Single-cell bacterial sequencing of invertible loci reveals promoter inversion rates as a strong determinant of population heterogeneity

Freeman Lan[1+], Jason Saba[1,2,3+], Yili Qian[1], Tyler D. 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

ABSTRACT

Single-cell heterogeneity can promote population-level fitness in response to unpredictable environmental conditions. In the human gut symbiont *Bacteroides* sp., the combinatorial phase variation of capsular polysaccharides (CPS) by promoter inversions generates diverse phenotypes across a population. By combining single-cell sequencing and stochastic modeling, we provide insights into fundamental questions about the regulation of heterogeneity in this health-relevant system. We show that *B. fragilis* populations can access diverse CPS promoter states and that promoter inversion rates are a major mechanism driving population composition. Exploiting control over the diversification mechanism, we show that populations with different initial compositions converge to a similar steady-state population-level compositions over time. More broadly, this approach could elucidate the population dynamics underlying stochastic genomic variation in a wide range of biological systems.

MAIN TEXT:

Gene regulation in bacteria is shaped by stochastic and deterministic processes, which enable populations to adapt to changing environmental conditions. 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 (Stapels *et al.*, 2018). The composition (fraction of the population occupying each state) of these phenotypically heterogeneous populations is influenced by the regulatory mechanisms encoded by the bacteria. Understanding how bacteria regulate the phenotypic heterogeneity at the population-level can reveal important insights into their mechanisms for adaptation.

Phase variation is a ubiquitous mechanism for generating phenotypic heterogeneity in bacterial populations (Gunther IV *et al.*, 2002; van der Woude and Bäumler, 2004; Bayliss, 2009; Cota, Blanc-Potard and Casadesús, 2012; Foley, 2015; Jiang *et al.*, 2019). This process works through diverse mechanisms of reversible genetic change to create subpopulations with distinct patterns of gene expression (Phillips *et al.*, 2019). For example, phase variation of *Bacteroides* CPS is known to create specialized subpopulations necessary for forming a biofilm (Béchon *et al.*, 2020). In addition, phase variation can create subpopulations with more promiscuous functions such as broad protection against different phage or antibiotics (Porter *et al.*, 2020; Porter *et al.*, 2017).

Despite the importance of phase variation to bacterial fitness in diverse environments, limitations in technology have hindered a fundamental understanding of the single-cell heterogeneity generated by combinations of multiple phase-variable loci. Previous studies of

phase variation have used fluorescent reporters that dissociate the loci from their natural genetic context and limit the number of loci that can be observed simultaneously (Saunders, Moxon and Gravenor, 2003; Criss, Kline and Seifert, 2005; Bayliss, 2009; Hung *et al.*, 2014). Furthermore, fluorescent reporters produce signals that are temporally disconnected from the genetic variation due to time delays in fluorescent protein maturation and decay (Krinos *et al.*, 2001).

In the abundant and prevalent genus in the human gut microbiome, *Bacteroides sp.*, recombinase-mediated promoter inversion creates diverse cell-surface capsular polysaccharide (CPS) phenotypes via reversible single-cell genetic heterogeneity affecting the expression of CPS biosynthetic operons (Krinos *et al*., 2001; Coyne *et al*., 2003). The expression of CPS of *Bacteroides fragilis*, a key human gut symbiont (Troy and Kasper, 2010), is controlled by the orientation of seven invertible promoters. These promoter orientations define discrete ON/OFF states of distinct CPS synthesis operons (**Fig. 1A**).

Since we have lacked tools to probe combinatorial promoter states within single cells, fundamental questions about phase variation in *Bacteroides sp.* have not yet been addressed. For example, it is not known whether the population composition of CPS states is strongly controlled or loosely influenced by rates of promoter inversion. In other words, do promoter inversions among CPS operons shape the population into a single or multiple distinct preferred compositions, or is the population weakly regulated with preference towards certain states? Alternatively, is the system dominated by stochastic processes such that no particular population composition is strongly favored (**Fig. 1B**)? A more detailed understanding of this fundamental question could provide insights into the contributions of stochastic processes, regulatory network architectures, and the role of history-dependence in shaping the dynamic behaviors of phenotypic diversification by combinatorial phase variation.

To answer these questions, we use ultrahigh-throughput single-cell targeted sequencing to determine the single-cell promoter states of all 7 phase-variable CPS promoters for multiple populations of *B. fragilis* over multiple timepoints. Exploiting synthetic control of the recombinase mediating promoter inversions, we show that populations starting with disparate promoter state compositions converge to a single steady-state promoter state composition. By fitting a stochastic dynamic model that assumes independent inversion rates to the time-series single-cell data, we show that the differential inversion rates of the promoters are major contributors to the observed dynamic behaviors. Our fitted model reveals the inversion rate parameters, which are overall biased towards the OFF state and not uniform across promoters.

To investigate the *B. fragilis* population diversification at the single-cell level, we used an ultrahigh-throughput, single-cell sequencing method referred to as Droplet Targeted Amplification

sequencing (DoTA-seq) **[to be available on BioArxiv soon]**. Although DoTA-seq can analyze combinations of genes in diverse gram-positive and gram-negative bacteria at the single-cell level, we simplified this protocol for *B. fragilis* (gram-negative) **(Methods)**. Briefly, single *B. fragilis* cells and random barcode oligos are encapsulated at limiting dilution into water-in-oil droplets at ~10 kHz rates along with multiplex PCR primer sets (**Fig 1C**). Multiplex PCR amplifies the seven CPS promoters tagged separately with a barcode sequence unique to each droplet to generate a single-cell amplicon sequencing library. The primer-annealing sites flank the inverted repeat sequences where promoter inversion occurs for each operon (Coyne *et al.*, 2003). Thus, each of the seven pairs of primers report the ON/OFF status of their respective promoter through the unique sequence of the amplicon generated (see Methods for detailed breakdown of the quality control and analysis pipeline, and **Supplementary Table 1** for a summary of the number of cells sequenced and passing quality controls for each library).

To determine the technical reproducibility of this method, we sequenced a single colony grown for 48 hours on an agar plate in technical replicates (~5000 cells per replicate). We detected 55 of 128 possible single-cell promoter states which represent 43% of all possible single-cell promoter states (**Fig 1D**). In a modified strain with the recombinase (*mpi*) responsible for inversion deleted (mpi8; Coyne *et al.*, 2003), 100% of the population displayed the same single-cell promoter state (~9000 cells) (**Fig 1E**). This indicates that the deletion of *mpi* locked the orientation of the promoters across the population, as expected. The variation between technical replicates was low for single-cell promoter states that displayed higher relative abundance. Consistent with this result, the coefficient of variation for technical replicates displayed a moderate negative correlation with the number of cells sequenced for each subpopulation **(Supplementary Fig 1),** suggesting some fraction of the variance could be attributed to random variation due to sampling (i.e. low abundance sub-populations may be inconsistently observed).

After demonstrating the robustness of the method, we investigated the variation in population-level promoter compositions across colonies (**Fig 1A**). To this end, we examined single-cell promoter orientations in five wild-type colonies picked at random, representing five independent populations **(Fig 2A)**. Across all five colonies, we observed 77 of 128 possible CPS combinations. Each colony displayed markedly different single-cell promoter states, with populations varying from primarily 0–4 promoters oriented on simultaneously (**Fig 2B**).

To understand global properties of promoter states across the population, we generated an undirected network that represents potential transitions between single-cell promoter states (**Fig 2C** and **Supplementary Fig 2**). Each single-cell promoter state is represented by a node

133 proportional in size to its relative abundance in the population. Nodes are connected if their
134 Hamming distance (HD) is equal to one, meaning that they are separated by a single promoter-
135 inversion (e.g., the HD between A--E--- and A-DE--- is 1).

136 While subpopulations with many (5-7) promoters turned off were frequently observed,
137 subpopulations with many promoters simultaneously on were rare, suggesting that *B. fragilis* limits
138 concomitant promoter-ON orientations. Although ~22-34% of the possible 128 states were
139 observed within a given colony, all subpopulations are connected to at least one other
140 subpopulation. This highly interconnected network suggests but does not conclusively define
141 derivative paths between subpopulations. In accordance with derivative development, however:
142 (i) larger (i.e., higher abundance) subpopulations, potentially derived earlier in population
143 development, had more time to produce many (>5) immediate neighbors **(Fig 2D)**. For
144 comparison, most subpopulations are connected to 3–5 immediate (HD = 1) neighbors
145 **(Supplementary Fig 3A)**; (ii) larger subpopulations were closer (i.e., lower HD) to the most
146 abundant combinatorial promoter state **(Fig 2E**, Spearman's Rho -0.65, $p < 2.2e-16$ between
147 subpopulation size and HD), despite most subpopulations being separated by a HD of 2–4
148 **(Supplementary Fig 3B)**; (iii) finally, most cells were only one promoter inversion away from the
149 most abundant promoter state (main promoter state) **(Fig 2B,E)**. The observed population-level
150 promoter state network properties could arise from two possibilities: (1) the population promoter
151 composition reflect a snapshot in early stages of a diversifying population, where the major
152 subpopulations are more closely related to the single-cell promoter state of the initial founding cell
153 or (2) the populations have reached steady-state and the observed promoter state composition
154 reflects multiple distinct steady-states that the system can converge to.

155 To determine whether these populations are in early stages of diversification or reflect
156 multiple steady-state compositions, we individually cultured the five colonies (**Fig 2B**) for two
157 weeks. The cultures were diluted every 8 or 16 hours alternately with different inoculum volumes
158 to prevent the cultures from entering late stationary phase. To track the temporal diversification
159 process, we performed DoTA-seq on the populations on day 3, 7, and 14 **(Fig 3A)**. During this
160 time, the Shannon diversity of individual cultures increased as new distinct single-cell promoter
161 states arose, and cells became more evenly distributed among different promoter states **(Fig 3B)**.
162 By day 14, each population individually exhibited on average 89 +/- 9 unique CPS states and
163 displayed 117 unique CPS states across all colonies **(Supplemental Table 2)**. Although these
164 populations had not yet reached steady-state (i.e. population-level promoter composition was still
165 changing over time), the population-level promoter compositions became increasingly similar as
166 time progressed (**Fig 3C**). These results suggest that populations from different initial starting

167 states eventually converge to a single steady-state population-level promoter composition.
168 However, the time to convergence was on the scale of weeks, which suggests that intrinsic
169 regulation weakly shapes population composition towards the preferred state in our experimental
170 conditions. In addition, the large diversity of single-cell promoter states of the populations at day
171 14 highlights the diverse repertoire of variation available to *B. fragilis* populations to cope with
172 environmental stresses.

173 The population dynamics and the steady-state promoter distributions could be driven by
174 regulation (e.g. promoter inversion rates) or variation in subpopulation fitness. We did not observe
175 statistically significant differences in the growth rates between different CPS locked strains when
176 grown in supplemented basal media (**Supplementary Fig 4).** Therefore, we considered the
177 contributions of the rate of recombinase-mediated promoter inversions to the population promoter
178 composition dynamics. In many bacterial species, the rates of phase-variation can change in
179 response to environmental cues affecting expression of *trans*-acting factors (Bayliss, 2009). In *B.*
180 *fragilis* specifically, expression of the Mpi recombinase may be regulated by additional phase-
181 variation mechanisms (Kuwahara et al., 2004). To eliminate these regulatory contributions, we
182 constructed a strain enabling precise control of expression of the recombinase. To this end, we
183 introduced a tightly regulated (i.e., low leaky expression), aTc-inducible *mpi* at an ectopic location
184 on the genome in an *mpi* deletion background in *B. fragilis*. In the presence of high aTc
185 concentrations, *mpi* can be expressed at high levels from this inducible promoter.

186 We first used this system to generate and isolate phase-locked variants starting from a
187 single colony (**Supplementary Fig 5**) that were locked into their respective single-cell promoter
188 states (A--E---, -B---G-, and --D---H) in the absence of aTc. In the presence of aTc, recombinase
189 expression is turned on and the promoter states diversify across the population as a function of
190 time. To investigate the diversification process over time, we grew the phase-locked strains to
191 mid-exponential-phase, induced recombinase expression, and then sampled these populations
192 over time after induction (**Fig 3D**). The induced populations showed a converging diversification
193 trajectory (**Fig 3E**) by approaching a single steady-state population composition 24 hours
194 following induction of recombinase (**Fig 3F**). Thus, high expression of recombinase (ie. increasing
195 rate of phase variation) resulted in a fast convergence of the population to the steady-state
196 composition, thus indicating a strong regulation of the population composition.

197 Computational modeling is a powerful way to gain mechanistic insight into complex
198 biological systems (Qian, Lan and Venturelli, 2021). However, its utilization is often impaired by
199 the complexity of biological systems combined with a lack of properly collected data resulting in

under-constrained model parameters. Single-cell data fitted to stochastic models provide additional parameter constraints that bulk data does not (Munsky, Fox and Neuert, 2015). Leveraging our single-cell data, we constructed a mechanistic computational model to describe the population-level promoter state composition as a dynamic process in which promoters flip independently and stochastically at different rates at the single-cell level. This process is represented as a continuous-time Markov model consisting of 128 discrete states (**Fig 4A, Supplementary Materials**). The system can be described by a series of chemical master equations that we solved to fit the model to the time-series data.

By fitting our model to single-cell promoter state time-series measurements of the synthetic strain, we evaluated the contribution of promoter inversion rates to the system's behavior in the absence of potential transcriptional regulation of the recombinase. This model was able to fit the synthetic strain time series data extremely well (Spearman's Rho ~0.98, p = 2.57e-275), suggesting that the assumptions of the model were consistent with the experimental system (**Fig 4B-C**). In addition, all model parameters were well-constrained by the time series data except D-ON and G-OFF **(Supplementary Table 3)**. To determine if the observed goodness of fit was due to the patterns observed in the experimental data, we shuffled the dataset labels and re-fit the model. The model did not fit well to the label-shuffled datasets **(Supplementary Fig 7)**. These results suggest that under synthetic control of the recombinase, the population diversification trajectories are driven primarily by the differential flipping rate constants for each promoter.

Using the fitted model, we extrapolated to the final steady-state distribution of subpopulations (**Fig 4D**). The population-level steady-state exhibited high diversity, with most single cells displaying 0-2 invertible promoters simultaneously ON, and was dominated by promoter combinations that have promoter A turned ON. Dominance of the promoter A in the ON state was reflected in the promoter inversion parameters, where only promoter A had a higher rate of flipping ON than OFF (**Fig. 4E)**. Promoters F, G, and H had higher inversion rates (both ON and OFF) than the other promoters, leading to a faster convergence of these promoters to steady state than the other promoters with smaller magnitude ON and OFF rates **(Supplementary Fig 8)**. The observed differences in kinetic rates could reflect differential roles that CPSs play in the biology of the bacterium.

Using our model, we explored whether the few promoter combinations (e.g. ABDEFG or -BDEFG) that were never observed in our experimental data sets were excluded due to (1) some additional factors (such as incompatibility between the capsules) precluding the existence of those states or (2) a direct consequence of the inversion rates. In the simulated steady state, which

233 represents the most diverse population, they were present at <10^{-5} frequency. This low frequency
234 would be unlikely to be detected in our datasets given that we sequenced ~5000 to ~10000 cells
235 for each population. This result, while not ruling out additional factors (1), is consistent with (2)
236 that the absence of some promoter combinations from our datasets can be explained simply by
237 the differences in inversion rates.

238 To evaluate the contribution of promoter inversion rates to the population dynamics of
239 wild-type *B. fragilis,* we fit the same stochastic model to time-series data of the 5 wild-type colonies
240 **(Fig 3A)**. The model displayed a moderately good fit to these data (Spearman's Rho ~ 0.8, p =
241 3.40e-132) **(Fig 4F)** which was significantly better than label-shuffled control data **(Fig 4G).** This
242 implies that the independent rates of promoter inversion accounted for a portion of the dynamics
243 in this experimental system. Certain posterior parameter distributions (F-OFF and G-OFF)
244 exhibited bimodality and were not well-constrained (>5% coefficient of variation in parameter
245 distribution) (**Supplementary Fig 6B, Supplementary Table 3**).

246 These results suggest that additional factors beyond those captured in the model (such
247 as regulation over recombinase levels or activity) contributes to the population dynamics in
248 wildtype populations, especially for promoters F and G. Consistent with this notion, promoters F
249 and G harbor extended inverted repeat DNA elements surrounding the promoter (Patrick et al.,
250 2003), potentially influencing recombinase activity or dependence on other regulatory factors.
251 Additional proteins or recombinases may also recognize specific CPS promoters, leading to
252 additional layers of regulation over specific CPSs. Further work will be required to elucidate these
253 additional layers of regulation over promoter inversion rates, especially in biologically relevant
254 environmental contexts (Kuwahara *et al.*, 2004). Alternatively, moderate differences in
255 subpopulation fitness levels could also impact growth over a two-week period, whereas these
256 impacts were small over fewer generation times in the case of synthetic strains. In sum, the
257 model's excellent goodness of fit to the synthetic strain temporal data indicates that promoter
258 inversion rates are a major mechanism driving the CPS diversity of *B. fragilis* populations. By
259 contrast, the model's moderate goodness of fit to the wildtype data suggests additional layers of
260 regulation are present, especially for promoters F and G.

261 It is well known that phase variation is a wide-spread mechanism of generating population
262 heterogeneity for diverse bacteria, but there is limited understanding of population composition,
263 dynamics, and regulation of the heterogeneity generated through phase variation (van der Woude
264 and Bäumler, 2004; Bayliss, 2009). By profiling the many phase variable CPS loci in *B. fragilis*
265 populations at the single-cell level as a function of time, we addressed fundamental questions

about the regulation of population heterogeneity in this system. Population dynamics are fundamentally shaped by different rates of promoter inversion, which determines both the steady-state population composition and how quickly the population reaches the steady-state. The convergence of this system to a single-steady-state as a function of time is in contrast to other diversification gene regulatory networks, which are strongly influenced by interlinked positive and negative feedback loops and exhibit multi-stable behavior (Venturelli *et al.*, 2015).

Our 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. Phase variation rates have been known to change according to environmental cues (Bayliss, 2009) and it will be an interesting question whether *B. fragilis* uses such a strategy to affect its population composition under physiologically relevant environmental conditions.This novel method of studying phase-variation as well as our findings can be extended to diverse bacteria, including pathogens, that use phase variation in key parts of their life cycle (Cowley, Myltseva and Nano, 1996; Snellings, Tall and Venkatesan, 1997; Lavitola *et al.*, 1999; Gunther IV *et al.*, 2002; Snyder *et al.*, 2010; Chopra-Dewasthaly *et al.*, 2012; Beare *et al.*, 2018; Safi *et al.*, 2019; Zhang *et al.*, 2019). Thus, our results set the stage for understanding how bacterial phenotypic diversity is used by bacteria to respond to environmental perturbations such antibiotic stress, microbial warfare, and phage and host immune cell attack.

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. This work was supported by the National Institutes of Health R35 GM12477409 and R21 AI156438, and Army Research Office W911NF-19-1-0269. J.S. was supported in part by the NIH Biotechnology Training Grant (T32 GM008349) and NIH F31 Graduate Fellowship (F31GM142153). F.L. was supported in part by the Burroughs Wellcome Fund Careers Award in the Scientific Interface. 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.

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 the stochastic promoter inversion computational models, wrote custom scripts

related to these models. 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.

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

Plating and harvesting of B. fragilis colonies

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

Long-term culturing of *B. fragilis* populations

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

Growth curves of wildtype and CPS promoter locked strains

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

Fitting of growth curves to data

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

Fabrication of microfluidic droplet maker

Microfluidic masters were fabricated in a clean room using soft lithography (Duffy *et al.*, 1998). SU-8 3000 photoresist (MicroChem) was spun on a silicon wafer (University Wafers) to achieve a thickness of 30 µm. Photolithography masks **(supplementary materials)** were ordered from CAD/ART and used to pattern the photoresist using a UV Led (ThorLabs) for 1 min 45s. The patterned photoresist was baked for 5 minutes at 95oC then developed using SU-8 developer (PGMEA) for 2 minutes then baked at 200oC for 2 minutes.

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

387

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

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

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

397 This PCR mixture is combined with 25µL of cells diluted with pre-injection buffer **(**10mM HEPES
398 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
399 mixture is 0.0025. This mixture is then injected into the droplet making microfluidics device using
400 syringe pumps (New Era) and 1mL syringes (BD) at a flowrate of 400 µL/hr and 800 µL/hr of
401 Biorad Evagreen droplet making oil (Biorad) to generate droplets.

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

405 After PCR, the coalesced droplets were removed using a pipette, and the emulsion was broken
406 on ice by adding 20 µL 500mM EDTA and 20 µL perfluoro-octanol (Sigma 370533), then vortexed
407 followed by a spin pulse centrifugation. The aqueous phase was transferred to another tube by
408 pipette and cleaned up using a Zymo cleanup and concentrator kit (Zymo Research D4013), then
409 subject to size selection using SPRI-select beads (Beckman Coulter B23317) using 0.7x volume
410 of beads. A further round of size selection was performed in 100 mM NaOH, 10% Ethanol, and
411 1.4x volume of beads to increase the purity of the library.

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

415

416 **Analysis of sequencing data**

417 The raw sequencing reads are obtained from the Miseq. Read 1 represents the targeted
418 amplicons. The first index read represents the unique cell barcode. The second index read is
419 used to multiplex libraries from different experiments. Demultiplexing of different libraries (index
420 2) is performed by the Miseq software. Cell barcode demultiplexing and quality control is
421 performed using a custom python script (R4-parser.py). The filtered reads were mapped to
422 custom built reference databases containing B. fragilis CPS promoter sequences and is available

423 on GitHub (see code availability) using Bowtie2 V2.3.4.1(Langmead and Salzberg, 2012) using
424 "--very-sensitive" presets. The mapped reads were analyzed using custom scripts as follows: The
425 mapped reads are organized into read groups consisting of reads with the same unique cell
426 barcode representing amplicons from the same droplet. Read groups with too few reads are
427 removed. The filtered groups are transformed into a table containing the barcode and a tally of
428 the number of reads that are mapped to each target (SAM-analysis.py). We further filtered the
429 reads by removing the read groups if the reads mapping to any promoter (sum of ON and OFF
430 orientation) are less than 1% of total reads for that group. Subsequently, if reads in any promoter
431 orientation were less than 1% of total reads for the barcode, we set the value to 0; this accounted
432 for expected noise. We next discretized the data by replacing values with 1 or 0, corresponding
433 to ON or OFF, respectively. We removed barcodes if neither orientation was found for any
434 promoter (i.e., if ON + OFF values = 0). We also removed read groups if both orientations were
435 found for any locus (ON + OFF = 2). The set of remaining read groups are used to determine the
436 frequencies of each promoter state combination in the population **(CPS-analysis.R).** This
437 processed list of frequencies for each promoter state is available in Supplementary Table 2.

438 All scripts are available on GitHub (see code availability).

439

440 **Computational modelling of promoter inversion dynamics**

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

453

454

455

456

457 **References**

458 Bayliss, C.D. (2009) 'Determinants of phase variation rate and the fitness implications of
459 differing rates for bacterial pathogens and commensals.', *FEMS microbiology reviews*, 33(3),
460 pp. 504–20. Available at: https://doi.org/10.1111/j.1574-6976.2009.00162.x.

461 Beare, P.A. *et al.* (2018) 'Genetic mechanisms of Coxiella burnetii lipopolysaccharide phase
462 variation', *PLOS Pathogens*. Edited by D.M. Monack, 14(3), p. e1006922. Available at:
463 https://doi.org/10.1371/journal.ppat.1006922.

464 Chopra-Dewasthaly, R. *et al.* (2012) 'Role of Vpma phase variation in *Mycoplasma agalactiae*
465 pathogenesis', *FEMS Immunology & Medical Microbiology*, 66(3), pp. 307–322. Available at:
466 https://doi.org/10.1111/j.1574-695X.2012.01010.x.

467 Cota, I., Blanc-Potard, A.B. and Casadesús, J. (2012) 'STM2209-STM2208 (opvAB): A Phase
468 Variation Locus of Salmonella enterica Involved in Control of O-Antigen Chain Length', *PLoS*
469 *ONE*. Edited by D. Chakravortty, 7(5), p. e36863. Available at:
470 https://doi.org/10.1371/journal.pone.0036863.

471 Cowley, S.C., Myltseva, S.V. and Nano, F.E. (1996) 'Phase variation in Francisella tularensis
472 affecting intracellular growth, lipopolysaccharide antigenicity and nitric oxide production',
473 *Molecular Microbiology*, 20(4), pp. 867–874. Available at: https://doi.org/10.1111/j.1365-
474 2958.1996.tb02524.x.

475 Coyne, M.J. *et al.* (2003) 'Mpi recombinase globally modulates the surface architecture of a
476 human commensal bacterium', *Proceedings of the National Academy of Sciences*, 100(18), pp.
477 10446–10451. Available at: https://doi.org/10.1073/pnas.1832655100.

478 Criss, A.K., Kline, K.A. and Seifert, H.S. (2005) 'The frequency and rate of pilin antigenic
479 variation in Neisseria gonorrhoeae.', *Molecular microbiology*, 58(2), pp. 510–9. Available at:
480 https://doi.org/10.1111/j.1365-2958.2005.04838.x.

481 Duffy, D.C. *et al.* (1998) 'Rapid Prototyping of Microfluidic Systems in Poly(dimethylsiloxane)',
482 *Analytical Chemistry*, 70(23), pp. 4974–4984. Available at: https://doi.org/10.1021/ac980656z.

483 Foley, J. (2015) 'Mini-review: Strategies for Variation and Evolution of Bacterial Antigens',
484 *Computational and Structural Biotechnology Journal*, 13, pp. 407–416. Available at:
485 https://doi.org/10.1016/j.csbj.2015.07.002.

486 Gunther IV, N.W. *et al.* (2002) 'Assessment of Virulence of Uropathogenic *Escherichia coli* Type
487 1 Fimbrial Mutants in Which the Invertible Element Is Phase-Locked On or Off', *Infection and*
488 *Immunity*, 70(7), pp. 3344–3354. Available at: https://doi.org/10.1128/IAI.70.7.3344-3354.2002.

489 Hung, M. *et al.* (2014) 'Modulating the frequency and bias of stochastic switching to control
490 phenotypic variation', *Nature Communications*, 5(1), p. 4574. Available at:
491 https://doi.org/10.1038/ncomms5574.

492 Jiang, X. *et al.* (2019) 'Invertible promoters mediate bacterial phase variation, antibiotic
493 resistance, and host adaptation in the gut', *Science*, 363(6423), pp. 181–187. Available at:
494 https://doi.org/10.1126/science.aau5238.

495 Krinos, C.M. *et al.* (2001) 'Extensive surface diversity of a commensal microorganism by
496 multiple DNA inversions', *Nature*, 414(6863), pp. 555–558. Available at:
497 https://doi.org/10.1038/35107092.

498 Kuwahara, T. *et al.* (2004) 'Genomic analysis of Bacteroides fragilis reveals extensive DNA
499 inversions regulating cell surface adaptation', *Proceedings of the National Academy of*
500 *Sciences*, 101(41), pp. 14919–14924. Available at: https://doi.org/10.1073/pnas.0404172101.

501 Langmead, B. and Salzberg, S.L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature*
502 *Methods*, 9(4), pp. 357–359. Available at: https://doi.org/10.1038/nmeth.1923.

503 Lavitola, A. *et al.* (1999) 'Intracistronic transcription termination in polysialyltransferase gene
504 (siaD) affects phase variation in Neisseria meningitidis', *Molecular Microbiology*, 33(1), pp.
505 119–127. Available at: https://doi.org/10.1046/j.1365-2958.1999.01454.x.

506 Munsky, B., Fox, Z. and Neuert, G. (2015) 'Integrating Single-Molecule Experiments and
507 Discrete Stochastic Models to Understand Heterogeneous Gene Transcription Dynamics',
508 *Methods (San Diego, Calif.)*, 85, pp. 12–21. Available at:
509 https://doi.org/10.1016/j.ymeth.2015.06.009.

510 Phillips, Z.N. *et al.* (2019) 'Phase-variable bacterial loci: how bacteria gamble to maximise
511 fitness in changing environments', *Biochemical Society Transactions*, 47(4), pp. 1131–1141.
512 Available at: https://doi.org/10.1042/BST20180633.

513 Qian, Y., Lan, F. and Venturelli, O.S. (2021) 'Towards a deeper understanding of microbial
514 communities: integrating experimental data with dynamic models', *Current Opinion in*
515 *Microbiology*, 62, pp. 84–92. Available at: https://doi.org/10.1016/j.mib.2021.05.003.

516 Safi, H. *et al.* (2019) 'Phase variation in *Mycobacterium tuberculosis glpK* produces transiently
517 heritable drug tolerance', *Proceedings of the National Academy of Sciences*, 116(39), pp.
518 19665–19674. Available at: https://doi.org/10.1073/pnas.1907631116.

519 Saunders, N.J., Moxon, E.R. and Gravenor, M.B. (2003) 'Mutation rates: estimating phase
520 variation rates when fitness differences are present and their impact on population structure',
521 *Microbiology*, 149(2), pp. 485–495. Available at: https://doi.org/10.1099/mic.0.25807-0.

522 Snellings, N.J., Tall, B.D. and Venkatesan, M.M. (1997) 'Characterization of Shigella type 1
523 fimbriae: expression, FimA sequence, and phase variation', *Infection and Immunity*, 65(6), pp.
524 2462–2467. Available at: https://doi.org/10.1128/iai.65.6.2462-2467.1997.

525 Snyder, L.A. *et al.* (2010) 'Simple sequence repeats in Helicobacter canadensis and their role in
526 phase variable expression and C-terminal sequence switching', *BMC Genomics*, 11(1), p. 67.
527 Available at: https://doi.org/10.1186/1471-2164-11-67.

528 Stapels, D.A.C. *et al.* (2018) 'Salmonella persisters undermine host immune defenses during
529 antibiotic treatment', *Science*, 362(6419), pp. 1156–1160. Available at:
530 https://doi.org/10.1126/science.aat7148.

531 Troy, E.B. and Kasper, D.L. (2010) 'Beneficial effects of Bacteroides fragilis polysaccharides on
532 the immune system', *Frontiers in bioscience : a journal and virtual library*, 15, pp. 25–34.

533 Venturelli, O.S. *et al.* (2015) 'Population Diversification in a Yeast Metabolic Program Promotes
534 Anticipation of Environmental Shifts', *PLOS Biology*, 13(1), p. e1002042. Available at:
535 https://doi.org/10.1371/journal.pbio.1002042.

536 van der Woude, M.W. and Bäumler, A.J. (2004) 'Phase and Antigenic Variation in Bacteria',
537 *Clinical Microbiology Reviews*, 17(3), pp. 581–611. Available at:
538 https://doi.org/10.1128/CMR.17.3.581-611.2004.

539 Zhang, A.-N. *et al.* (2019) 'Choosing Your Battles: Which Resistance Genes Warrant Global
540 Action?', *bioRxiv*, (October), p. 784322. Available at: https://doi.org/10.1101/784322.

541

542

543 **Figure captions:**

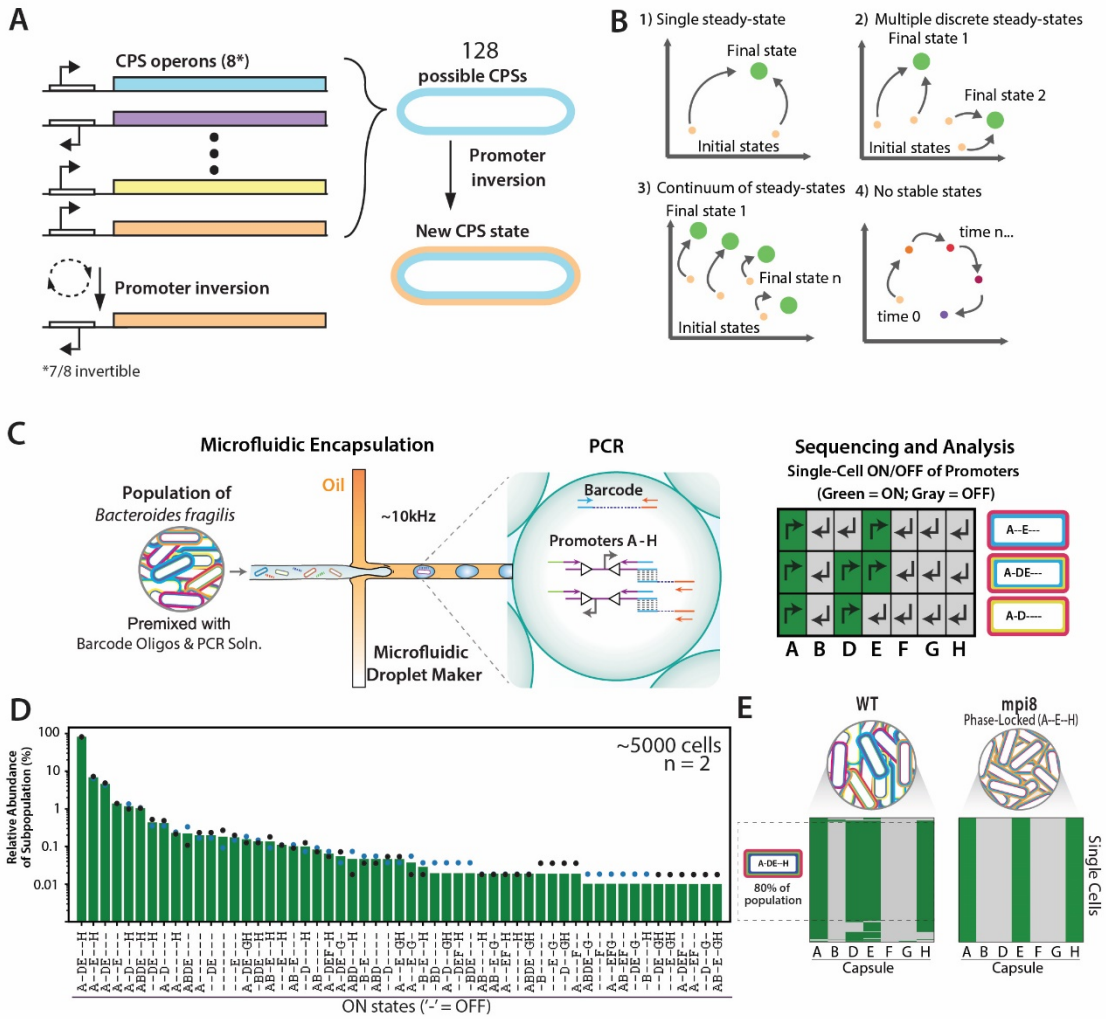
544 Figure 1. DoTA-seq reveals bacterial subpopulations defined by multiple phase variable loci. A)
545 Summary schematic of *B. fragilis* CPS loci. Stochastic inversions of the promoters generates
546 diversity in CPS expression. B) Cartoon schematic of hypothetical population behaviors of the
547 CPS system under continuous growth. Populations could 1) converge to a single stable state, 2)
548 converge to multiple stable states, 3) exist in a continuum of stable states, or 4) remain in flux
549 with no stable states. C) Cartoon representation of the ultrahigh-throughput single-cell targeted
550 sequencing workflow for B. fragilis CPS promoters. We first premix cells from a colony or liquid
551 culture with barcode oligos, high-fidelity PCR mix, and target-loci primers (14 primers for the 7
552 invertible CPS promoters) containing adaptor sequences that overlap the barcode oligo. Cells in
553 this mix are then encapsulated into picolitre emulsion droplets at limiting dilution so that most
554 droplets contain 1 or 0 cells in a Poisson distribution. The barcode oligos also exist at limiting
555 dilution so that droplets containing cells also contain 1 or 0 barcodes. Droplets that contain both
556 a barcode oligo and a cell are selectively amplified and linked to genomic loci via overlap
557 extension PCR. The cell is lysed during the initial PCR denaturing step. Only amplicons linking
558 the barcode and loci contain ends that can be sequenced. After library prep, sequencing reveals
559 the promoter orientations of all targeted loci for thousands of cells. Right: promoter orientation
560 heatmap for 3 example cells. Coverage in the 'on' orientation for a particular CPS promoter is
561 green. Dashes ('-') represent promoters that are oriented off. D) The CPS promoter states of a
562 single colony of B. fragilis, shown as a bar plot. Dots represent range of two technical replicates.
563 E) Heatmaps as described in (A) for thousands of cells, with groups of rows clustered to reflect
564 subpopulations with alternative combinatorial promoter orientations. A control strain with the
565 primary recombinase gene (*mpi*) deleted displays a population with 100% of cells having the
566 expected promoter configuration (A--E--H) for this particular isolate.

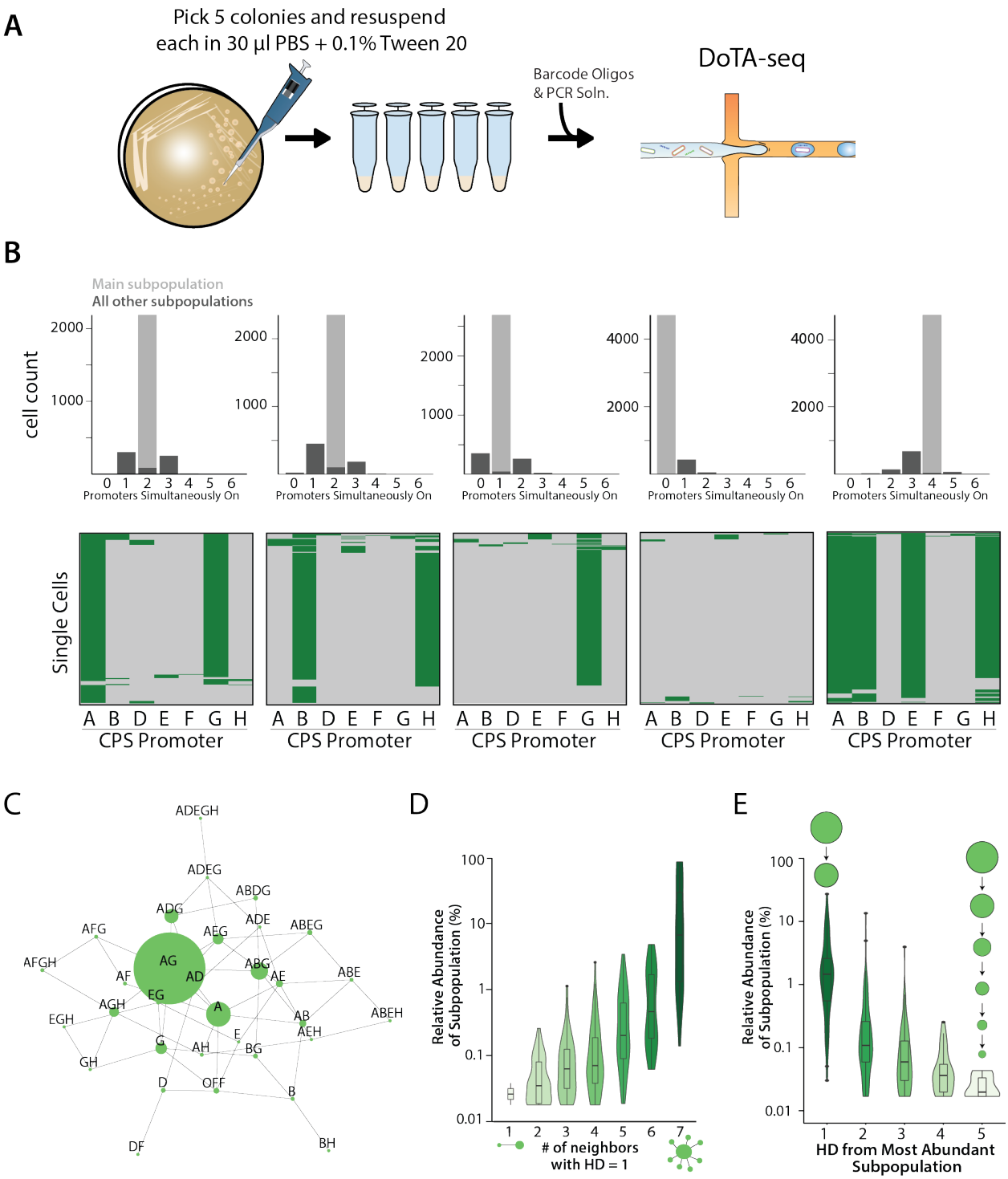
567 Figure 2. *B. fragilis* colonies display diverse compositions with similar features. A) Cartoon
568 representation of experiment. A single glycerol stock was streaked onto an agar plate, and 5
569 colonies are picked for sequencing after 48 hours of growth. B) Heatmaps from 5 wild-type
570 colonies. Above: Histogram of cells having *n* promoters simultaneously oriented on. C)
571 Representative network representation of single-cell promoter state variants in colony 1. By
572 connecting subpopulations with a Hamming Distance (HD) of 1 (i.e., separated by a single
573 promoter inversion event), we can generate a network representation of the subpopulations. Each
574 node represents one subpopulation. Node size represents relative abundance. For network
575 graphs of all colonies, see supplementary figure 2. D) Violin plot of the size of subpopulations with
576 different numbers of neighbors (where HD = 1). All colonies show the characteristic large
577 subpopulation surrounded by progressively smaller populations. E) Violin plot of the size of the
578 subpopulations versus the distance from the most abundant (main) subpopulation showing that
579 most of the large populations are found close (lower HD) to the main subpopulation.

580 Figure 3. Population-level promoter state composition converge to a single steady-state
581 composition over time. A) Cartoon schematic of the experiment. The 5 colonies from figure 2 are
582 grown in broth media for 14 days. Samples are taken and single-cell sequenced at multiple
583 timepoints. B) Shannon diversities of the 5 wildtype populations originating from the 5 colonies
584 over each timepoint. All 5 populations show increasing diversity at every point sampled, but the
585 rates of increase differ between different colonies. Day 0 represents the populations as colonies
586 on a plate. C) The populations show a converging trajectory over time. Right: Population-level

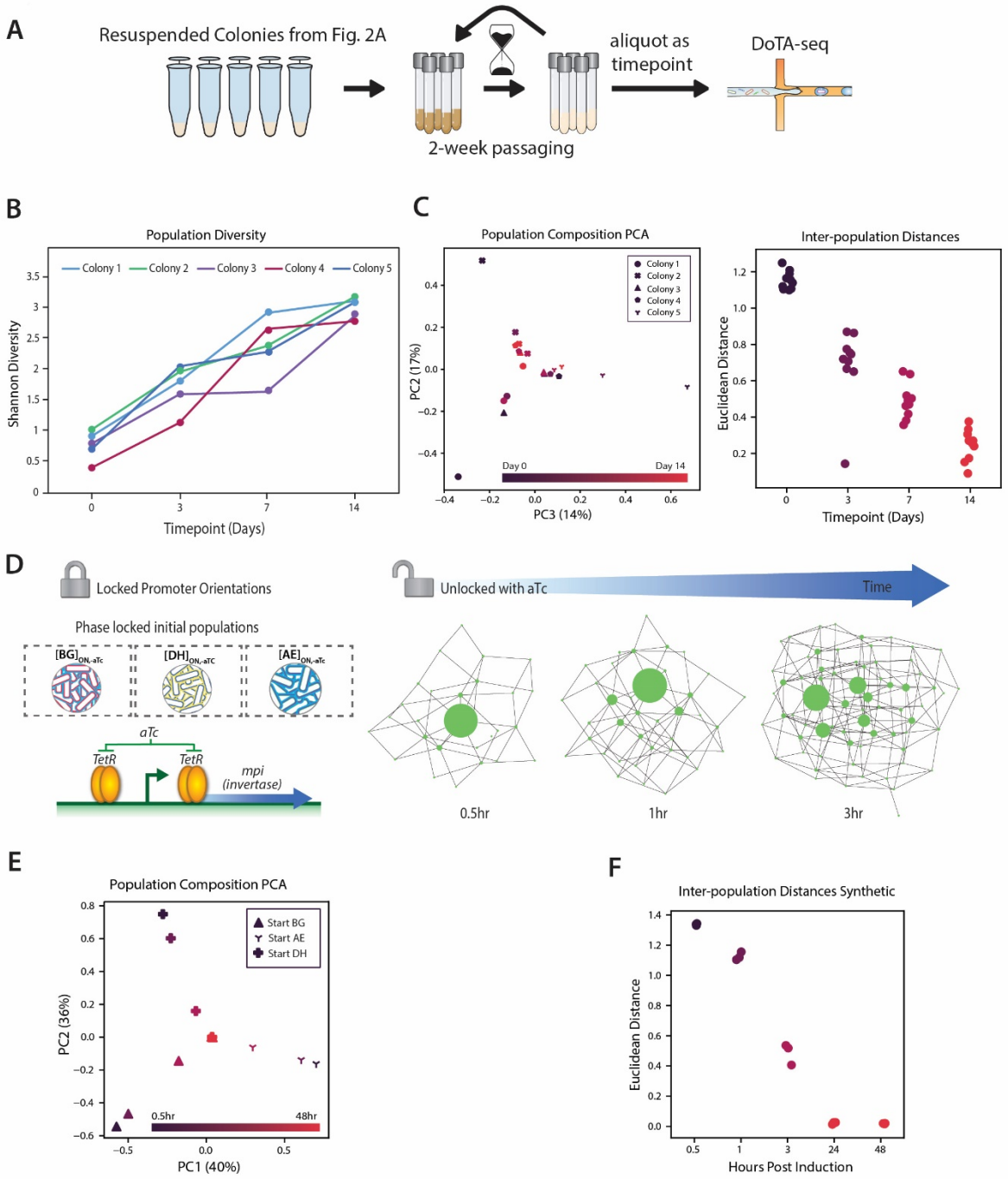
promoter state compositions are visualized by plotting on principal component axes. Marker shapes represent different populations. Marker colors represent different timepoints. Left: The pair-wise Euclidean distance between all 5 populations at each timepoint. The trend of decreasing Euclidean distance shows a converging trajectory of the populations. D) Cartoon representation of the induced promoter inversion experiment. The native recombinase gene (*mpi*) responsible for CPSs promoter inversions is placed under an aTc-inducible promoter. Inversion is induced in phase-locked variants during log phase by addition of aTc at 100 ng/μL. Samples are taken at precise timepoints afterwards. E) Induced populations also show a converging trajectory when visualized by plotting on principal component axes. Marker shapes represent different populations. Marker colors represent different timepoints. F) Pairwise Euclidean distances between the 3 induced populations at different timepoints post induction. By 24 hours the 3 populations have already converged.

Figure 4. A stochastic dynamic model of promoter inversions captures the population dynamics of CPS phase variation. A) A simplified representation (only two promoters) of the stochastic dynamic model for single-cell promoter states which is used to describe the population-level promoter compositions at different timepoints. In the simplified representation, the cell transitions between 4 different states by inverting promoters on and off one at a time with different kinetic rate parameters for each type of promoter inversion. A and B represent different operon promoters in the ON state and "-" represents the promoter in the OFF state. In the complete model there are 7 promoters, 14 kinetic parameters, and 128 different states. B) Comparison of model predictions and experimental data for each of the 128 possible promoter combination states for each timepoint of the induction experiment, showing excellent agreement between model and data. Rho represents the Spearman correlation coefficient. C) Model predicted and experimental population data is visualized by plotting the first 2 components of a Principal Component Analysis, showing close concordance between the populations. D) Network graph representation of the population composition for the model-predicted steady-state population. Subpopulations are connected if 1 HD apart. Node size represents relative subpopulation size. E) Kinetic rate parameters for promoter inversion derived from the model. Error bars represent standard deviations of the Monte-Carlo Marko-chain parameter distribution. For histograms of parameter distributions see supplemental figure 7. F) Comparison of model predictions and experimental data for the model trained on 14-day wild-type time-series data. Rho represents Spearman's correlation coefficient between experimental data and model predictions. G) Comparison of Spearman's correlation between model prediction and experimental data for the wild-type strain time-series and label-shuffled control datasets.


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 discrete 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
Single-Cell ON/OFF of Promoters
(Green = ON; Gray = OFF)
Population of
Bacteroides fragilis
Premixed with
Barcode Oligos & PCR Soln.
Oil
~10kHz
Microfluidic
Droplet Maker
Barcode
Promoters A-H
A-E---
A-DE---
A-D----
A B D E F G H
D
Relative Abundance
of Subpopulation (%)
~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

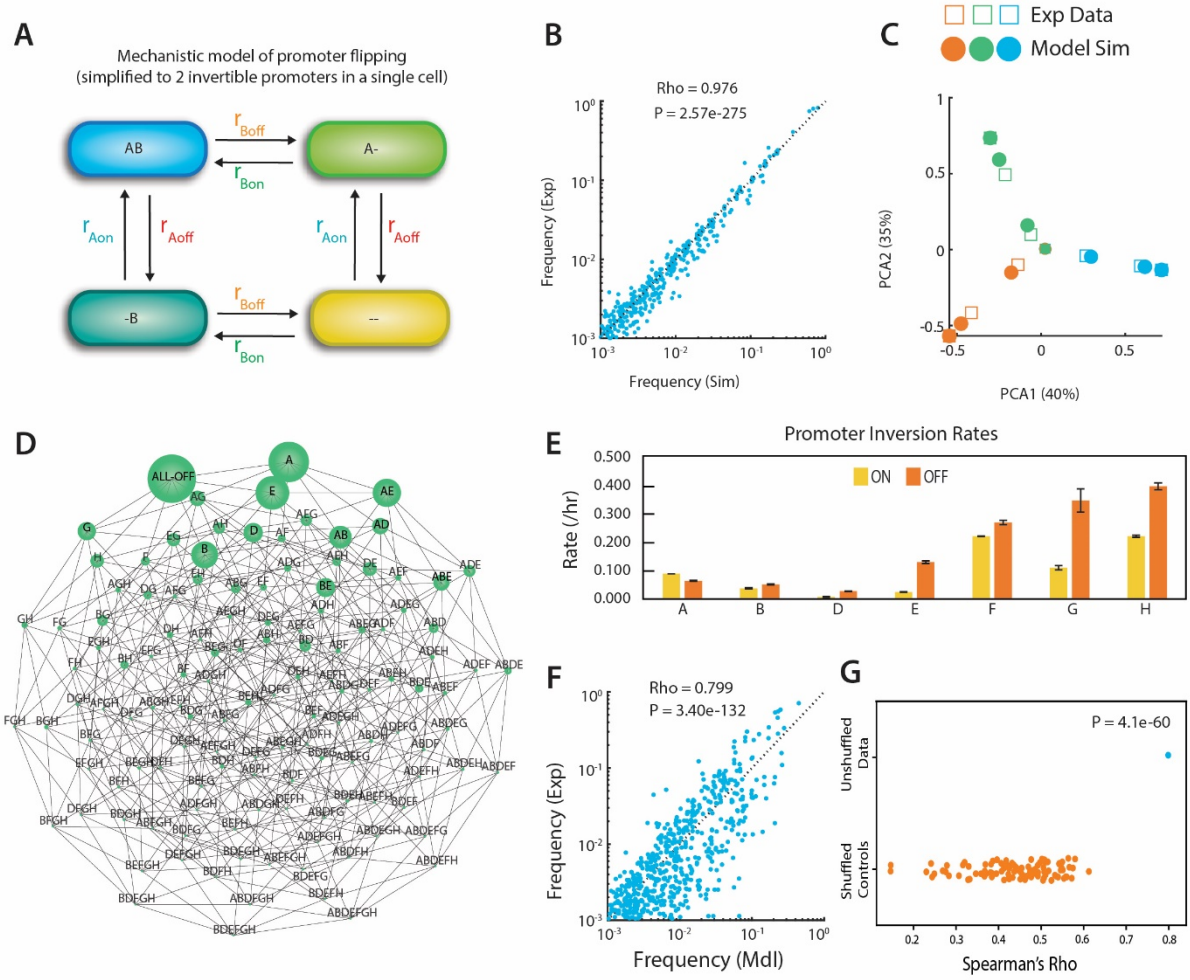

A
Pick 5 colonies and resuspend
each in 30 µl PBS + 0.1% Tween 20
Barcode Oligos
& PCR Soln.
DoTA-seq
B
Main subpopulation
All other subpopulations
cell count
Promoters Simultaneously On
Single Cells
A B D E F G H
CPS Promoter
C
ADEGH
ADEG
ABDG
ADG
ADE
ABEG
AFG
AEG
AFGH
AG
ABG
AF
AD
AE
ABE
EG
AGH
A
EGH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

623


A
Resuspended Colonies from Fig. 2A
2-week passaging
aliquot as timepoint
DoTA-seq
B
Population Diversity
Colony 1
Colony 2
Colony 3
Colony 4
Colony 5
Shannon Diversity
Timepoint (Days)
C
Population Composition PCA
PC2 (17%)
PC3 (14%)
Day 0
Day 14
Inter-population Distances
Euclidean Distance
Timepoint (Days)
D
Locked Promoter Orientations
Unlocked with aTc
Time
Phase locked initial populations
[BG]ON,-aTc
[DH]ON,-aTc
[AE]ON,-aTc
aTc
TetR
TetR
mpi (invertase)
0.5hr
1hr
3hr
E
Population Composition PCA
Start BG
Start AE
Start DH
PC2 (36%)
PC1 (40%)
0.5hr
48hr
F
Inter-population Distances Synthetic
Euclidean Distance
Hours Post Induction

624

625


A
Mechanistic model of promoter flipping
(simplified to 2 invertible promoters in a single cell)
AB
A-
-B
--
r Boff
r Bon
r Aon
r Aoff
B
Rho = 0.976
P = 2.57e-275
Frequency (Exp)
Frequency (Sim)
C
Exp Data
Model Sim
PCA2 (35%)
PCA1 (40%)
D
ALL-OFF
E
Promoter Inversion Rates
ON
OFF
Rate (/hr)
A
B
D
E
F
G
H
F
Rho = 0.799
P = 3.40e-132
Frequency (Exp)
Frequency (Mdl)
G
P = 4.1e-60
Unshuffled Data
Shuffled Controls
Spearman's Rho