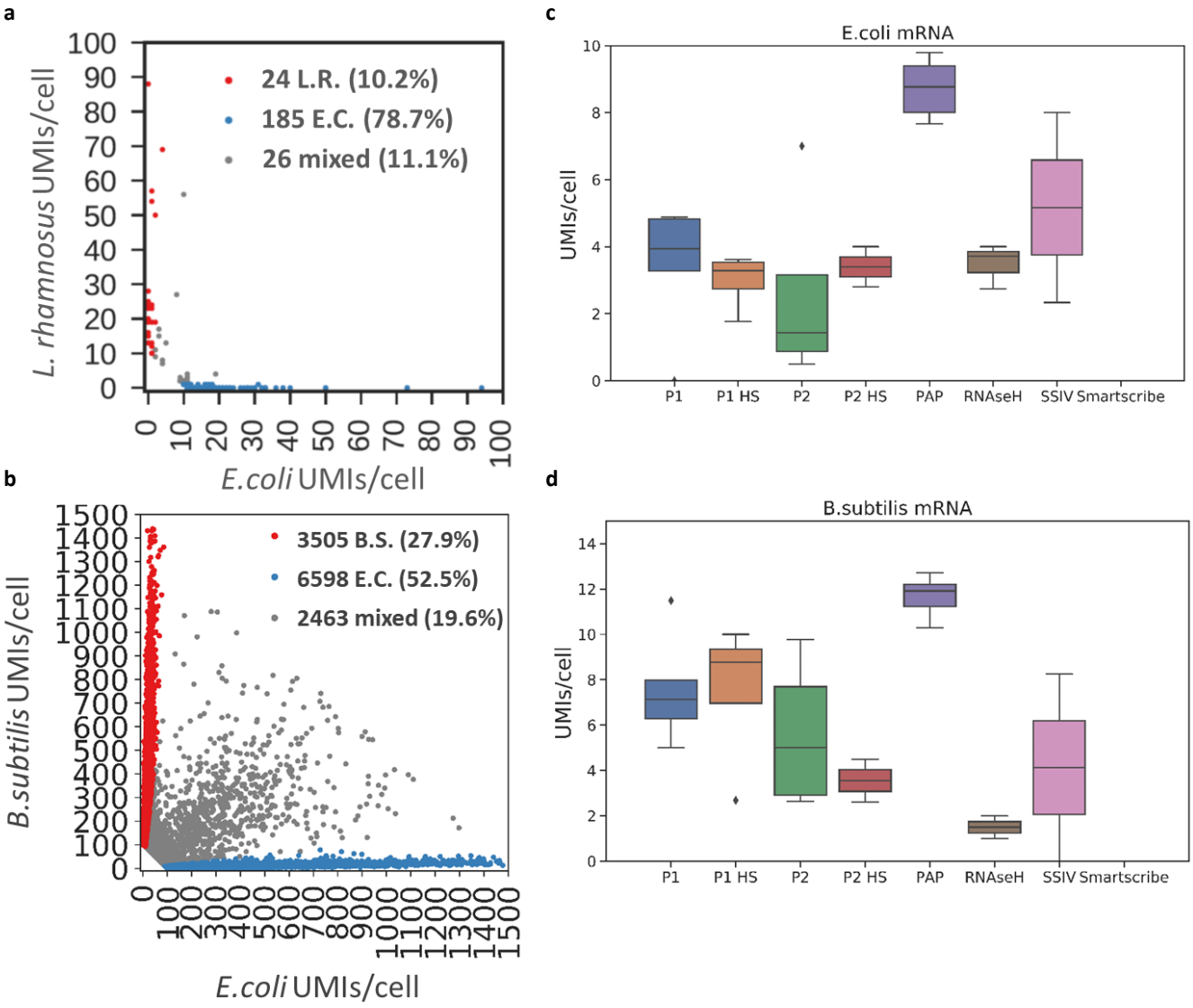


Figure 3.1 – figure supplement 2. microSPLiT optimization progression. **(a)** A barnyard plot showing each putative single-cell transcriptome as a dot aligned to the combined *L. rhamnosus* –*E. coli* (E.C.) genome. The cells with <10 reads were filtered out. UMI – unique molecular identifier. (**b**) A barnyard plot showing each putative single-cell transcriptome as a dot aligned to the combined B.S. –E.C. genome. The cells with <100 reads were filtered out. The mixed cells are expected to originate from the physical aggregates traveling through barcoding reactions together. **(c,d)** Sample test of differential conditions for permeabilization and mRNA enrichment for **(c)** *E. coli* and **(d)** *B. subtilis* read counts. UMI – unique molecular identifier. Conditions tested: "P1" - 0.5%Triton-X100 + Lysozyme (5mg\ml) for both species; "P2" - 0.5%Triton-X100 for *E. coli* and 0.5%Triton-X100 + Lysozyme (20mg\ml) for *B. subtilis*; "HS" – heat shocked samples; "PAP" – *E. coli* Poly(A) Polymerase treatment; "RNAseH" – reverse transcription (RT) with rRNA-specific probes followed by RNAseH digestion; "SSIV" – SuperScript IV reverse transcriptase (ThermoFisher) instead of Maxima reverse transcriptase (ThermoFisher); "Smartscribe" – SMARTScribe reverse transcriptase (Takara Bio) instead of Maxima reverse transcriptase (ThermoFisher). For the latter three conditions "P1" permeabilization was used.

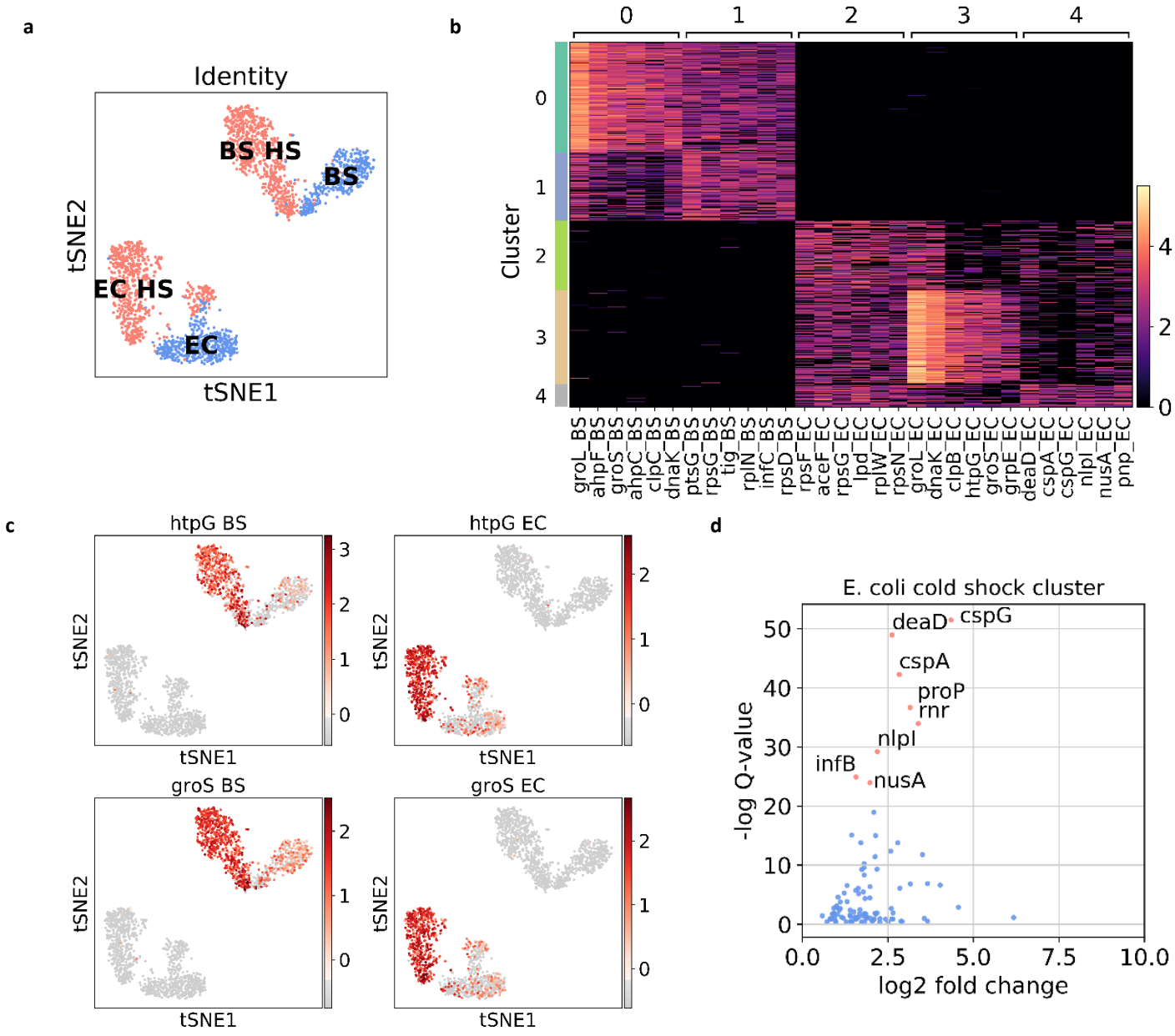


Figure 3.1 -- figure supplement 3. Transcriptional responses of *E. coli* and *B. subtilis* cells to heat shock. **(a)** t-SNE plot showing cluster assignments based on the first barcode label ("HS" indicates heat shocked samples) and the genes expressed in the cluster ("EC" – *E. coli*; "BS" – *B. subtilis*). **(b)** Top 6 genes enriched for each cluster in Figure 1F). **(c)** Expression of select classical heat shock markers for both *E. coli* and *B. subtilis* overlayed on the t-SNE. **(d)** Genes enriched in cluster 4, *E. coli* cold shock cluster.

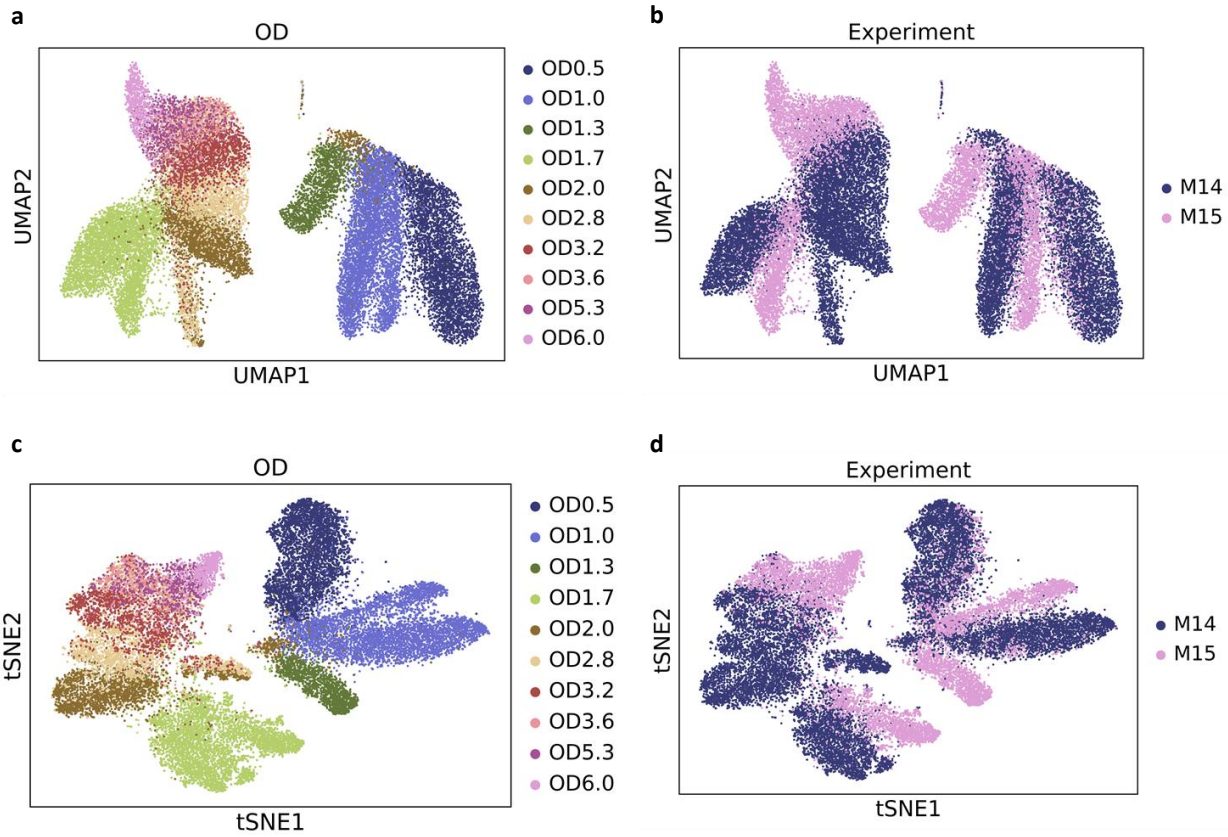


Figure 3.2 – figure supplement 1. 25,214 *B. subtilis* cells sampled across the growth curve. **(a,b)** UMAP embedding of data from the rich medium growth experiment, colored by either OD label (a) or by experiment (b, "M14" – experiment 1, "M15" – experiment 2). **(c,d)** As in (a,b), but with the t-SNE embedding.

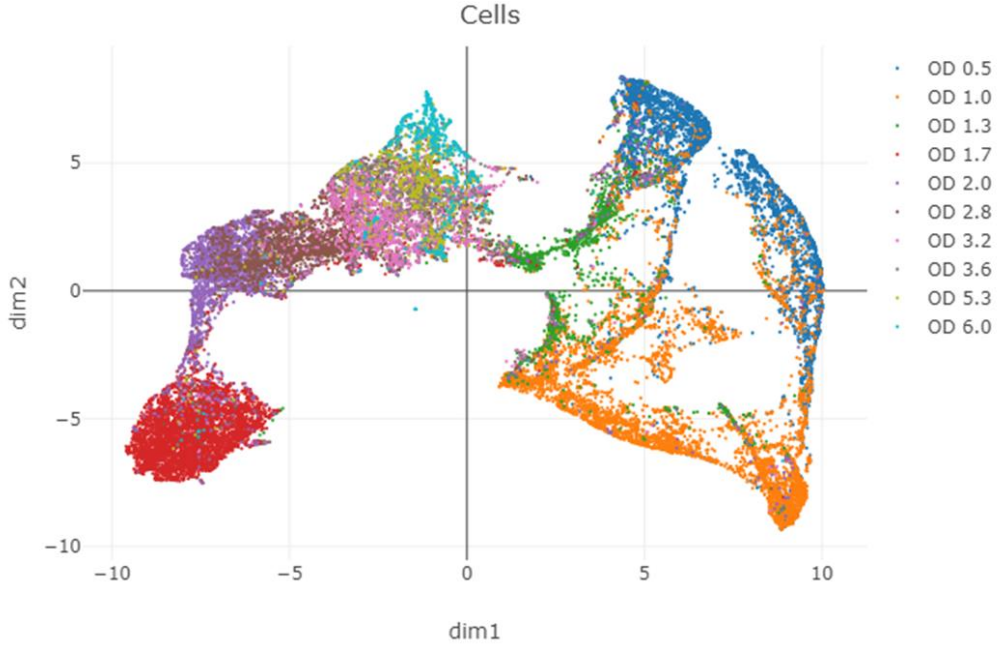


Figure 3.2 – figure supplement 2. UNCURL processing and embedding of data from the rich medium growth experiment, colored by the OD label.

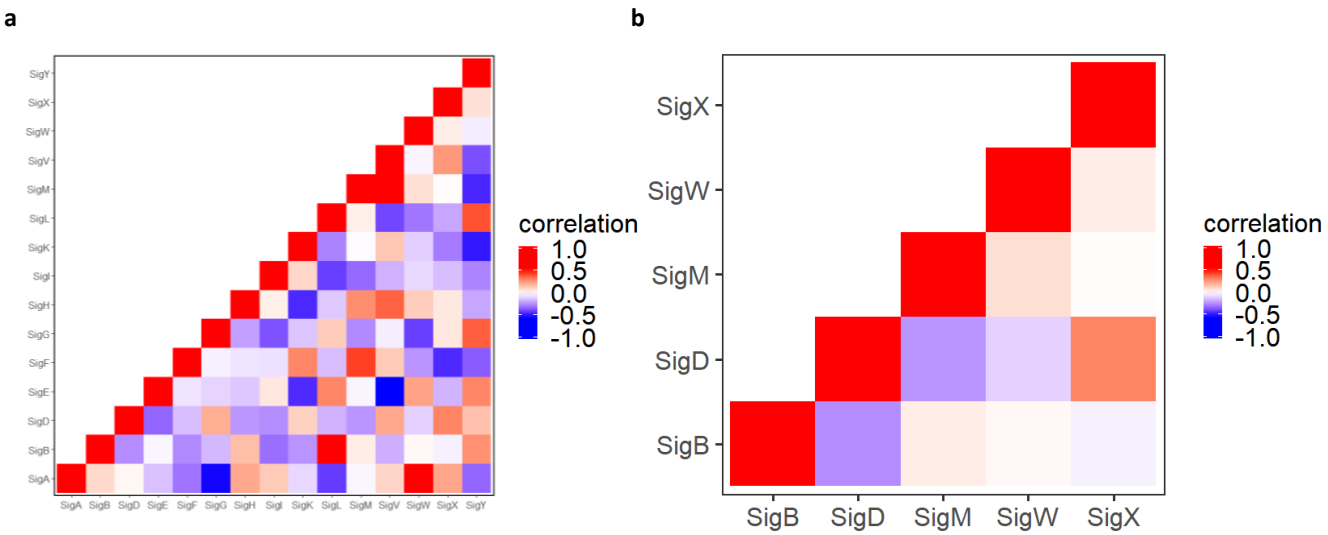


Figure 3.2 – figure supplement 3. Expression correlation of genes regulated by alternative sigma factors. The mean expression for each sigma regulon was calculated and normalized for each of the 25,214 cells and correlated to the mean expression of the other regulons. (**a**) shows a correlation heatmap for the 16 central alternative sigma factors. (**b**) shows the correlations of a select set of sigma factors described in (Park *et al*. 2018).

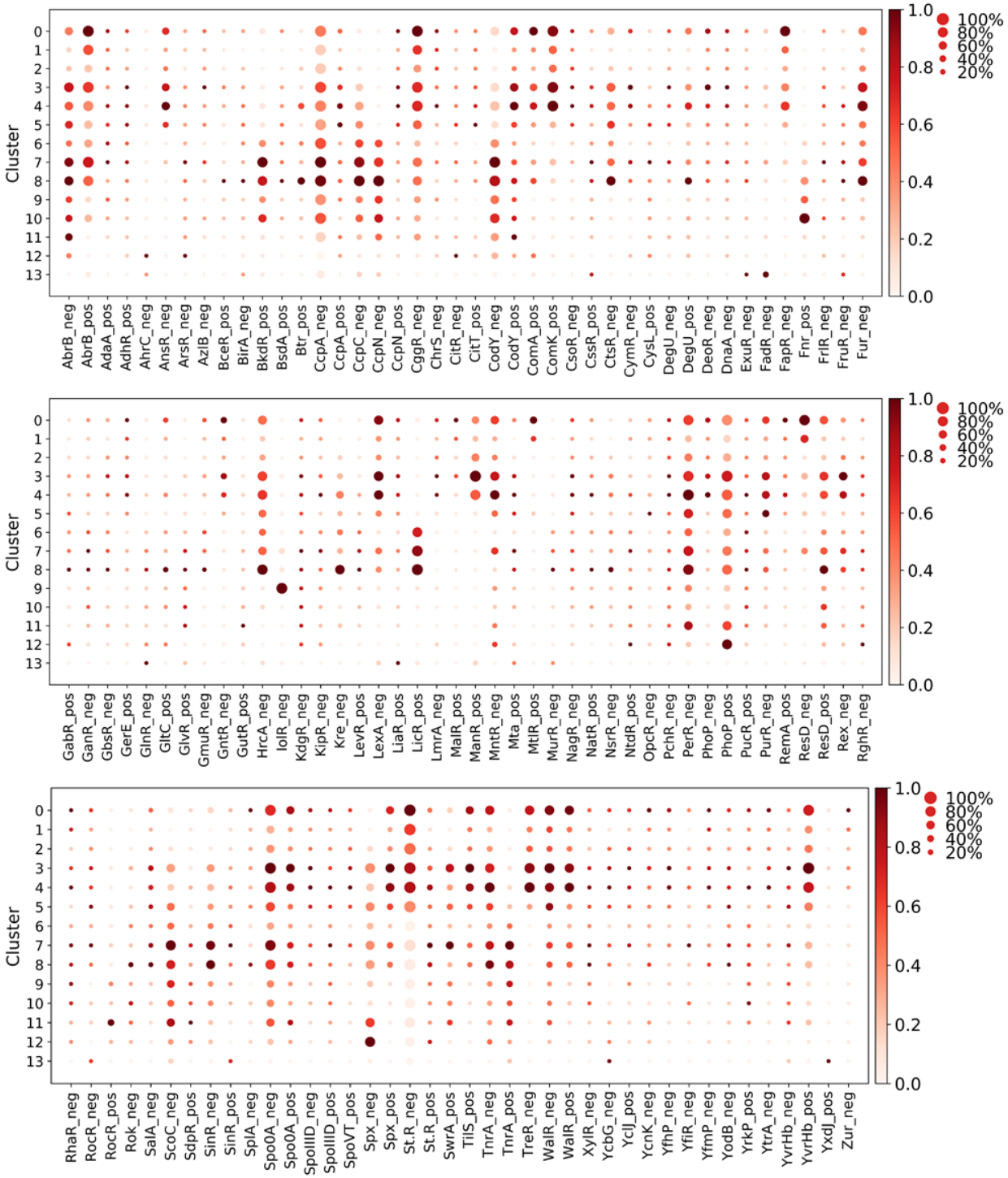
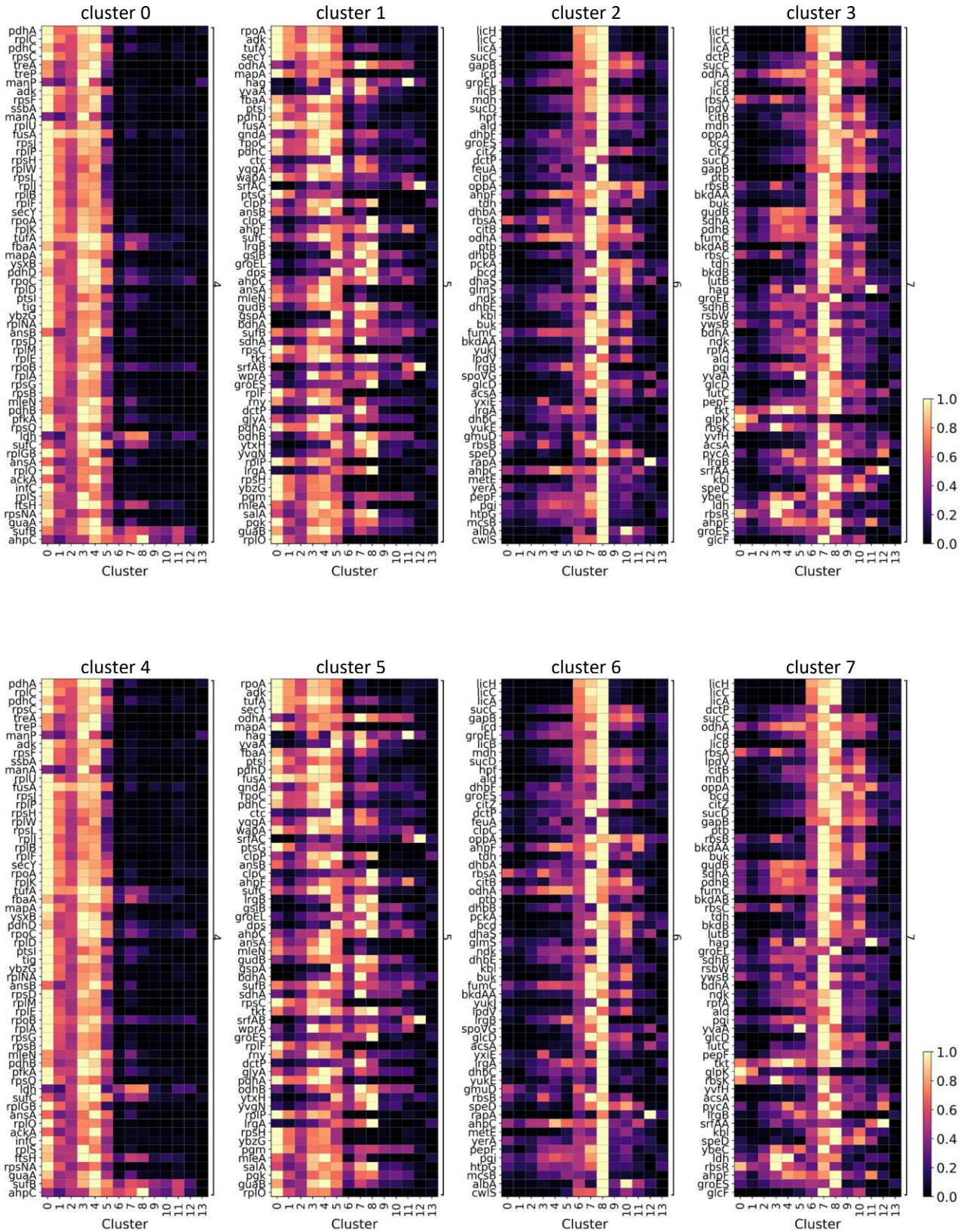


Figure 3.2 – figure supplement 4. Normalized detected transcriptional regulators (TR) activity per cluster (see Materials and Methods). The size of each dot indicates the proportion of cells in the cluster in which the sigma factor is active, while the color indicates the average activity. “Neg” for a given TR indicates that the activity was calculated for the genes in the respective regulon negatively regulated by this TR, and “pos” indicates the activity was calculated for the genes whose expression is positively regulated by the given TR.


cluster 0
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
cluster 6
cluster 7
Cluster
1.0
0.8
0.6
0.4
0.2
0.0

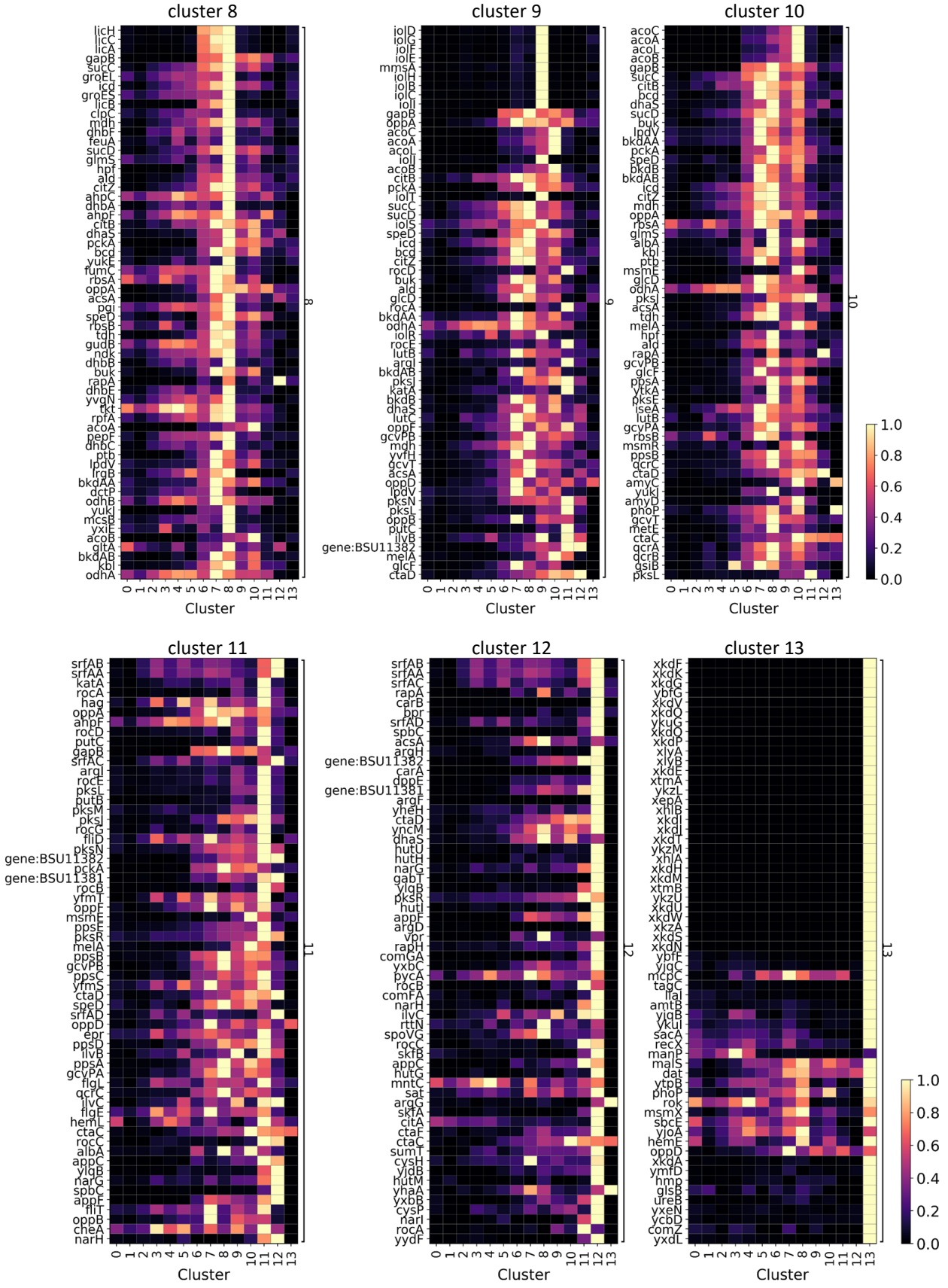


Figure 3.3 – figure supplement 5. Top 60 expressed genes for clusters 0-13

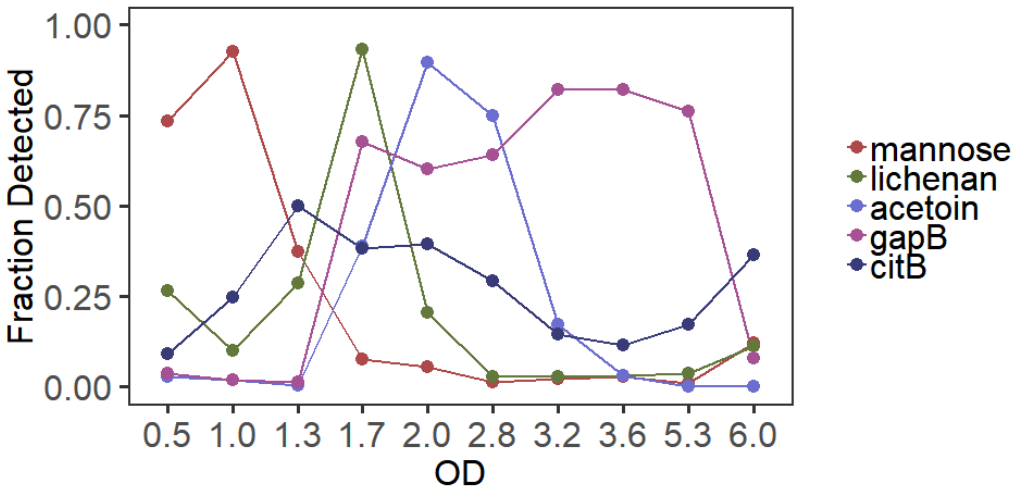


Figure 3.3 – figure supplement 2. Fraction of cells expressing the indicated carbon metabolism genes or operons as a function of OD. The genes used to generate the curves are listed in Table 3.2.

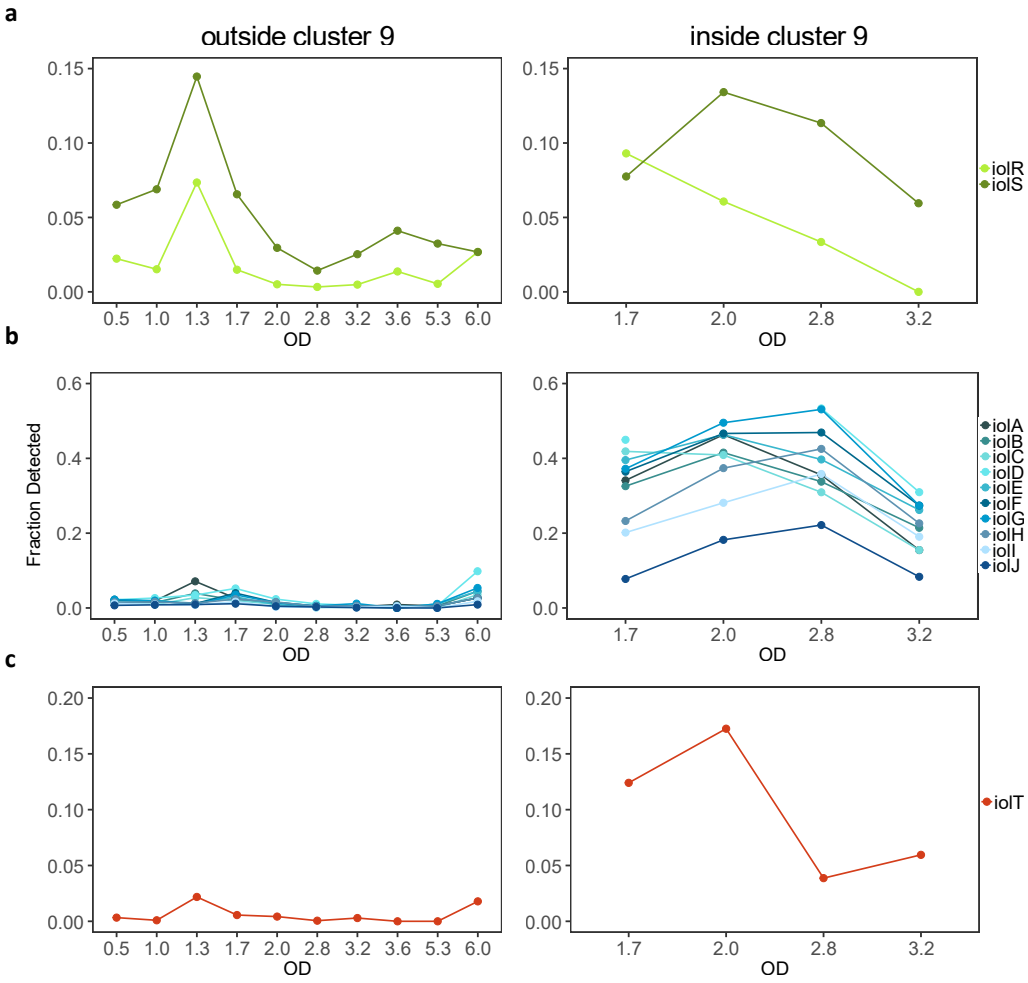


Figure 3.4 – figure supplement 1. Fraction of cells detected to express each of the three inositol utilization operons. (**a**) iolRS, (**b**) iolAJ, (**c**) iolT operon expressing cells fraction plotted for cells either outside or inside cluster 9 across OD measurements. Cells were classified as expressing if we detected at least one transcript from any of the genes in operon. Note that within cluster 9, only ODs 1.7-3.2 are shown as only those samplings had meaningful (>5) numbers of cells included in this cluster.

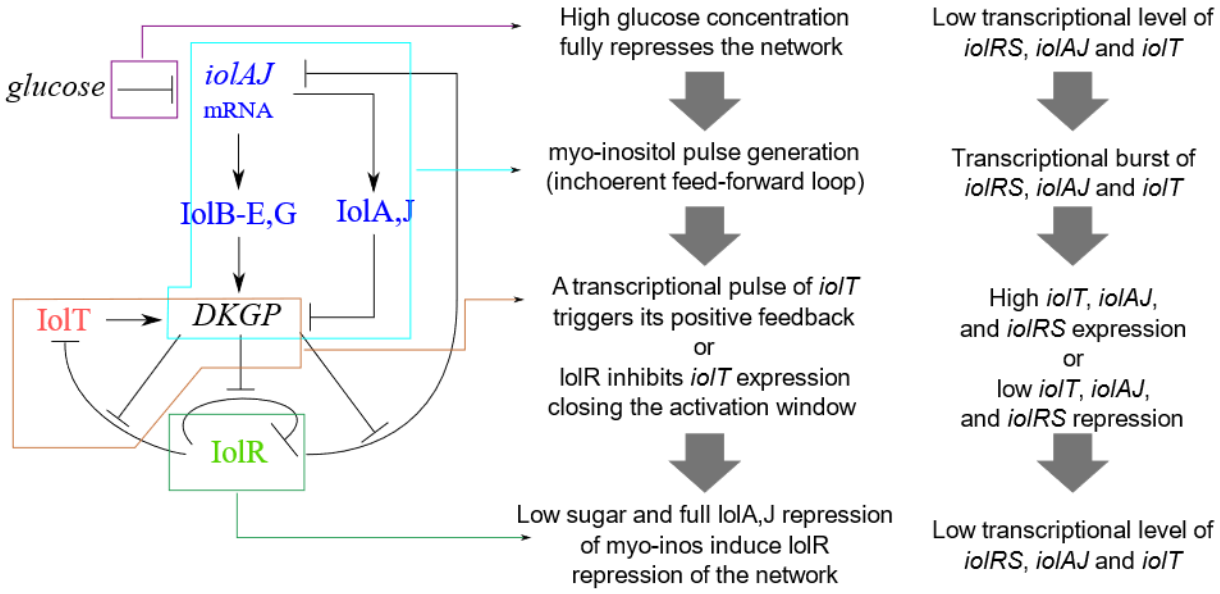


Figure 3.4 – figure supplement 2. Myo-inositol utilization pathway diagram and model description with potential intracellular states that explain the expression heterogeneity we observe.

Figure 3.5 – figure supplement 1. Fraction of cells expressing the indicated Skf and Sdp family toxins as a function of OD. The genes used to generate the curves are listed in Table 3.2.