High-throughput single-cell sequencing of multiple invertible promoters reveals a strong determinant of bacterial population heterogeneity

Freeman Lan[1+], Jason Saba[1,2,3+], Yili Qian[1], Tyler Ross[1,4], Robert Landick[1,2*] and Ophelia S Venturelli[1,2,4,5*]

[1]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53726

[2]Department of Bacteriology, University of Wisconsin-Madison, WI 53726

[3]Microbiology Doctoral Training Program, University of Wisconsin-Madison, WI

[4]Department of Biomedical Engineering, University of Wisconsin-Madison, WI

[5]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison WI

[+]Denotes equal contribution

*To whom correspondence should be addressed: venturelli@wisc.edu, landick@wisc.edu

Summary sentence: High-throughput single cell sequencing of phase variation reveals regulation as a major driver of population diversification

ABSTRACT

Population heterogeneity can promote bacterial fitness in response to unpredictable environmental conditions. Human gut symbiont *Bacteroides* spp., displays variability in single cell surface architectures by combinatorial regulation of promoter inversions that drive expression of capsular polysaccharides (CPS). Using high-throughput single-cell sequencing, we reveal population heterogeneity generated through combinatorial promoter inversion and show that *B. fragilis* populations can access diverse CPS promoter states. Combining our data with stochastic computational modeling, we demonstrate that the rates of promoter inversion regulated by a broadly conserved serine recombinase are a major mechanism shaping population heterogeneity. Exploiting control over the expression of the recombinase, we devise a method for creating phase-locked variants and show that populations with different initial compositions converge to a similar steady-state composition over time. Our approach can be used to interrogate single-cell phase variable states of diverse microbes including bacterial pathogens.

MAIN TEXT:

Gene regulation in bacteria is shaped by stochastic and deterministic processes. When faced with changing and uncertain conditions, establishing a phenotypically diverse population *a priori* to "hedge" against different environmental stresses can promote population survival. For example, the establishment of persister populations in bacterial pathogens can promote survival under antibiotic treatment (1). The composition (fraction of the population occupying each state) of these heterogeneous populations is influenced by the underlying regulatory mechanisms. Understanding how bacteria regulate the phenotypic heterogeneity at the population-level can reveal important insights into their mechanisms for survival and adaptation.

Phase variation is a ubiquitous mechanism for generating phenotypic heterogeneity in host-associated bacteria such as pathogens and symbionts. (2–8). This process works through diverse mechanisms, such as reversible DNA recombination, to create subpopulations with distinct patterns of gene expression (9). In symbiotic *Bacteroides spp.,* phase variation of capsular polysaccharide (CPS) generates specialized and multi-functional subpopulations having alternative propensities for biofilm development (10), resistance to antibiotics (11), and protection from diverse phages (11,12).

Despite the importance of phase variation to bacterial fitness in diverse environments, limitations in technology have hindered a fundamental understanding of the landscape of single-cell states generated by combinatorial phase variation. Additionally, previous studies of phase variation used fluorescent reporters that disconnect the loci from their natural genetic context and limit the number of loci that can be observed simultaneously (4,13–15). Furthermore, fluorescent

68 reporters produce signals that are temporally disconnected from phase variation due to time
69 delays in fluorescent protein maturation and decay (16).
70 In the abundant and prevalent genus in the human gut microbiome, *Bacteroides spp.*, a
71 broadly conserved serine recombinase controls promoter inversion at 7–13 CPS biosynthetic loci
72 (16–18). *Bacteroides fragilis* (*B. fragilis*), a key symbiont, has seven invertible CPS loci (A, B, D,
73 E, F, G, and H), the orientations of which determine the landscape of surface architectures for
74 single cells. An eighth CPS promoter C is constitutively in the ON state to inhibit formation of lower
75 fitness acapsular subpopulations. Omitting C for conciseness in labeling, promoter orientations
76 define discrete ON/OFF states of distinct CPS synthesis operons (e.g., A-E--- denotes A and E in
77 the ON state and B, D, F, G, H in the OFF state) (**Figure 1A**).
78 Because tools to quantify multiple promoter states within a large number of single cells
79 have not been available, fundamental questions about combinatorial phase variation in any
80 bacterial species remain unanswered. Here, we use a new single-cell sequencing method to
81 investigate the contribution of regulation to combinatorial CPS promoter states within a *B. fragilis*
82 population. Is the population composition of CPS states tightly constrained or loosely influenced
83 by the intrinsic mechanisms of phase variation (in this case, rates of recombination)? In this
84 system, what is the landscape of promoter state compositions that are generated by phase-
85 variation (**Figure 1B**)? Are there one or multiple steady-state compositions or is the system
86 dominated by stochastic processes such that no particular population composition is strongly
87 favored? Can the inversion rates be resolved using single-cell time-series measurements of
88 single-cell states across the population and are these rates coupled or independent? A more
89 detailed and quantitative understanding of this system could provide insights into the contributions
90 of stochastic processes, regulatory network architectures, and the role of history-dependence in
91 shaping the dynamic behaviors of phenotypic diversification by combinatorial phase variation.
92
93 *High-throughput single-cell sequencing of multiple invertible promoters*
94 To investigate *B. fragilis* population diversification at the single-cell level, we used an ultrahigh-
95 throughput, single-cell sequencing method referred to as Droplet Targeted Amplification
96 sequencing (DoTA-seq) (19). DoTA-seq can analyze combinations of genes in diverse gram-
97 positive and gram-negative bacteria at the single-cell level. Here we simplified the protocol for
98 profiling the 7 CPS promoter ON/OFF states in *B. fragilis* (gram-negative) **(Methods)**. Briefly,
99 single *B. fragilis* cells and random barcode oligos are encapsulated at limiting dilution into water-
100 in-oil droplets at ~10 kHz rates along with multiplex PCR primer sets (**Figure 1C**). Multiplex PCR
101 amplifies the seven CPS promoters tagged separately with a barcode sequence unique to each

102 droplet to generate a single-cell amplicon sequencing library. The primer-annealing sites flank the
103 inverted repeat sequences where promoter inversion occurs for each operon (17). Thus, each of
104 the seven pairs of primers report the ON/OFF status of their respective promoter through the
105 unique sequence of the amplicon generated (see Methods for detailed breakdown of the quality
106 control and analysis pipeline, and **Supplementary Table 1** for a summary of the number of cells
107 sequenced and passing quality controls for each library).

108 To determine the technical reproducibility of this method, we sequenced a single colony
109 grown for 48 hours on an agar plate in technical replicates (~5000 cells per replicate). We
110 detected 55 of 128 possible single-cell promoter states within this population, which represent
111 43% of all possible single-cell promoter states (**Figure 1D**). In a modified strain with the
112 recombinase (*mpi*) responsible for inversion deleted (mpi8) (17), the entire population displayed
113 the same single-cell promoter state (100% of ~9000 cells) (**Figure 1E**). Therefore, the deletion
114 of *mpi* locked the orientation of the promoters across the population. The variation between
115 technical replicates was low for most single-cell promoter states, but increased for states with low
116 relative abundance. Consistent with this result, the coefficient of variation for technical replicates
117 displayed a moderate negative correlation with the number of cells sequenced for each
118 subpopulation **(Supplementary Figure 1).** This implies that some fraction of the variance could
119 be attributed to random variation due to sampling (i.e. low abundance sub-populations may be
120 inconsistently observed).

121

122 *Patterened development of population composition in wild type B. fragilis colonies*

123 After establishing the robustness of the method, we investigated the variation in population-level
124 promoter compositions across multiple colonies (populations derived from single cells) (**Figure**
125 **1A**). To this end, we examined single-cell promoter orientations in five wild-type colonies picked
126 at random, representing five independent populations **(Figure 2A)**. Across all five colonies, we
127 observed 77 of 128 possible CPS combinations. Each colony displayed markedly different
128 population compositions with 0–4 invertible promoters oriented in the ON state (**Figure 2B**).

129 To understand the global properties of the population compositions across each colony,
130 we generated an undirected network that represents the relationships between single-cell
131 promoter states (**Figure 2C** and **Supplementary Figure 2**). Each single-cell promoter state is
132 represented by a node proportional in size to its relative abundance in the population. Nodes are
133 connected if their Hamming distance (HD) is equal to one (i.e. separated by a single promoter-
134 inversion). For example, the HD between combinatorial states A--E--- and A-DE--- is 1.

Aggregate analysis of the networks for all colonies revealed patterns in population composition. While subpopulations with many (5-7) promoters turned off were frequently observed, subpopulations with many (5-7) promoters simultaneously on were rare. This pattern suggests that *B. fragilis* limits concomitant promoter-ON orientations. In addition, all observed subpopulations (accounting for ~22-34% of possible single-cell states) are connected (1 HD away) to at least one other observed subpopulation. This is consistent with the possibitility that the rates of inversion are much slower than the rate of cell division. Further, larger (i.e., higher abundance) subpopulations had many (>6) immediate (HD = 1) neighbors **(Figure 2D)**. For comparison, most other lower abundance subpopulations were connected to ~3–5 immediate neighbors **(Supplementary Figure 3A)**. This pattern is consistent with the notion that high abundance populations arose earlier in population development and gave rise to their lower abundance neighbors.

Higher abundance subpopulations were closer (i.e., lower HD) to the most abundant subpopulation **(Figure 2E**, Spearman's Rho -0.65, $p < 2.2e-16$ between subpopulation size and HD). Finally, most of the cells were only one promoter inversion away from the most abundant promoter state (main promoter state) **(Figure 2B,E)**. These observed patterns in population-level promoter state network properties could arise from two possibilities. First, the population composition reflects a snapshot in early stages of a diversifying population, where the major subpopulations are more closely related to the single-cell promoter state of the initial founding cell. Alternatively, the observed promoter-state composition reflects the longer-term steady-state to which each colony converged to independently.

Populations with different initial promoter state compositions display converging trajectories over time

To determine whether these populations are in early stages of diversification or reflect multiple steady-state compositions, we individually cultured the five colonies for two weeks in liquid media (**Figure 2B**). The cultures were diluted every 8 or 16 hours alternately with different inoculum volumes to prevent the cultures from entering late stationary phase (**Methods**). To track the temporal diversification process, we performed DoTA-seq on the populations on day 3, 7, and 14 **(Figure 3A)**. The Shannon diversity of individual cultures increased over time as new and distinct single-cell promoter states arose, and the population became more evenly distributed among different combinatorial promoter states **(Figure 3B)**. By day 14, each population exhibited on average 89 +/- 9 unique single-cell states and displayed 117 unique single-cell states across all

colonies **(Supplemental Table 2)**. Although these populations had not yet reached steady-state (i.e. population-level promoter composition was still changing over time), the population compositions became increasingly similar to each other as time progressed (**Figure 3C**). These results suggest that populations from different initial starting states eventually converge to a single steady-state population composition. However, the time for convergence to a steady-state was on the scale of weeks, which suggests that weak forces shape the population composition towards a single preferred state under our culture conditions. In addition, the large number of single-cell promoter states detected in the populations on day 14 highlights the diverse repertoire of variation available to *B. fragilis* populations to cope with environmental stresses.

The population dynamics and the steady-state promoter distributions could be driven by regulation (e.g., promoter inversion rates), variation in subpopulation fitness, or a combination of these factors. We first considered the contribution of the recombinase-mediated promoter inversion rates to the population dynamics. In many bacterial species, the rates of phase-variation can change in response to environmental cues (4). In *B. fragilis* specifically, expression of the Mpi recombinase may be regulated by additional phase-variation mechanisms (20). To eliminate these potential additional layers of regulation, we constructed a strain enabling synthetic control of expression of the recombinase. To this end, we introduced a tightly regulated (i.e., low leaky expression), aTc-inducible *mpi* at an ectopic location on the genome in an *mpi* deletion background in *B. fragilis*. In the presence of high aTc concentrations, *mpi* can be expressed at high levels from this inducible promoter **(Fig. 3D)**.

We first used this system to generate and isolate phase-locked variants (**Supplementary Figure 5**) in different single-cell promoter states (A--E---, -B---G-, and --D---H) in the absence of aTc. In the presence of aTc, recombinase expression is turned on and the promoter states diversify across the population as a function of time. To investigate the diversification process over time, we cultured the phase-locked strains to mid-exponential-phase, induced recombinase expression, and then sampled these populations over time after induction (**Figure 3D**). The induced populations showed converging population composition trajectories (**Figure 3E**) and approached a single steady-state population composition 24 hours following induction of recombinase (**Figure 3F**). Thus, high expression of recombinase resulted in a fast convergence of the population to the steady-state composition, imposing a strong constraint on population composition.

To evaluate the contribution of fitness differences for each variant to the observed population dynamics, we determined the growth rates of the phase-locked variants under our

culturing conditions. We did not observe statistically significant differences in growth rates between different CPS locked strains when grown in our culturing medium ($p > 0.19$, one-factor ANOVA) (**Supplementary Figure 4)**. This suggests that variation in subpopulation fitness was not a strong contributor to the population dynamics in our experimental conditions**.**

Computational modeling reveals a major mechanism governing population dynamics

Computational modeling is a powerful way to gain mechanistic insight into complex biological systems (21). However, the development of computational models can be challenged by the complexity of biological systems combined with a lack of high quality data needed to sufficiently constrain model parameters. Single cell data can provide constrain stochastic model parameters better than bulk measurements (22). Leveraging our single-cell data, we constructed a stochastic mechanistic model to describe the population composition as a dynamic process in which promoters flip independently and stochastically at the single-cell level. This process is represented as a continuous-time Chemical Master Equation (CME) model consisting of 128 discrete states (**Figure 4A, Supplementary Materials**). The parameters of this model were estimated by fitting the analytical solution of the CME to the time-series single cell data (**Methods**).

We fit the model to the single-cell time series data to evaluate the contribution of promoter inversion rates to the observed population dynamics. This model fit the time series measurements of promoter states in the *mpi* synthetically regulated strain extremely well (Spearman's Rho ~0.98, $p = 2.57e\text{-}275$), suggesting that the assumptions of the model were consistent with the experimental observations (**Figure 4B-C**). In addition, all model parameters were well-constrained by the time series data except the promoter inversion rates D-ON and G-OFF. The posterior parameter distributions for these parameters displayed coefficients of variation greater than 5%, indicating that these parameters were moderately constrained **(Supplementary Table 3)**. To determine if the observed goodness of fit was due to the patterns observed in the experimental data, we randomly shuffled the dataset labels and re-fit the CME model. The model did not fit well to the label-shuffled datasets **(Supplementary Figure 7)**. In sum, the population diversification trajectories of the *mpi* synthetically regulated strain are driven primarily by independent and differential promoter inversion rates.

To investigate the longer-term behavior of the population beyond the experimentally measured time points, we determined the steady-state of the CME model (**Figure 4D**). The steady-state exhibits high diversity, with most single cells displaying 0-2 invertible promoters

234 simultaneously ON. In addition, the steady-state is dominated by promoter combinations with
235 promoter A in the ON state. Dominance of the promoter A in the ON state is reflected in the
236 inferred promoter inversion rates, where only promoter A has a higher rate of flipping ON than
237 OFF (**Fig. 4E)**. Promoters F, G, and H display higher inversion rates (both ON and OFF) than the
238 other promoters, leading to a faster convergence of these promoters to steady state than the other
239 promoters with smaller magnitude ON and OFF rates **(Supplementary Figure 8)**.

240 The observed differences in kinetic rates could reflect different roles that CPSs play in the
241 biology of the bacterium. For example, capsular polysaccharide A (PSA) has been shown to be
242 essential for colonization (23), pathogenesis (24), and was found on average to be oriented ON
243 in more than 50% of WT populations in every tested condition (17,25). This is consistent with the
244 inferred higher ON than OFF rate for promoter A. Capsular polysaccharide G (PSG) was
245 previously shown to be highly context dependent, only detected in high proportions in the context
246 of a microbial community *in vivo* (25). Here, promoter G had faster kinetic rates but was biased
247 towards the OFF state. Thus, the ON and OFF rates may influence sensitivity of *B. fragilis* CPS
248 compositions to environmental conditions.

249 To evaluate the contribution of promoter inversion rates to the population dynamics of
250 wild-type *B. fragilis,* we fit the CME model to time-series measurements of the 5 wild-type
251 populations **(Figure 3A)**. The model displayed a reasonable fit to these data (Spearman's Rho ~
252 0.8, p = 3.40e-132) **(Figure 4F)** and the fit was substantially better than the label-shuffled control
253 data (p = 4.1e-60, one sample t-test) **(Figure 4G).** This result implies that the independent rates
254 of promoter inversion could explain a portion of the dynamics in this experimental system. The
255 posterior distributions for certain parameters (F-OFF and G-OFF) exhibited bimodality and were
256 not well-constrained (>5% coefficient of variation in parameter distribution). This suggests that
257 additional factors beyond those captured in the model (such as regulation of recombinase levels
258 or activity) contribute to the population dynamics in wildtype populations, especially for promoters
259 F and G (**Supplementary Figure 6B, Supplementary Table 3**). Consistent with this notion,
260 promoters F and G harbor extended inverted repeat DNA elements surrounding the promoter
261 (26), potentially influencing recombinase activity or dependence on other regulatory factors.
262 Although Mpi is required for inversion, Mpi may interact with other regulatory proteins to modulate
263 inversion at different loci, providing an additional layer of regulation over specific CPSs (23)**.**
264 Further work is required to elucidate these additional layers of regulation of promoter inversion
265 rates, especially in physiologically relevant environmental contexts (20).

In sum, the model's excellent fit to time series measurements of the *mpi* synthetically regulated strain indicates that independent promoter inversion rates are a major mechanism driving the population-level CPS promoter state composition of *B. fragilis* populations. By contrast, the model's moderately good fit to the wildtype data suggests that additional layers of regulation are present, especially for promoters F and G.

Summary and outlook

Phase variation has been shown to be a wide-spread mechanism of generating population heterogeneity for diverse bacteria (3,4). However, we have a limited quantitative understanding of the diversification process, the landscape of single-cell states and the driving regulatory networks. By profiling the phase variable CPS loci in *B. fragilis* populations at the single-cell level over time, we uncovered the quantitative contribution of promoter inversion rates to the observed dynamic behaviors. The population dynamics are fundamentally shaped by different rates of promoter inversion, which influences the steady-state population composition and how quickly the population reaches the steady-state. The convergence of this system to a single-steady-state is in contrast to other diversification gene regulatory networks, which are strongly influenced by interlinked positive and negative feedback loops and exhibit bistable behavior (27). These results provide a foundation for further understanding a human gut symbiont in the context of their natural life cycles, such as through colonization of the human gut. This novel method of studying phase-variation, as well as our findings can be extended to other diverse bacteria, including pathogens, that use phase variation in their life cycle (2,28–35). Thus, our results set the stage for studying how population heterogeneity is used by bacteria to respond to environmental perturbations such antibiotic stress, microbial warfare, and phage and host immune cell attack.

METHODS

Plasmids and Strains

B. fragilis NCTC9343 was obtained from ATCC. Lyophilized culture was resuspended in SBM and frozen in 25% glycerol. All *B. fragilis* cultures were grown at 37°C in an anaerobic chamber (Coylabs) with an atmosphere of 2.5 ± 0.5% H2, 15 ± 1% CO2 and balance N2. We note that production of a stock following outgrowth single colony, rather than directly from lyophilized culture as done here, will generate a stock with significantly lower initial diversity.

To generate the inducible recombinase strains locked under different promoter orientations, we integrated chromosomally the *mpi* recombinase gene under the control of a tetracycline inducible

300 promoter in a *mpi* deletion strain. pNBU2_erm-TetR-P1T_DP-GH023 was a gift from Andrew
301 Goodman used for creating the tetracyline-inducible recombinase strain (Addgene plasmid #
302 90324 ; http://n2t.net/addgene:90324 ; RRID:Addgene_90324). Briefly, we cloned mpi (also
303 known as ssr2) into pNBU2_erm-TetR-P1T_DP-GH023 via Gibson assembly, transformed into
304 E. coli strain RL1752 via electroporation, and sequence-verified using whole-plasmid sequencing
305 (pJS041). pJS041 was next transformed via electroporation into the donor strain E. coli BW29427
306 and plated on LB agar containing 100 µg/mL ampicillin and 150 µM DAP. Overnight cultures were
307 created by inoculation of a single colony of donor (pJS041/E. coli BW29427) or recipient (Δmpi
308 M44) colonies into LB containing 100 µg/mL ampicillin and 150 µM DAP or Supplemented Basal
309 Medium (SBM), respectively. In the morning, 10 µL of overnight donor culture was used to
310 inoculate 5 mL LB containing the same supplements, while 250 µL of overnight recipient culture
311 was used to inoculate 20 mL of SBM (2% proteose peptone, 0.5% yeast extract, 0.5% NaCl, 1.5%
312 agar, supplementing 0.5% glucose, 0.5% K2HPO4, 0.05% cysteine, 5 µg/mL hemin, and 0.25
313 µL/100 mL vitamin K1 after autoclaving). The donor culture was grown to mid-late log phase while
314 recipient culture was grown mid-log phase, then the donor culture was pelleted by centrifugation
315 at 4000xg for 5 minutes, washed by resuspension in 5 mL LB and centrifugation. The donor pellet
316 was then resuspended with the 20 mL recipient culture, pelleted at 3000g for 10 minutes, then
317 resuspended in 1 mL standard SOC medium containing 5 µg/mL hemin and 5.5 µM vitamin K1
318 (SOC+HK), spotted on BHIS (37g/L with 5 µg/mL hemin and 0.25 µL/100 mL vitamin K1) agar
319 containing 150 µM DAP, and incubated upright, overnight, and aerobically at 37°C. The next day,
320 matings were scraped from plates and resuspended in 2 mL of SOC+HK. Dilutions were plated
321 on BHIS agar containing 200 µg/mL gentamicin and 5 µg/mL erythromycin. Colonies were
322 screened for their ability to express *mpi* under induction of 100 ng/µL anhydrotetracycline by SDS-
323 PAGE.

324

325 **Plating and harvesting of B. fragilis colonies**

326 Cells were streaked out directly from frozen glycerol stock at the lab bench on to SBM agar plates
327 and the plates were incubated for two days (>36 hours) before colonies were harvested using a
328 sterile pipette tip and resuspended in PBS + 0.1% Tween 20 by pipetting up and down.

329

330 **Long-term culturing of *B. fragilis* populations**

331 Five colonies were resuspended in 30 µL PBS + 0.001% Tween20. 10 µL of each suspension
332 was used to inoculate separate 1 mL overnight cultures. The remaining suspension was used for
333 microfluidic analysis of colonies. In the morning, timepoints were taken from each of the five
334 cultures: each culture was briefly vortexed, then two 100 µL aliquots were mixed with equal
335 volumes of 50% glycerol (technical replicates) and flash-frozen in a dry ice-ethanol bath. Each
336 day for 2 weeks, cultures were passaged twice: in the morning (16 hours after inoculation), 10 µL
337 of the overnight culture was used to inoculate 1 mL of SBM (1:100); in the evening (8-9 hours
338 later), 1 µL of saturated culture was used to inoculate 1 mL of SBM (1:1000).

339

340 **Growth curves of wildtype and CPS promoter locked strains**

341 Overnight cultures of each strain are inoculated from frozen glycerol stocks into SBM media. The
342 next day, the OD600 of each overnight culture is measured on the Nanodrop One (Thermofisher)
343 using the cuvette setting, and a 96-well flat bottom assay plate (Midsci 781602) is prepared with
344 200 µL of SBM media and each culture is added to two wells to an initial OD600 of 0.01. The plate
345 is incubated in a Tecan F-200 plate reader at 37°C in the anaerobic chamber, taking an OD600
346 reading every 30 minutes for 48 hours.

347

348 **Fitting of growth curves to data**

349 The raw OD600 readings are truncated to remove the death phase and the values normalized to
350 between 0 and 1 for each individual curve. This is to account for potential differences in OD600
351 per cell for each CPS variant. A logistics growth ODE model (d[X]/dt = (r + alpha-X)*X) where X
352 is the normalized abundance, r is the growth rate, and alpha is the self-inhibitory parameter) is fit
353 to this data using least squares fit function in python. The growth rate is taken as the exponential
354 growth parameter in the logistics model. Scripts used in this analysis can be found on github (See
355 code availability).

356

357 **Fabrication of microfluidic droplet maker**

358 Microfluidic masters were fabricated in a clean room using soft lithography (36). SU-8 3000
359 photoresist (MicroChem) was spun on a silicon wafer (University Wafers) to achieve a thickness
360 of 30 µm. Photolithography masks **(supplementary materials)** were ordered from CAD/ART and
361 used to pattern the photoresist using a UV Led (ThorLabs) for 1 min 45s. The patterned
362 photoresist was baked for 5 minutes at 95oC then developed using SU-8 developer (PGMEA) for
363 2 minutes then baked at 200oC for 2 minutes.

364 PDMS devices were created by curing PDMS elastomer (Sylgard-184) at a 1:11 crosslinker to
365 elastomer ratio over the silicon master. These devices were cut out using a scalpel and holes
366 punched using a 0.75mm reusable biopsy punch (World Precision Instruments). PDMS devices
367 were bonded to glass slides using a plasma cleaner (Harris Plasma), and then treated with
368 Aquapel (Aquapel Glass Treatment) to render them hydrophobic.

369

370 **Single cell sequencing of *B. fragilis* populations**

371 *B. fragilis* cells were pelleted by centrifugation, washed with 1 mL of PBS + 0.1% Tween 20, then
372 resuspended in 100-500 µL of PBS + 0.1% Tween 20. OD600 readings were taken of the
373 resuspension using the pedestal mode of the Nanodrop One (Thermo Scientific), which is used
374 as proxy for cell concentrations for input into the DoTA-seq workflow.

375 A PCR reaction mix is prepared as follows: 25 µL Q5 Ultra II PCR mastermix, 1µL of primers
376 mixture consisting of P7(10 µM)-Barrev(1 µM), 0.7µL of Inverton targeting primers mixture (20 µM
377 total), 0.5 µL of Barcode Oligos (1 pM) which has been freshly diluted from 500 pM with TE buffer.
378 See supplementary table 4 for a list of oligonucleotides used in this paper.

379 This PCR mixture is combined with 25µL of cells diluted with pre-injection buffer **(**10mM HEPES
380 pH 7.5, NaCl 25 mM, EDTA 0.1 mM, Tween-20 2% v/v) such that the final OD of cells in the final

mixture is 0.0025. This mixture is then injected into the droplet making microfluidics device using syringe pumps (New Era) and 1mL syringes (BD) at a flowrate of 400 µL/hr and 800 µL/hr of Biorad Evagreen droplet making oil (Biorad) to generate droplets.

Droplets are collected into a PCR tube (Fisher Scientific #14222292) and thermo-cycled in a Biorad CX-100 thermocycler as follows: 98oC for 2 minutes, and 40 cycles of 98oC and 30 seconds at 65oC for 5 minutes, followed by 72oC for 10 minutes and hold at 12oC.

After PCR, the coalesced droplets were removed using a pipette, and the emulsion was broken on ice by adding 20 µL 500mM EDTA and 20 µL perfluoro-octanol (Sigma 370533), then vortexed followed by a spin pulse centrifugation. The aqueous phase was transferred to another tube by pipette and cleaned up using a Zymo cleanup and concentrator kit (Zymo Research D4013), then subject to size selection using SPRI-select beads (Beckman Coulter B23317) using 0.7x volume of beads. A further round of size selection was performed in 100 mM NaOH, 10% Ethanol, and 1.4x volume of beads to increase the purity of the library.

All libraries were sequenced on the Miseq using V3 150cycle kits with custom sequencing primers **(supplementary table 4)**, using 155 cycles for read 1, 18 cycles for the I7 index, and 8 cycles for the I5 index.

Analysis of sequencing data

The raw sequencing reads are obtained from the Miseq. Read 1 represents the targeted amplicons. The first index read represents the unique cell barcode. The second index read is used to multiplex libraries from different experiments. Demultiplexing of different libraries (index 2) is performed by the Miseq software. Cell barcode demultiplexing and quality control is performed using a custom python script (R4-parser.py). The filtered reads were mapped to custom built reference databases containing B. fragilis CPS promoter sequences and is available on GitHub (see code availability) using Bowtie2 V2.3.4.1(37) using "--very-sensitive" presets. The mapped reads were analyzed using custom scripts as follows: The mapped reads are organized into read groups consisting of reads with the same unique cell barcode representing amplicons from the same droplet. Read groups with too few reads are removed. The filtered groups are transformed into a table containing the barcode and a tally of the number of reads that are mapped to each target (SAM-analysis.py). We further filtered the reads by removing the read groups if the reads mapping to any promoter (sum of ON and OFF orientation) are less than 1% of total reads for that group. Subsequently, if reads in any promoter orientation were less than 1% of total reads for the barcode, we set the value to 0; this accounted for expected noise. We next discretized the data by replacing values with 1 or 0, corresponding to ON or OFF, respectively. We removed barcodes if neither orientation was found for any promoter (i.e., if ON + OFF values = 0). We also removed read groups if both orientations were found for any locus (ON + OFF = 2). The set of remaining read groups are used to determine the frequencies of each promoter state combination in the population **(CPS-analysis.R).** This processed list of frequencies for each promoter state is available in Supplementary Table 2. All scripts are available on GitHub (see code availability).

Computational modelling of promoter inversion dynamics

For a single cell, the promoter flipping process is represented mathematically as a continuous-time Markov process consisting of 128 discrete states. In an infinitesimal time interval, the propensity for a cell to switch from state *x* to state *y* with a single promoter change is a product of (1) the flipping rate constant of that promoter and (2) the probability that the cell is currently in state *x*. We assume that this stochastic flipping process is ergodic: the frequency of a subpopulation at time *t* represents the probability of an individual cell to be in this state. This assumption allows us to fit the analytical solution of this model to experimental subpopulation frequencies to infer the flipping rate constants. Parameter inference is performed using a Markov Chain Monte Carlo approach, which accounts for uncertainties in measurements of subpopulation frequencies (ie. subpopulations close to stochastic limit of detection). For a detailed description of the modelling workflow, please see **supplementary materials BF-modelling.pdf**. All scripts are available on GitHub.

Code availability

All scripts and code used in analysis are available on Github at **https://github.com/lanfreem/Bfragilis-CPS.git**.

Data availability

Raw sequencing data and OD600 plate-reader data are available on Zenodo at the time of publication.

Acknowledgements

We would like to thank Professor Laurie Comstock at the University of Chicago for supplying published strains and offering feedback on the preliminary data. J.S. would like to thank Dr. Mike Wolfe for discussion and guidance for data processing. J.S. was supported in part by the NIH Biotechnology Training Grant (T32 GM135066 and T32 GM008349) and NIH F31 Graduate Fellowship (F31GM142153). J.S. is supported additionally by the SciMed Graduate Research Scholars Fellowship – support for this fellowship is provided by the Graduate School, part of the Office of Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison, with funding from the Wisconsin Alumni Research Foundation and the UW-Madison. F.L. was supported in part by the Burroughs Wellcome Fund Careers Award in the Scientific Interface. This work was supported by the National Institutes of Health R35 GM124774, R35 GM12477409 and R21 AI156438, and Army Research Office W911NF-19-1-0269.

Author Contributions

O.S.V, R.L., F.L. and J.S. conceived of the study. F.L. and J.S. conceived and performed experiments and conceived and wrote custom scripts for data analysis and processing. F.L. conceived of and executed the microfluidics technology protocol, and generated sequencing data and sequencing analysis. J.S. conceived of and executed the method for generating and isolating phase-locked variants. F.L., J.S., and T.R. performed microfluidics experiments. Y.Q. conceived and executed computational modeling, wrote custom scripts related to these models and

performed analyses of these data. F.L. conceived of and executed other computational models. O.S.V., R.L., F.L., J.S. and Y.Q. analyzed the data. O.S.V., F.L., and J.S. wrote the manuscript. O.S.V., R.L., F.L., J.S. and Y.Q. contributed to the design of Figures. O.S.V. and R.L. supervised the study and secured funding.

References

1. Stapels DAC, Hill PWS, Westermann AJ, Fisher RA, Thurston TL, Saliba AE, et al. Salmonella persisters undermine host immune defenses during antibiotic treatment. Science. 2018 Dec 7;362(6419):1156–60.

2. Gunther IV NW, Snyder JA, Lockatell V, Blomfield I, Johnson DE, Mobley HLT. Assessment of Virulence of Uropathogenic *Escherichia coli* Type 1 Fimbrial Mutants in Which the Invertible Element Is Phase-Locked On or Off. Infect Immun. 2002 Jul;70(7):3344–54.

3. van der Woude MW, Bäumler AJ. Phase and Antigenic Variation in Bacteria. Clin Microbiol Rev. 2004 Jul;17(3):581–611.

4. Bayliss CD. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. FEMS Microbiol Rev. 2009 May;33(3):504–20.

5. Cota I, Blanc-Potard AB, Casadesús J. STM2209-STM2208 (opvAB): A Phase Variation Locus of Salmonella enterica Involved in Control of O-Antigen Chain Length. Chakravortty D, editor. PLoS ONE. 2012 May 11;7(5):e36863.

6. Foley J. Mini-review: Strategies for Variation and Evolution of Bacterial Antigens. Comput Struct Biotechnol J. 2015 Jan 1;13:407–16.

7. Jiang X, Brantley Hall A, Arthur TD, Plichta DR, Covington CT, Poyet M, et al. Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. Science. 2019 Jan;363(6423):181–7.

8. Seib KL, Jen FEC, Scott AL, Tan A, Jennings MP. Phase variation of DNA methyltransferases and the regulation of virulence and immune evasion in the pathogenic Neisseria. Pathog Dis. 2017 Aug 31;75(6):ftx080.

9. Phillips ZN, Tram G, Seib KL, Atack JM. Phase-variable bacterial loci: how bacteria gamble to maximise fitness in changing environments. Biochem Soc Trans. 2019 Aug 30;47(4):1131–41.

10. Béchon N, Mihajlovic J, Vendrell-Fernández S, Chain F, Langella P, Beloin C, et al. Capsular Polysaccharide Cross-Regulation Modulates Bacteroides thetaiotaomicron Biofilm Formation. mBio. 2020 Jun 23;11(3):e00729-20.

11. Porter NT, Canales P, Peterson DA, Martens EC. A Subset of Polysaccharide Capsules in the Human Symbiont Bacteroides thetaiotaomicron Promote Increased Competitive Fitness in the Mouse Gut. Cell Host Microbe. 2017 Oct 11;22(4):494-506.e8.

12. Porter NT, Hryckowian AJ, Merrill BD, Fuentes JJ, Gardner JO, Glowacki RWP, et al. Phase-variable capsular polysaccharides and lipoproteins modify bacteriophage susceptibility in Bacteroides thetaiotaomicron. Nat Microbiol. 2020 Sep;5(9):1170–81.

13. Saunders NJ, Moxon ER, Gravenor MB. Mutation rates: estimating phase variation rates when fitness differences are present and their impact on population structure. Microbiology. 2003 Feb 1;149(2):485–95.

14. Criss AK, Kline KA, Seifert HS. The frequency and rate of pilin antigenic variation in Neisseria gonorrhoeae. Mol Microbiol. 2005 Oct;58(2):510–9.

15. Hung M, Chang E, Hussein R, Frazier K, Shin JE, Sagawa S, et al. Modulating the frequency and bias of stochastic switching to control phenotypic variation. Nat Commun. 2014 Aug 4;5(1):4574.

16. Krinos CM, Coyne MJ, Weinacht KG, Tzianabos AO, Kasper DL, Comstock LE. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. Nature. 2001 Nov;414(6863):555–8.

17. Coyne MJ, Weinacht KG, Krinos CM, Comstock LE. Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. Proc Natl Acad Sci. 2003 Sep 2;100(18):10446–51.

18. Johnson RC. Site-specific DNA Inversion by Serine Recombinases. Microbiol Spectr. 2015 Feb 19;3(3):1–36.

19. Lan F, Saba J, Ross TD, Zhou Z, Krauska K, Anantharaman K, et al. Massively parallel single-cell sequencing of genetic loci in diverse microbial populations [Internet]. bioRxiv; 2022 [cited 2022 Nov 28]. p. 2022.11.21.517444. Available from: https://www.biorxiv.org/content/10.1101/2022.11.21.517444v1

20. Kuwahara T, Yamashita A, Hirakawa H, Nakayama H, Toh H, Okada N, et al. Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. Proc Natl Acad Sci. 2004 Oct 12;101(41):14919–24.

21. Qian Y, Lan F, Venturelli OS. Towards a deeper understanding of microbial communities: integrating experimental data with dynamic models. Curr Opin Microbiol. 2021 Aug 1;62:84–92.

22. Munsky B, Fox Z, Neuert G. Integrating Single-Molecule Experiments and Discrete Stochastic Models to Understand Heterogeneous Gene Transcription Dynamics. Methods San Diego Calif. 2015 Sep 1;85:12–21.

23. Liu CH, Lee SM, VanLare JM, Kasper DL, Mazmanian SK. Regulation of surface architecture by symbiotic bacteria mediates host colonization. Proc Natl Acad Sci. 2008 Mar 11;105(10):3951–6.

24. Coyne MJ, Tzianabos AO, Mallory BC, Carey VJ, Kasper DL, Comstock LE. Polysaccharide Biosynthesis Locus Required for Virulence of Bacteroides fragilis. Infect Immun. 2001 Jul;69(7):4342–50.

539 25. Troy EB, Carey VJ, Kasper DL, Comstock LE. Orientations of the Bacteroides fragilis
540 Capsular Polysaccharide Biosynthesis Locus Promoters during Symbiosis and Infection. J
541 Bacteriol. 2010 Nov;192(21):5832–6.

542 26. Patrick S, Parkhill J, McCoy LJ, Lennard N, Larkin MJ, Collins M, et al. Multiple inverted
543 DNA repeats of Bacteroides fragilis that control polysaccharide antigenic variation are
544 similar to the hin region inverted repeats of Salmonella typhimurium. Microbiology.
545 149(4):915–24.

546 27. Venturelli OS, Zuleta I, Murray RM, El-Samad H. Population Diversification in a Yeast
547 Metabolic Program Promotes Anticipation of Environmental Shifts. PLOS Biol. 2015 Jan
548 27;13(1):e1002042.

549 28. Cowley SC, Myltseva SV, Nano FE. Phase variation in Francisella tularensis affecting
550 intracellular growth, lipopolysaccharide antigenicity and nitric oxide production. Mol
551 Microbiol. 1996 May;20(4):867–74.

552 29. Snellings NJ, Tall BD, Venkatesan MM. Characterization of Shigella type 1 fimbriae:
553 expression, FimA sequence, and phase variation. Infect Immun. 1997 Jun;65(6):2462–7.

554 30. Lavitola A, Bucci C, Salvatore P, Maresca G, Bruni CB, Alifano P. Intracistronic transcription
555 termination in polysialyltransferase gene (siaD) affects phase variation in Neisseria
556 meningitidis. Mol Microbiol. 1999 Jul;33(1):119–27.

557 31. Snyder LA, Loman NJ, Linton JD, Langdon RR, Weinstock GM, Wren BW, et al. Simple
558 sequence repeats in Helicobacter canadensis and their role in phase variable expression
559 and C-terminal sequence switching. BMC Genomics. 2010;11(1):67.

560 32. Chopra-Dewasthaly R, Baumgartner M, Gamper E, Innerebner C, Zimmermann M,
561 Schilcher F, et al. Role of Vpma phase variation in *Mycoplasma agalactiae* pathogenesis.
562 FEMS Immunol Med Microbiol. 2012 Dec;66(3):307–22.

563 33. Beare PA, Jeffrey BM, Long CM, Martens CM, Heinzen RA. Genetic mechanisms of
564 Coxiella burnetii lipopolysaccharide phase variation. Monack DM, editor. PLOS Pathog.
565 2018 Feb 26;14(3):e1006922.

566 34. Safi H, Gopal P, Lingaraju S, Ma S, Levine C, Dartois V, et al. Phase variation in
567 *Mycobacterium tuberculosis glpK* produces transiently heritable drug tolerance. Proc Natl
568 Acad Sci. 2019 Sep 24;116(39):19665–74.

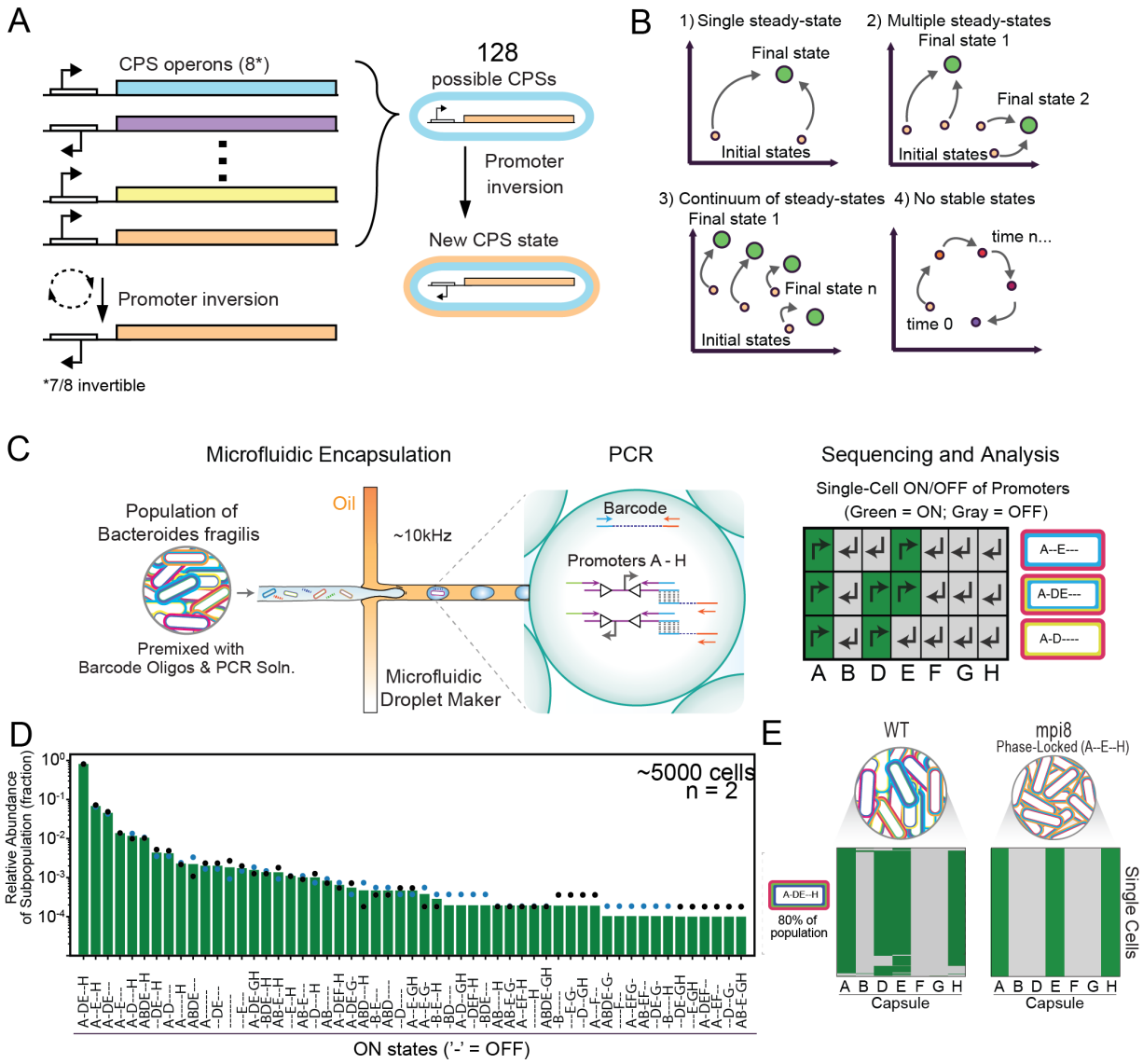
569 35. Zhang AN, Li L, Yin X, Dai CL, Groussin M, Poyet M, et al. Choosing Your Battles: Which
570 Resistance Genes Warrant Global Action? bioRxiv. 2019 Oct;(October):784322.

571 36. Duffy DC, McDonald JC, Schueller OJA, Whitesides GM. Rapid Prototyping of Microfluidic
572 Systems in Poly(dimethylsiloxane). Anal Chem. 1998 Dec;70(23):4974–84.

573 37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012
574 Apr;9(4):357–9.

575

576

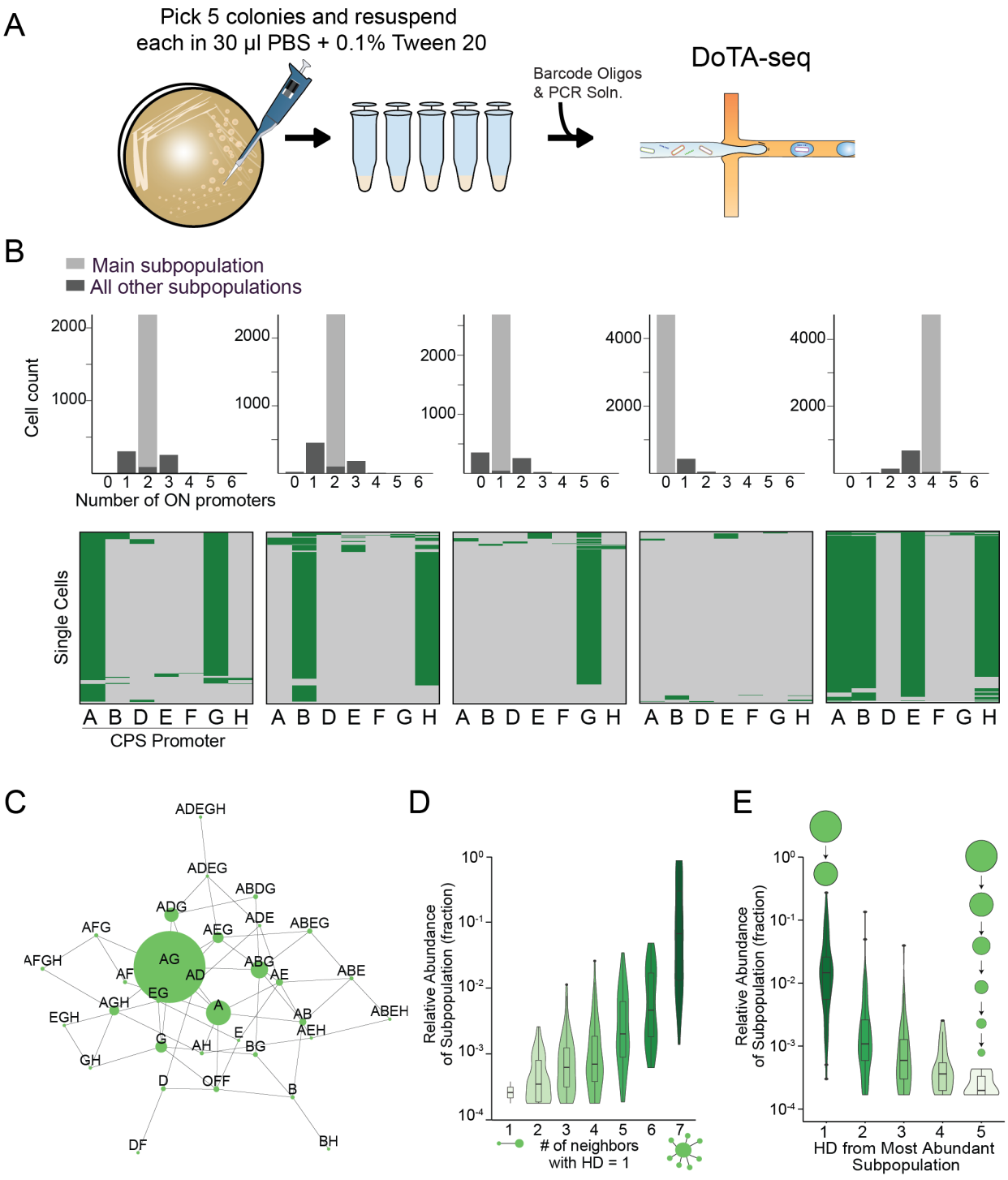

A
CPS operons (8*)
128
possible CPSs
Promoter
inversion
New CPS state
Promoter inversion
*7/8 invertible
B
1) Single steady-state
Final state
Initial states
2) Multiple steady-states
Final state 1
Final state 2
Initial states
3) Continuum of steady-states
Final state 1
Final state n
Initial states
4) No stable states
time n...
time 0
C
Microfluidic Encapsulation
PCR
Sequencing and Analysis
Population of
Bacteroides fragilis
Premixed with
Barcode Oligos & PCR Soln.
Oil
~10kHz
Microfluidic
Droplet Maker
Barcode
Promoters A - H
Single-Cell ON/OFF of Promoters
(Green = ON; Gray = OFF)
A--E---
A-DE---
A-D----
A B D E F G H
D
Relative Abundance
of Subpopulation (fraction)
~5000 cells
n = 2
ON states ('-' = OFF)
E
WT
mpi8
Phase-Locked (A--E--H)
A-DE--H
80% of
population
Single Cells
A B D E F G H
Capsule
A B D E F G H
Capsule

577

578

579 **Figure 1. High-throughput single cell sequencing of capsular polysaccharide promoters**
580 **(CPS) in a key human gut symbiont *Bacteroides fragilis* (*B. fragilis*) reveals bacterial**
581 **subpopulations defined by multiple phase variable loci. (A)** Summary schematic of *B. fragilis*
582 CPS loci. Promoter inversions generates diversity in CPS expression. **(B)** Cartoon schematic of
583 the promoter state landscape of the CPS system under continuous growth. Populations could 1)
584 converge to a single stable state, 2) converge to multiple stable states, 3) exist in a continuum of
585 stable states, or 4) remain in flux with no stable states. **(C)** Cartoon representation of the ultrahigh-
586 throughput single-cell targeted sequencing workflow for *B. fragilis* CPS promoters. We first premix
587 cells from a colony or liquid culture with barcode oligos, high-fidelity PCR mix, and target-loci
588 primers (14 primers for the 7 invertible CPS promoters) containing adaptor sequences that
589 overlap the barcode oligo. Cells in this mix are then encapsulated into picolitre emulsion droplets
590 at limiting dilution so that most droplets contain 1 or 0 cell according to a Poisson distribution. The
591 barcode oligos also exist at limiting dilution so that droplets containing cells also contain 1 or 0
592 barcodes. Droplets that contain both a barcode oligo and a cell are selectively amplified and linked
593 to genomic loci via overlap extension PCR. The cell is lysed during the initial PCR denaturing
594 step. Only amplicons linking the barcode and loci can be sequenced. After DNA library
595 preparation, sequencing reveals the promoter orientations of all targeted loci for thousands of
596 cells. Right: promoter orientation heatmap for 3 representative cells. Coverage in the 'on'
597 orientation for a particular CPS promoter is green. Dashes ('-') represent promoters that are
598 oriented off. **(D)** The CPS promoter states of a single colony of *B. fragilis*, shown as a bar plot.
599 Data points represent two technical replicates. **(E)** Heatmaps as described in (A) for thousands
600 of cells, with groups of rows clustered to reflect subpopulations with alternative combinatorial
601 promoter orientations. A control strain with the primary recombinase gene (*mpi*) deleted displays
602 a population with 100% of cells having the expected promoter configuration (A--E--H) for this
603 particular isolate.

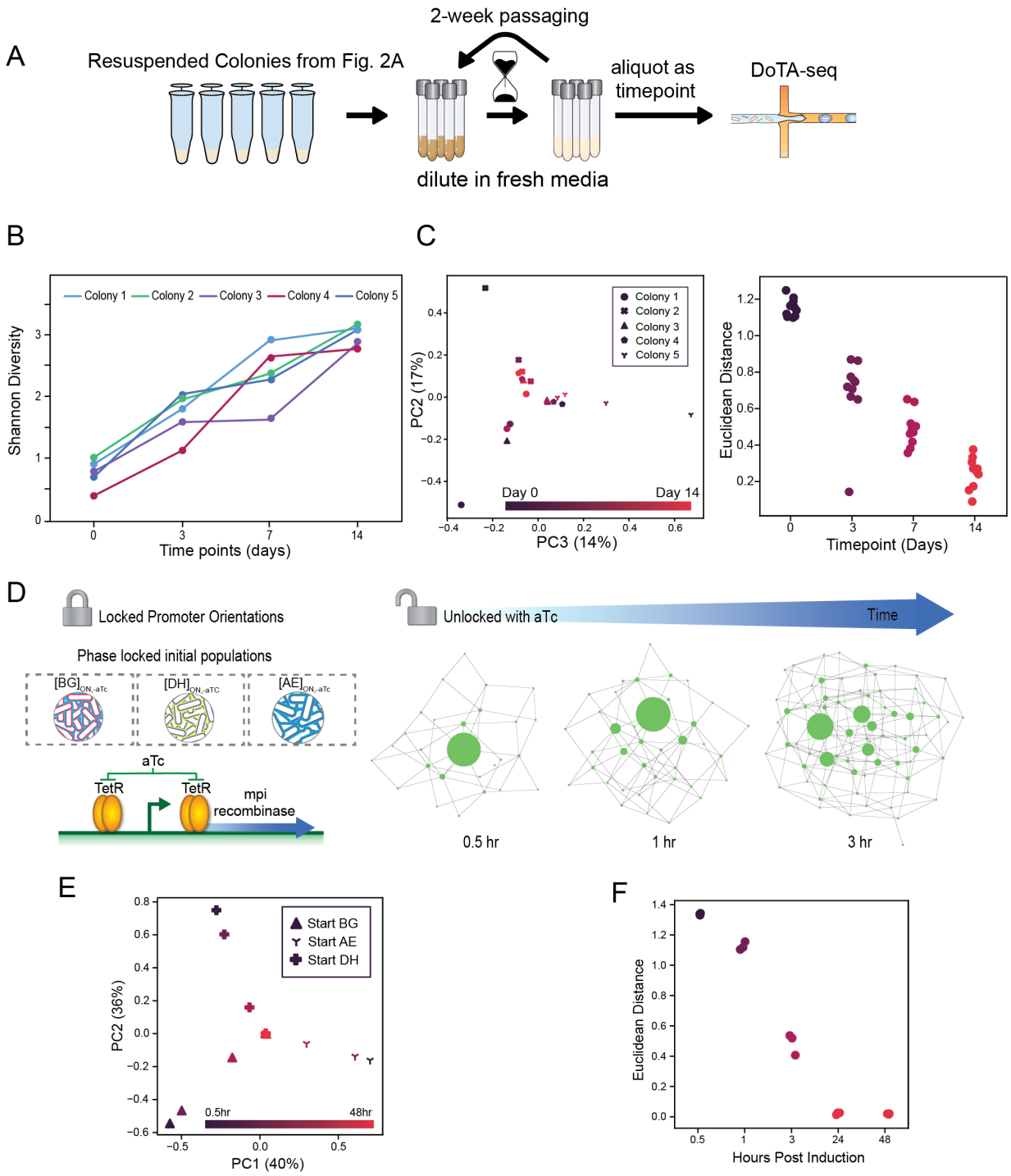
604


A
Pick 5 colonies and resuspend
each in 30 µl PBS + 0.1% Tween 20
Barcode Oligos
& PCR Soln.
DoTA-seq
B
Main subpopulation
All other subpopulations
Cell count
Number of ON promoters
Single Cells
A B D E F G H
CPS Promoter
C
ADEGH
ADEG
ABDG
ADG
ADE
ABEG
AFG
AEG
AFGH
AG
ABG
AF
AD
AE
ABE
EG
AGH
A
AB
ABEH
EGH
E
AEH
G
AH
BG
GH
D
OFF
B
DF
BH
D
Relative Abundance
of Subpopulation (fraction)
# of neighbors
with HD = 1
E
HD from Most Abundant
Subpopulation

605

606

Figure 2. High-throughput single cell sequencing of promoter states in *B. fragilis* colonies reveals diverse compositions with similar features. **(A)** Cartoon representation of experiment. A single glycerol stock was streaked onto an agar plate, and 5 colonies were picked for sequencing after 48 hours of growth. **(B)** Heatmaps of promoter states for 5 wild-type colonies. Top: Histogram of cells having *n* promoters simultaneously oriented on. The light gray bar represents the size of the most abundant state (i.e. main population). **(C)** Representative network representation of single-cell promoter states of colony 1. By connecting subpopulations with a Hamming Distance (HD) of 1 (i.e., separated by a single promoter inversion event), we can generate a undirected network representation of the subpopulations. Each node represents one subpopulation. Node size represents relative abundance. For network graphs of all colonies, see Supplementary Figure 2. **(D)** Violin plot of the distribution of subpopulations with different numbers of neighbors (where HD = 1). All colonies show the characteristic main subpopulation surrounded by progressively lower abundance subpopulations. **(E)** Violin plot of the distribution of subpopulations versus the distance from the most abundant subpopulation showing that most of the large populations are found close (lower HD) to the main subpopulation.


A
Resuspended Colonies from Fig. 2A
2-week passaging
dilute in fresh media
aliquot as timepoint
DoTA-seq
B
Colony 1
Colony 2
Colony 3
Colony 4
Colony 5
Shannon Diversity
Time points (days)
C
PC2 (17%)
PC3 (14%)
Day 0
Day 14
Euclidean Distance
Timepoint (Days)
D
Locked Promoter Orientations
Unlocked with aTc
Time
Phase locked initial populations
aTc
TetR
mpi recombinase
0.5 hr
1 hr
3 hr
E
Start BG
Start AE
Start DH
PC2 (36%)
PC1 (40%)
0.5hr
48hr
F
Euclidean Distance
Hours Post Induction

623

624

Figure 3. Population-level promoter state composition converges to a single steady-state composition over time. (A) Cartoon schematic of the experiment. The 5 colonies characterized in Figure 2 were grown in liquid media for 14 days and diluted periodically to prevent culture from reaching late stationary phase. Samples are taken and promoter states were characterized using high-throughput single-cell sequencing at multiple timepoints. **(B)** Shannon diversities of the 5 wildtype populations originating from the 5 colonies over each timepoint. Shannon diversity increases with time for all colonies. Day 0 represents the populations as colonies on a plate. **(C)** The populations show a converging trajectory over time to a single steady-state. Right: Population-level promoter state compositions are visualized bty principal component analysis (PCA). Marker shapes represent different populations. Marker colors represent different timepoints. Left: The pairwise Euclidean distance between all 5 populations at each timepoint. The trend of decreasing Euclidean distance shows a converging trajectory of the populations to a single steady-state. **(D)** Cartoon representation of the induced promoter inversion experiment. The native recombinase gene (*mpi*) responsible for promoter inversions was driven by an aTc-inducible promoter. Inversion is induced in phase-locked variants during log phase by the addition of 100 ng/μL aTc into the cultures. **(E)** aTc-induced cultures also show a converging trajectory based on PCA. Marker shapes represent different cultures. Marker colors represent different timepoints. **(F)** Pairwise Euclidean distances between the 3 aTc-induced populations at different timepoints post induction. By 24 hours, the 3 populations converged to a steady-state.

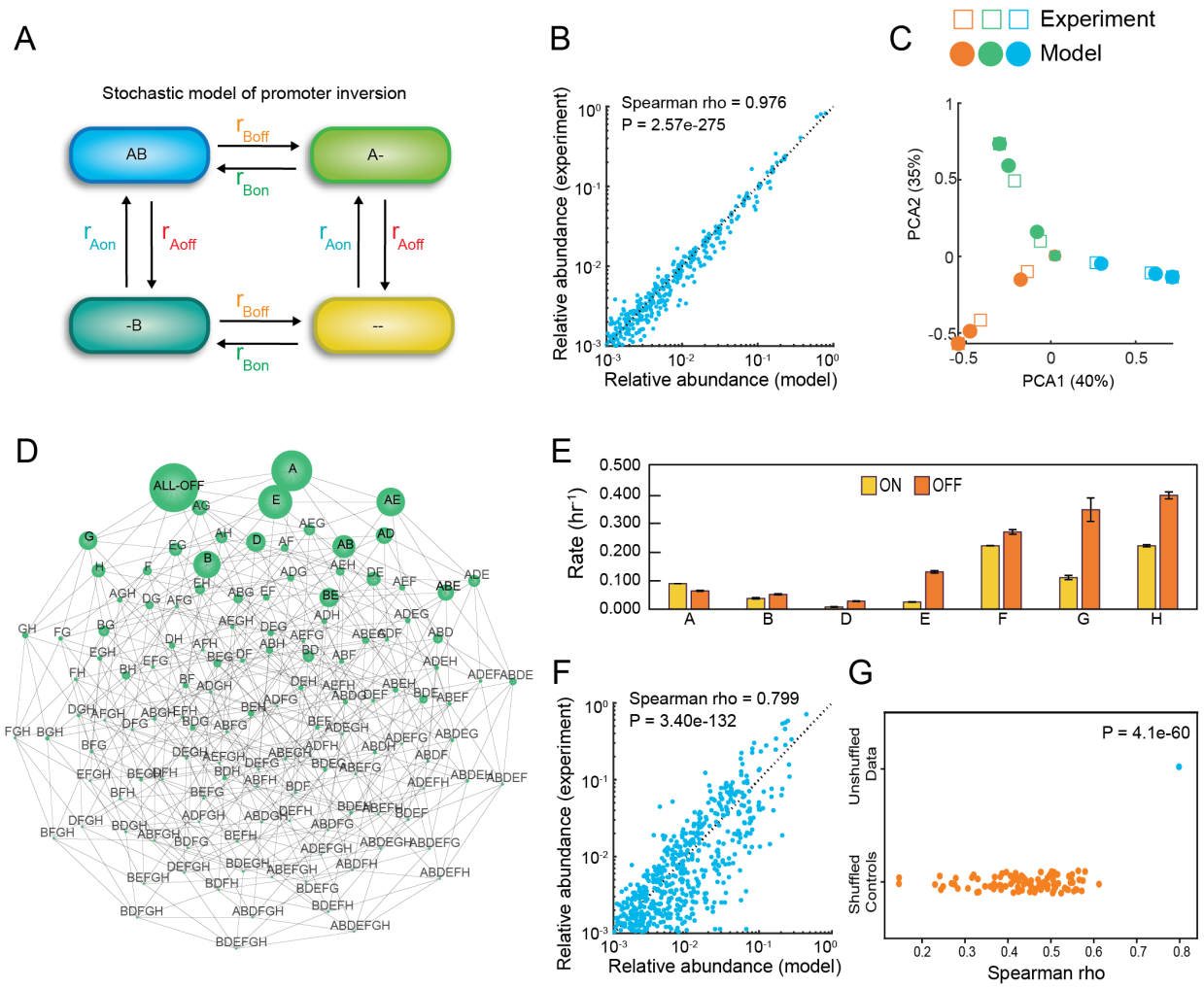


Figure 4. A stochastic dynamic model of promoter inversions demonstrates that promoter inversion rates are a major variable shaping the temporal changes in promoter state diversification. (A) Schematic (only two promoters) of the stochastic dynamic computational model for capturing temporal changes in single-cell promoter states. In the simplified representation, the cell transitions between four different states by inverting promoters on and off with different kinetic rate parameters. A and B represent different operon promoters in the ON state and "-" represents the promoter in the OFF state. In the model, there are 7 promoters, 14 kinetic parameters, and 128 different states. **(B)** Scatter plot of promoter state relative abundance of the 128 possible promoter combination states predicted by model versus experimentally measured for each time point of the aTc-inducible experiment. Spearman rho represents the Spearman correlation coefficient. **(C)** Model predictions and experimental data visualized using Principal Component Analysis (PCA). **(D)** Undirected network graph of the model's prediction of the promoter state composition. Subpopulations are connected by an edge if their hamming distance is equal to one. Node size represents relative relative abundance of the subpopulation. **(E)** Inferred kinetic rate parameters for promoter inversions derived from the model fitted to the aTc-inducible time-series data. Error bars represent 1 s.d. of the Monte-Carlo Marko chain parameter distribution. Posterior parameter distributions are shown in Supplementary Figure 7. **(F)** Scatter plot of promoter state relative abundance of the 128 possible promoter combination

states predicted by model versus experimentally measured for each time point for the 14-day wild-type experiment. **(G)** Categorial scatter plot of the Spearman's correlation of a model fit to the label shuffled controls versus wild-type time-series data.