

# Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium

Denise M. Wolf<sup>1,ϕ</sup>, Lisa Fontaine-Bodin<sup>1,δ,ϕ</sup>, Ilka Bischofs<sup>δ,ϕ</sup>, Gavin Price<sup>δ,ϕ</sup>, Jay Keasling<sup>δ,γ,ϕ</sup>, and Adam P. Arkin<sup>δ,ϕ</sup>,

<sup>ϕ</sup>Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720;  
Departments of <sup>δ</sup>Bioengineering and <sup>γ</sup>Chemical Engineering, University of California, Berkeley, CA 94720;

**Corresponding authors:** Denise Wolf [dmwolf@lbl.gov](mailto:dmwolf@lbl.gov) and Adam Arkin [aparkin@lbl.gov](mailto:aparkin@lbl.gov)

## Abstract

Memory is usually associated with higher organisms rather than bacteria. However, evidence is mounting that many regulatory networks within bacteria are capable of complex dynamics and multi-stable behaviors that have been linked to memory in other systems. Moreover, it is recognized that bacteria that have experienced different environmental histories may respond differently to current conditions. These “memory” effects may be more than incidental to the regulatory mechanisms controlling acclimation or to the status of the metabolic stores. Rather, they may be regulated by the cell and confer fitness to the organism in the evolutionary game it participates in. Here, we propose that history-dependent behavior is a potentially important manifestation of memory, worth classifying and quantifying. To this end, we develop an information-theory based conceptual framework for measuring both the persistence of memory in microbes and the amount of information about the past encoded in history-dependent dynamics. This method produces a phenomenological measure of cellular memory without regard to the specific cellular mechanisms encoding it. We then apply this framework to a strain of *Bacillus subtilis* engineered to report on commitment to sporulation and degradative enzyme (AprE) synthesis and estimate the capacity of these systems and growth dynamics to ‘remember’ 10 distinct cell histories prior to application of a common stressor. The analysis suggests that *B. subtilis* remembers, both in short and long term, aspects of its cell history, and that this memory is distributed differently among the observables. While this study does not examine the mechanistic bases for memory, it presents a framework for quantifying memory in cellular behaviors and is thus a starting point for studying new questions about cellular regulation and evolutionary strategy.

**Key words:** memory; information theory; phenotypic diversification strategy; bacteria; game theory; *Bacillus subtilis*

---

<sup>1</sup> These authors contributed equally to this work.

# Introduction

Your average bacterium is unlikely to recite  $\pi$  to 15 places or compose a symphony. Yet evidence is mounting that these ‘simple’ cells contain complex control circuitry capable of generating multi-stable behaviors and other complex dynamics that have been conceptually linked to memory in other systems. And though few would call this phenomenon memory in the ‘human’ sense, it has long been known that bacterial cells that have experienced different environmental histories may respond differently to current conditions [1-3]. Though some of these history-dependent behavioral differences may be physically necessary consequences of the prior history, and thus some might argue insignificant, other behavioral differences may be controllable and therefore selectable and even fitness enhancing manifestations of memory.

In this paper we take the potentially controversial view that history-dependent behavior, whether short or long term, controlled or incidental, reflects a form of memory [4-6]. Because bacterial dynamics at every level of resolution operate within the limitations and potentials of nonlinear physical and biochemical dynamical systems, they must exhibit at least very short-term transient memory, and potentially longer term memory. The type of memory (and its significance) depends on which features of cell history are ‘remembered’, and at what resolution; whether or not the system eventually ‘forgets’ its past, and if so, how long this forgetting takes; the mechanisms in the cell responsible for memory storage, encoding, and retrieval; and whether or not this memory provides a fitness advantage in a natural environment. In cellular systems, environmental memory has been noted to be inherent in everything from the selective history of mutation, epigenetic inheritance via chromatin modification in neurons and DNA methylation in chemotaxing bacteria [7], genetic and epigenetic phase variation mechanisms controlling surface features of pathogenic bacteria [8,9], cellular proliferation and survival in the immune system, and in switch-like feedback systems in regulatory networks spanning signal transduction, metabolism and gene expression [10-21]. There is also a growing body of work focusing on synthetically designing and constructing network motifs and systems that are capable of showing some types of dynamic memory [22,23]. These and many other studies in synthetic and natural systems suggest that even the simplest first-order chemical reactions have at least transient memory of initial conditions, and more complex mechanisms involving history-dependent changes in the concentrations, states and localization of proteins and other regulatory network elements can encode a wide range of input information and store it for amounts of time ranging from minutes to days or longer [4,16,24,25]. The state dynamics of such systems contain the memory of past controlling inputs, and even of past environmental conditions if one is to interpret more broadly [5,26].

In metazoans, the ability of somatic cells to remember their fates is key to development and thus to organismal fitness. The same can be said for other types of metazoan cells like those found in the immune system that use a memory of past states to modify future behavior. In principle at least, memory, whether short- or long-term, can feasibly confer an evolutionary advantage in microbes as well. For instance, Hoffer *et. al.* suggest that in *E. coli* a form of ‘memory’ of past phosphate limitation leads to a faster response to successive periods of phosphate limitation, and that this faster response may be survival enhancing [5]. It has also been suggested that pathogenic bacteria use cross-talk encoded memory to balance the demands of immune avoidance with a sequential, compartment to compartment infection lifecycle [8,9]. More abstractly, the dynamic implementation of cellular behaviors can be viewed as a selected, ‘winning’ (or at least stable) strategy in an evolutionary game [12,27]. In game theory, information creates advantage [28-30], and information about the past as well as the present creates even greater advantage. Thus if bacterial cells are able to store information about past experience in some type of memory, and use this memory to modulate their behavior, this opens up the possibility of playing game strategies with memory, a provably superior family of strategies compared to those without memory [31-35]. Even if the memory capacity of the system is short term, but on the order of environmental fluctuations, it could conceivably impact fitness and therefore play a role in an evolved adaptive behaviour [28].

Given the potential ubiquity and significance of bacterial memory, we propose that quantifying history dependent behavior in microbes could be an important piece of the puzzle of bacterial regulation, survival strategy, and evolution. To this end, we developed an information-theory based conceptual framework for

thinking about and measuring both the persistence of memory in microbes and the amount of information about the past encoded in these dynamics. This method produces a phenomenological measure of cellular memory *without* regard to the specific cellular mechanisms encoding it. We then applied this framework to the bacterium *B. subtilis*. *B. subtilis* presents an excellent model organism for this study because of its exquisite sensitivity to environmental conditions, its known mechanisms of bistability and other hysteretic switch-like regulatory stress response mechanisms and architectures, and its developmental decision to sporulate that strongly resembles eukaryotic memory-associated processes determining developmental cell fate ([10,36-40], Fig. 1). Also, certain aspects of *B. subtilis* behavior, such as spore coat composition, have already been associated with environmental memory [41-43], and though much suggests that there should be memory, how these response dynamics depend on past conditions prior to application of a stress has not been systematically examined.

In our experiments, we quantified the ability of three *B. subtilis* stress response systems – sporulation, degradative enzyme synthesis, and growth - to ‘remember’ 10 distinct cell histories prior to application of a common stressor. We chose to observe commitment to sporulation (via reporter fusion to  $P_{SPOIIE}$ ) because the sporulation decision is bistable, and bistability is associated with memory [9,11,16,44]. We added the reporter for degradative enzyme synthesis (measured by a fluorescent reporter fused to the AprE promoter) because though it shares many common controllers with sporulation, its expression pattern is quite different and not believed to be bistable or probabilistic. We wondered whether any history-dependence in sporulation control would be mirrored in AprE control. Finally, we chose to observe growth (as measured by  $OD_{600}$ ) because it is perhaps the most accessible measure of cellular health and fitness and is an integrator of many other aspects of cell function, thus it may show interesting differences depending on cell history. One can imagine that there might be a strong fitness incentive toward memory in *B. subtilis*. If cells could use a memory of past conditions to ‘predict’ future conditions, and delay sporulation, an expensive process, if the environment is likely to improve or accelerate sporulation if the starvation period is likely to be long, they might improve their odds for long-term survival.

## Results

### ***Information Theoretic Memory Framework***

#### **‘Adaptive’ memory experiment**

A complete quantification of biologically relevant memory would involve first perturbing the cell with all possible sequences of complex environmental inputs it might experience in the wild in each of its growth modes, then measuring all cellular responses to these perturbations, and, finally, quantifying the degree and distribution of history-dependence in these responses.

Here we assume a simple approximation of this scenario, in which each sample of a biological system is subjected to one of many conditions prior to time  $t_0$ , and then observed in a common condition after  $t_0$  (see Fig. 2 and Definition (1) in Appendix S1 in Supplementary Information). We call this an ‘adaptive’ memory experiment because it roughly simulates a temporal shift in the environment requiring adaptation or acclimation, and to differentiate it from the more classical memory experiments in physics, engineering and cell biology designed to identify hysteretic loops [45-47]. While we do not identify such loops here, multistability is suggested by the appearance of long term memory in our experiments. More complex environmental history trajectories could feasibly unravel more memory effects.

We are interested in whether past conditions can be inferred from observations of behavior in current conditions. *The assumption here is that history-dependent behavior is a manifestation of memory, and that the better the possible inference about prior conditions from current measurements, the more memory there is within the system.*

## Adapting communication metrics to memory

To quantify this intuitive concept of history-dependence as memory, we use concepts from information theory [48] in the tradition of Landauer's use of informational entropy to estimate human memory capacity [49], and the extensive body of work characterizing memory in individual neurons [50-53].

By interpreting the random variable  $Y$  as behavior in current conditions, and the random variable  $M$  as past cellular history prior to time  $t_0$ , the mutual information  $I(M;Y) = H(M) - H(M|Y)$  of  $M$  relative to  $Y$  provides a measure of memory in informational entropy bits (see [48], Fig. 3, and Definition (2) in Appendix S1 for details, including the definition of informational entropy  $H$ ). Roughly speaking, from this perspective  $I(M;Y)$  captures how much uncertainty about past conditions can be reduced by observations of behavior in current conditions. Worded differently,  $I(M;Y)$  captures how much information about past conditions can be inferred from observations of behavior in current conditions. *The better the possible inference about prior conditions (and thus the higher the bit count of  $I(M;Y)$ ), the more memory there is within the system.*

## Short term vs. long term memory

Memory, or history-dependent behavior, can manifest across multiple time scales. Short term, or transient, memory is stored by the system for some time, and then 'forgotten' (see Fig. 4a,d). Systems may also have either 'effective' long term memory if the transient dynamics are long compared to environmental fluctuations, or 'true' asymptotic memory if the stationary state of the system depends on initial conditions, as occurs in nonlinear systems with multiple attractors (see Figs. 4b,c,and e). For an example of the latter, the state of a bistable switch encodes an asymptotic memory of the last switching event.

Because in many systems the significance, mechanistic origin, and function of memory likely depends on how long it lasts, and in particular whether it can be classified as short-term or long-term, we distinguish between the two types of memory and quantify them separately. From an information perspective, we say that an external observer of an adaptive memory experiment with *a priori* knowledge of the probability distribution over cell histories detects short-term memory in this system if observing measurements of some fraction of the short-term behaviour of the system after time  $t_0$  leads to a reduction in uncertainty about the history of the system prior to time  $t_0$ . In this case, we say that the cells exhibit  $I_{\text{trans}}(M;Y; t_{\text{trans}}) \equiv I(M;Y(t = t_0:t_0+ t_{\text{trans}}))$  bits of **short term memory** in the observable  $Y$  over the period from  $t_0$  to  $t_0+t_{\text{trans}}$ , where  $t_{\text{trans}}$  is a time before the signal approaches its steady state (Definition (3) in Appendix S1). Likewise, long-term memory is detected if observing measurements of the system behavior near an apparent steady state after time  $t_0$  leads to a reduction in uncertainty about the history of the system. Here we say the cells exhibit  $I_{\text{asym}}(M;Y) \equiv I(M;Y(t=t_0+t_{\text{asym}}:\infty))$  bits of **long term memory** in the observable response  $Y$  during the experiment, where  $t_{\text{asym}}$  is the time it takes for the signal to settle (Definition (4) in Appendix S1).

## Memory quantification normalized

The above metrics for short term and long term memory are absolute measures, in that they give a bit count for an answer. Though these absolute numbers can be useful, it is also useful to measure memory in relative terms, compared to the total amount of memory that *could* be observed in a perfectly retentive system given the

limitations of the experiments. To address this issue, we define **short-term memory fidelity** to be  $P_{\text{trans}}(M; Y; t_{\text{trans}}) \equiv I(M; Y(t = t_0:t_0+t_{\text{trans}}))/H(M)$  and **long-term memory fidelity** to be  $P_{\text{asym}}(M; Y) \equiv I(M; Y(t = t_0+ t_{\text{asym}} : \infty))/H(M)$ , where  $H(M)$  is the entropy over all the past conditions that were applied in the experiment. These normalized mutual information metrics, measures between 0 and 1 of the fraction of uncertainty about the past conditions tested that is reduced by knowledge of future cellular response, have also been called the coefficients of constraint [54] (see Definition (5) in Appendix S1).

## Quantifying memory in higher dimensions

In addition to analyzing each observable individually, we are interested in calculating the short and long term memory exhibited by the combined behavior of multiple observables. To do so, the above definitions are easily extended to the case of multiple observables by letting  $Y$  be a vector  $\bar{Y} = (Y_1, \dots, Y_n)$  and calculating  $I_{\text{asym}}(M; (Y_1, \dots, Y_n))$  and  $I_{\text{trans}}(M; (Y_1, \dots, Y_n); t_{\text{trans}})$  and the memory fidelity of each. This combined-memory estimation is interesting because it allows one to address the question of whether combining information from multiple read-outs leads to extra memory beyond what is present in any of the individual read-outs, and if so, how much. This issue is related to the size of the memory, and the dimension it occupies within a cell's state space.

An inequality governing the informational entropy of a vector pair of variables  $(X, Y)$  is as follows:  $\max(H(X), H(Y)) \leq H(X, Y) \leq H(X) + H(Y)$  [54]. Thus, we know that the memory exhibited by any pair of observables must be greater or equal to the bit count of the most retentive pathway of the pair, and less than or equal to the sum of the bit counts of the two pathways. If two pathways are controlled independently, their combined behavior could produce the upper limit on memory in the higher-dimensional space, whereas if the pathways are controlled by a common signal or if one pathway hierarchically controls the other, the lower limit might be realized. To quantify this concept, we define **memory orthogonality** between two pathway readouts  $Y1$  and  $Y2$  to be:  $Mem_{\text{orth}}(M; (Y1, Y2)) \equiv (I(M; (Y1, Y2)) - \max(I(M; Y1), I(M; Y2))) / \min(I(M; Y1), I(M; Y2))$ , where  $M$  is cell history and  $I$  is mutual information.  $Mem_{\text{orth}}$  equals 1 if the two variables combined as a vector yield the upper bound of memory, and 0 if the two variables in combination yield the lower bound (see Definition (6) in Appendix S1).

## Implementation

For the calculations above, listed more formally in Appendix S1 in Supplemental Information, we need to estimate probability distributions over the past cell histories being tested and the responses of the cells to each history. For past conditions/histories, we enforce a uniform probability of observation of each condition by running each experiment (condition  $i \Rightarrow$  response  $i$ ) a fixed number of times. For responses, we cluster trajectories from the different conditions and the probability of a response is simply the histogram of trajectories over clusters. The probability of prior environment given cluster membership is enumerated in a similar way. Details of the entire analysis algorithm can be found in Materials and Methods.

## Caveats

The above information-based metrics and simple associated analysis algorithm (see Materials and Methods) are useful in that they transform the 'lay' questions – "Do cells 'remember' past experiences and use these memories to modify future stress response dynamics?" and "If so, is this 'memory' short term or long term, and how much is there?" – into well-defined queries about information and uncertainty yielding quantitative estimates of microbial memory in informational entropy *bits*.

However, any attempt to quantify or qualify memory is fundamentally limited by the possibility of unobservable states (see Fig. 4c), uncontrolled and unobservable inputs, poor choice of input combinations and sequences, and measurement errors and distortions. Here we assume most such limitations, discussed in more detail in Supplementary Information (Section S1), are inherent in the estimation of memory processes and most likely to result in information loss and thus *underestimates* of the ability of the system to ‘remember’ the cell histories tested by the experimental compendium. Therefore we interpret quantifications of memory within our *B. subtilis* compendium as lower bound estimates.

## Experiment and Overview of Analysis

**Memory experiment on *B. subtilis*:** To test for history dependent behavior – ‘memory’ - in *B. subtilis*, we engineered a fluorescently labeled strain of *Bacillus subtilis* to report on commitment to sporulation and degradative enzyme synthesis: the KEE strain (*PspoIIIE-gfp*, *PaprE-dsred cmp*, see Materials and Methods for details on strain construction). The *spoIIIE* promoter (*PspoIIIE*), our sporulation reporter, controls expression of *spoIIIE*, a gene encoding a serine phosphatase specifically expressed upon commitment to sporulation and therefore considered a good sporulation commitment signal [55,56]. The *aprE* promoter (*PaprE*), our degradative enzyme synthesis reporter, controls expression of the extracellular protease subtilisin naturally produced by *B. subtilis* cells at the end of exponential growth [57].

With the KEE reporter strain, we used our framework to estimate, in informational entropy bits, the capacity of these stress response pathways and of the cell growth dynamics to ‘remember’ 10 distinct cell histories prior to application of a common stressor. Specifically, we first grew three replicate cultures in one of two media, Luria Broth medium (LB) or growth medium (GM) [58], to one of five different densities (all still in exponential growth, ranging from  $OD_{600} = [0.1:1]$ , see Table 1, where  $OD_{600}$  is the optical density of the culture at 600nm), for a total of ten cell histories. Thus in the first stage of the experiment, a clonal population of cells was divided into 30 groups, each of which experienced one of the 10 cell histories consisting of growth in one of two media to one of five cell densities over a fixed period of time (see Materials and Methods for details).

We chose to combine different media with growth to different densities as our set of cell histories because growth media can impact cell state, as can growth of cultures to different densities over a fixed period of time. Cells deplete nutrients and respond to the environment and its dynamics with changes in metabolic fluxes, post-translational modifications, gene expression, quorum signaling and synthesis of storage compounds. GM medium (also called CH medium) is a rich medium with casein hydrolysate as the sole carbon source [58]. LB medium is a much richer and more complex medium than GM and therefore sustains more rapid growth. We assumed that any resulting history-dependent differences in cell state at time  $t_0$  might lead to different history-dependent behaviors in the common medium after  $t_0$ .

After experiencing one of the 10 different cell histories, cells were then pelleted and resuspended at an intermediate density ( $OD_{600}=0.5$ ) in a common stress medium, in this case, sporulation salts starvation medium (SM) [58]. The resuspension time is denoted  $t_0$ . Thus, regardless of past experiences, all cells observed after  $t_0$  were subjected to starvation conditions starting at  $t_0$  in a fixed-density, fixed-size population.

Our three observables  $Y$  after  $t_0$  consisted of two fluorescent reporters, one for sporulation initiation and another for degradative enzyme synthesis (strain KEE (*PspoIIIE-gfp*, *PaprE-dsred cmp*)), and optical density of the culture as a proxy for cell growth ( $OD_{600}$ ), measured at the bulk population level every 15 minutes for 24 hours starting at time  $t_0$  (see Fig. 4 for time series, and Materials and Methods for details on strain construction and experiments). Thus, with 30 cultures – three for each of the 10 cell histories – and three observables per culture measured every 15 minutes for 24 hours in the common stress medium starting at  $t_0$ , the memory data compendium for this set of experiments consists of  $30 \times 3 \times 96 = 8,640$  measurements arranged in a 90 by 96 matrix.

**Data analysis overview:** The resulting memory data compendium was then analyzed for short- and long-term memory in each output signal individually and in all possible combinations of the three signals by applying the memory quantification algorithm described in detail in Materials and Methods and illustrated in the flow chart in Supplementary Information Section S2.

To briefly summarize, in order to estimate how much short-term and long-term memory was manifested in the behavior of the reporters, we sought to calculate the mutual information between the behavior of the cells after  $t_0$  and the history of the cells before  $t_0$ . This calculation required that we estimate the joint probability density between cellular behavior after  $t_0$  and cell history prior to  $t_0$ . Given constraints on the amount of data and other considerations described in detail in Section S3 of Supplementary Information, we took a clustering approach to this problem. That is, we first clustered the response of the pathway reporter as a way of dividing the trajectories into groups with common, distinct behaviors. The resulting assignment of each trajectory to a cluster was then used to calculate the frequency of co-occurrence of each behavioral class and each possible cell history. From this histogram we estimated the requisite joint probability distribution, which was then used to calculate the mutual information between cell history and the behavior of the observable, and thus arrive at an estimate for memory.

We performed this procedure on the 30 trajectories (3 replicates for each of the 10 cell histories tested) of each of the three observables, using both the short term (first 11 hours of measurements, during which the signal was still dynamically varying - see Materials and Methods for more details on our choice of analysis intervals) and long-term response (last three hours of measurements, from 21 to 24 hours, by which time the signals have remained flat for several hours) in order to estimate short-term and long-term memories manifested in each individual signal. To calculate the short-term and long-term memory in the combined activities of multiple signals, we took the same approach, with the one difference being that the clustering step captured the combined behavior of multiple readouts (Step 3 in the algorithm in Materials and Methods). All bit counts were then normalized to calculate memory fidelities and orthogonalities, as defined in Appendix S1, in order to estimate in relative terms how much of the total possible memory each system ‘remembers’, and how much ‘extra’ memory is embedded in the higher-dimensional spaces formed by multiple pathways.

Since the 30 populations were subjected to 10 different (within error) past conditions  $M=(Medium1, Density1)$  in equal proportions, the informational entropy of the cell history space  $M$  is  $H(M) = -\log_2(1/10) = 3.3219$  bits. Thus, without prior knowledge there are 3.3219 bits of information about cell history at most that can be recovered from observation of these three outputs, either individually or in combination and on any time scale.

## Experimental Results

### A qualitative overview of history-dependence

The *B. subtilis* stress responses measured by the three observables (Figure 5) appear neither memoryless nor in possession of a perfect memory of the cell histories tested. They do not appear to be memoryless because not all signals from a given observable follow a common trajectory (within noise bounds) irrespective of past history of the cells. Nor does the memory of any observable appear to be perfect, because though there are ten distinct cell histories prior to time  $t_0$ , there appear to be fewer than ten distinct dynamics per observable in response to the starvation stressor administered at time  $t_0$ . By eye, there appear to be more distinct behaviors in the short term than in the long term. Also, different cell histories group together for different observables. This means that we expect a higher bit count estimate of short term memory than long term memory, and different amounts of memory and of different aspects of cell history in the three pathway observables.

## All observables exhibit short-term memory of cell history, with sporulation exhibiting the most and growth dynamics the least

The transient behavior (first 11 hours) of the SpoIIIE (sporulation) reporter clusters into five distinct classes of behavior (different onset times and sigmoidal vs. more pulsatile expression), whereas the transient behavior of the AprE (degradative enzyme synthesis) reporter clusters into three classes (different onset times and different expression levels) and the growth reporter into just two classes (some vs. almost no growth) (see left panels of Fig. 6a,b,c). The mutual information between the resulting clustering vectors and the cell history vector captures how well the different behavioral classes of each observable correspond to different cell histories. Performing this calculation, we estimate  $I_{\text{trans(spo)}} = 1.96 \text{ bits}$  of short-term memory in the sporulation reporter;  $I_{\text{trans(AprE)}} = 1.4855 \text{ bits}$  of short-term memory in the degradative enzyme synthesis reporter, and  $I_{\text{trans(OD)}} = 1 \text{ bit}$  of short-term memory in the growth dynamics reporter OD<sub>600</sub>. Thus, all three observables exhibit short-term memory of the cell histories tested, with the sporulation reporter exhibiting the most memory and growth dynamics the least.

Dividing these absolute bit counts by the entropy of the cell history space, we estimate the short-term memory fidelities of sporulation initiation, degradative enzyme synthesis, and growth dynamics to be  $P_{\text{trans(spo)}} = I_{\text{trans(spo)}}/H(M) = 1.96/3.3219 = 0.59$ ,  $P_{\text{trans(AprE)}} = I_{\text{trans(AprE)}}/H(M) = 1.48/3.3219 = 0.45$ , and  $P_{\text{trans(OD)}} = I_{\text{trans(OD)}}/H(M) = 1/3.32 \approx 0.3$ , respectively. This means that if one were to observe all 30 short-term responses of one of the three reporters after  $t_0$  but not told which history corresponds to which trajectory, 59% of the uncertainty about cell history prior to time  $t_0$  could be reduced by observation of the transient sporulation reporter dynamics after time  $t_0$ , 45% of this uncertainty about the past could be reduced by observation of the degradative enzyme synthesis reporter dynamics after  $t_0$ , and only 30% of this uncertainty could be reduced by observation of the growth dynamics after  $t_0$ . More intuitively, one could say that 59%, 45% and 30% of the cell histories tested are ‘remembered’ by the short-term dynamics of the sporulation, degradative enzyme synthesis, and growth reporters, respectively (see Fig. 7 and Table S1).

## All observables exhibit long-term memory of cell history, though at a lower bit count than short-term memory

Though short term memory can be important—because even short term behavioral differences may have fitness consequences [59], especially if they are on the order of environmental fluctuations [28,60]—long term memory is generally the first thing that comes to mind when memory is discussed [61-64]. One might expect long term memory in *B. subtilis* stress responses - sporulation control especially - because of the feedback topologies in their regulatory circuitry and reportedly bistable behaviors [10,36-39].

To estimate the long term memory in each individual pathway we first clustered the final segment of the 30 time series of each reporter (from 21 to 24 hours after  $t_0$ ) to estimate the number of distinct long-term behaviors for each of the three pathway reporters (results = 2 unequal-sized clusters for each reporter, as shown in Fig. 6, though the cluster sizes and associated cell histories differ across reporters). We then calculated the mutual information between the clustering results and the cell history vector to arrive at lower bound estimates of  $I_{\text{asym(spo)}} = 0.8813 \text{ bits}$ ,  $I_{\text{asym(AprE)}} = 0.72 \text{ bits}$ , and  $I_{\text{asym(OD)}} = 0.97 \text{ bits}$  of long-term memory in the networks controlling sporulation initiation, AprE synthesis and growth dynamics, respectively. Thus, like a switch, there appear to be two, stable, long term behaviors for each pathway reporter, though the probability of converging to each is not equal or the same across reporters, as is reflected by distinct bit counts of less than 1 (if half the past histories lead to one attractor and the other half of the histories lead to the other, there would be 1 bit of asymptotic memory).

Dividing these absolute numbers by the entropy of the cell history space, we estimate the long-term memory fidelities of sporulation initiation, degradative enzyme synthesis, and growth to be  $P_{\text{asym(spo)}} = I_{\text{asym(spo)}}/H(M) = 0.8813/3.3219 = 0.265$ ,  $P_{\text{asym(AprE)}} = 0.22$ , and  $P_{\text{asym(OD)}} = 0.29$ , respectively. Thus, approximately 25% of the uncertainty about cell history prior to the onset of starvation is reduced by knowledge of any one of the three long-term reporter dynamics in the starvation environment. To summarize, all three observables exhibit around 1 bit of long-term memory of the histories tested, though of different aspects of cell history as will be shown below. One bit is a significant amount but much less than the nearly 2 bits of memory seen in the most retentive short-term response.

## Different observables remember different aspects of cell history to different degrees

The above memory estimates are in a sense high-level, because each of the 10 distinct cell histories is treated identically. By drilling down a level of resolution to the component parts of the cell histories – initial nutrient composition of the media and cell density reached in that media (which can also feasibly affect both the nutritional composition of the medium and cell state while in log phase) – we can investigate which aspects of cell history are remembered by the observables and for how long.

In the short term, all three observables have a perfect memory of whether they were grown in LB or GM, and only a partial memory of their density in this medium. Put more formally, if we consider growth medium in isolation and calculate the mutual information between growth medium prior to time  $t_0$  and transient response of the three reporters to starvation after time  $t_0$ , we see that a history of growth in LB can be distinguished from a history of growth in GM with 100% memory fidelity ( $P_{\text{trans(Spo)}}(\text{Medium}I; Y; t_{\text{trans}}=11 \text{ hrs}) = P_{\text{trans(AprE)}}(\text{Medium}I; Y; t_{\text{trans}}=11 \text{ hrs}) = P_{\text{trans(OD)}}(\text{Medium}I; Y; t_{\text{trans}}=11 \text{ hrs}) = 1$ , where  $\text{Medium}I$  is a random variable representing growth medium prior to time  $t_0$ , and can take on the values GM or LB). In contrast, the ability of the pathways to remember the population density reached prior to  $t_0$  (and any changes in cell state these differences in cell density create) is less simple. With a history of growth in GM, the cell density prior to the onset of starvation at  $t_0$  is not ‘remembered’ by the short-term *B. subtilis* sporulation, degradative enzyme synthesis, or growth dynamics responses, even transiently (0% memory fidelity), as all responses are indistinguishable within noise ( $I_{\text{trans}}(\text{Density}I/\text{Medium}I = \text{GM}; Y; t_{\text{trans}}=11 \text{ hrs}) = 0$ ). However, when grown in LB, the cell density prior to  $t_0$  is remembered with 80% memory fidelity by the transient sporulation dynamics and with 60% memory fidelity by the transient AprE dynamics ( $P_{\text{trans(Spo)}}(\text{Density}I/\text{Medium}I = \text{LB}; Y; t_{\text{trans}}=11 \text{ hrs}) = 0.8$ ;  $P_{\text{trans(AprE)}}(\text{Density}I/\text{Medium}I = \text{LB}; Y; t_{\text{trans}}=11 \text{ hrs}) = 0.6$ ).

In the long term, all three observables have only a partial memory of which medium they were grown in, and to what density. Like in the transient memory case, past growth medium is remembered better than past cell density, but unlike in the transient memory case, there is no perfectly clean dividing line separating out the long-term responses to the two growth media histories. For example, given observations of the long-term behavior of the sporulation reporter, a history of growth in LB can be distinguished from a history of growth in GM with only 39% memory fidelity ( $I_{\text{asym(spo)}}(\text{Medium}I; Y)/H(\text{Medium}I) = 0.39$ ), whereas cell densities (grouped into five classes, (D,-1,-2,-3 and -4)) prior to  $t_0$  are remembered even less well, with only 12.1% memory fidelity ( $I_{\text{asym(spo)}}(\text{Density}I; Y)/H(\text{Density}I) = 0.121$ ). A similar pattern can be seen in the long-term memories of the other two reporters. Interestingly, though each reporter exhibits two possible long-term behaviors, the clusters are different sizes and the histories that correspond to each behavioral cluster are different for different pathways. As will be shown in the next section, these differences lead to the possibility of an increased memory capacity in the higher dimensional space defined by the combined activities of multiple pathways.

## There is more long-term memory in the combined activity of the observables than is present in any individual observable

Interestingly, analysis of the *transient memory* of the pairs of pathway readouts (Spo, AprE), (Spo, OD<sub>600</sub>), and (AprE, OD<sub>600</sub>) shows *no increase in memory in the higher dimensional space* than is found in the most retentive pathway in the dyad (see Figure 7). For example, we estimate the transient memory found in the pair (AprE, OD<sub>600</sub>) to be 1.4855 bits, which is the same bit count found in AprE alone ( $Mem_{orth(trans)}(M;(AprE, OD_{600})) = (1.4855-1.4855)/(1.4855) = 0$ ). Likewise, the three-dimensional readout (Spo, AprE, OD<sub>600</sub>) shows no more transient memory than is found in the sporulation pathway (1.96 bits), its most retentive member.

However, the same conclusion does not follow for asymptotic memory. *Every pair of pathway readouts contains more asymptotic memory than either constituent signal, and the triple pathway readout contains more asymptotic memory (at 1.57 bits) than any of the constituent pairs* (see Figure 7). This implies that the long term behavior of our three observables occupies a relatively high dimensional space, with each subsystem responding differently to aspects of past conditions. For example, though the AprE pathway is estimated to have only 0.7219 bits of asymptotic memory and the growth measure OD<sub>600</sub> has only 0.971 bits of asymptotic memory, the pair (AprE, OD<sub>600</sub>) has 1.371 bits of asymptotic memory ( $Mem_{orth(asym)}(M;(AprE, OD_{600})) = (1.371-0.971)/(0.7219)=0.554$ , or 55.4% of the maximum). Put more concretely, the asymptotic behavior of the AprE signal alone ‘remembers’ two classes of cell history: the first a history of growth in rich medium to higher densities and the second all other histories in the compendium. Whereas observations of the asymptotic behavior of the growth signal allow distinction between two *different* classes of cell history; the first growth in rich medium to all densities greater than the lowest tested (-4), and the second all other histories in the compendium. Viewed together as a combined vector in a higher dimensional space, the asymptotes of the pair (AprE, OD<sub>600</sub>) permit distinction between three classes of cell history: growth in rich medium to higher cell densities, growth in rich medium to low (but not lowest) and intermediate cell densities, and, finally, growth in rich medium to the lowest density or growth in poorer medium to any density. Adding the sporulation signal increases the information storage yet again, by adding another discernable class, leading to a total long-term combinatorial storage of 1.57 bits. Thus, because the different cellular systems in *B. subtilis* remember different aspects of prior history, the combined activity of multiple pathways is able to combinatorially store more information about the past than can any individual pathway. However, the total asymptotic memory is still somewhat less than the total transient memory (1.57 vs. 1.96 bits). (For a complete accounting of cell history memory over all signal combinations, and for the mutual information between all pairs of signals, including the transient and asymptotic responses of each signal, see Figure 7 and Table S1.)

## Discussion

Though evidence that bacterial cells are able to remember their histories and use these memories to alter their behavior in a fitness enhancing manner would not raise expectations that bacteria could recite  $\pi$  or write music, it *would* enrich the motifs-modules-games view of bacterial regulation [12] by adding game strategies *with memory* to the repertoire of microbes. This exploratory paper does not provide evidence that *B. subtilis*, or any other microbe, is intelligent or is playing an evolved, fitness-enhancing memory strategy. Rather, in this work we propose that the familiar phenomenon of history-dependent behavior in microbes reflects a form of memory worth studying systematically and quantifying, and that doing so sets the foundation for understanding both the mechanisms and function of memory in cell behavior and fitness. To this end we formulated a conceptual information-theory based framework for measuring microbial memory, thereby introducing tools that begin to observe and quantify the relationship between past cell history and future cell behavior from a new angle. This method produces a phenomenological measure of cellular memory without regard to the specific cellular mechanisms encoding it.

We then applied these tools to a simple set of medium-shift experiments on *B. subtilis*, in the process demonstrating that *B. subtilis* does ‘remember’, both in the short and long term, aspects of its cell history, and that this memory is distributed differently among the observables. More short term than long term memory was evident, with short-term sporulation dynamics exhibiting the most memory and long-term degradative enzyme AprE synthesis dynamics the least. As expected, some but not all of the history-dependence between the sporulation and AprE reporters is shared (AprE has 75% of the short-term and 80% of the long-term memory estimated for sporulation). We also illustrated how to quantify memory in multiple combined variables, in the process showing that because the different cellular systems in *B. subtilis* remember different aspects of prior history to different degrees, the combined activity of multiple pathways is able to combinatorially store more information about the past than can any individual pathway. Of the two components of cell history varied in our compendium – past growth medium and the cell density reached in this medium, which can alter cell state even in log phase – growth medium appeared to be better remembered by *B. subtilis*, with past density remembered best when originally grown in the medium richest with nutrients, LB. Admittedly we do not yet know whether the memory we have observed is fitness enhancing and evolved or just incidental, or what molecular mechanisms or artifacts are responsible for the observed pattern of memory storage. Rather, these simple experiments and the surrounding analysis and framework demonstrate what could be the beginning of a larger memory program, and indicate that memory in cellular behaviors may be a rich area for further exploration.

## Ideas for a more complete memory-in-microbes research program

A more complete program for investigating memory in bacteria would encompass at least three lines of inquiry, essentially the ‘what’, ‘how’, and ‘why’ of bacterial memory. The first line of inquiry (what), for which this study is an example, is the quantification of environmental memory in a microbe. This study could be extended by resolving the population-averaged behavior analyzed in this paper into single-cell measurements and memory classification and quantification. Given that sporulation is thought to be a stochastically triggered bistable developmental process [10,36-39], one might expect the population-averaged measurements (Figure 5.b) to resolve into bimodal distributions of high and low GFP-expressing cells. And since AprE synthesis control is believed to be more deterministic and analog, one might expect more monomodal distributions. Preliminary data from flow cytometry analysis support this expectation, at least for some histories and time points (see Figure S1 in Supplemental Information), but further work is needed to determine for what conditions and pathways memory at the single-cell level can be classified as stochastic, and the form and quantification of this stochasticity. An exploration of the memory characteristics of other cellular players active in these and interacting networks, and the space of their environmental sensitivity, with the goal of estimating the ‘true’ memory capacity of the system, are other possible extensions of this work.

A second line of inquiry (how) would build upon the first by elucidating the causal basis for any observed environmental memory. Though many genetic and epigenetic bacterial switching mechanisms have been elucidated [8,10,16], still unclear is exactly how different types of environmental and intercellular signals might be encoded and remembered within cellular circuitry for varying lengths of time, a question addressable through mutant studies and modeling. On the ‘meta’ level one could ask whether memory is stored within single cells, population distributions, or in the larger state space defined by the cell-environment interaction through distributions of nutrients, waste products, enzymes, signaling molecules, biofilm generating conditions, and so on. A third line of inquiry could focus on the ‘why’s’ of environmental memory. Is environmental memory, if it exists, controlled or incidental: evolutionarily advantageous, deleterious, or neutral? Is there evidence that memory-modulation of phenotype expression control does not provide a fitness advantage in the *present* but rather in a *future* implicitly anticipated from past experiences, thus implying an internal model of environmental dynamics (in analogy to the internal model principle in control [65])? We suspect that answers to these ‘why’ questions could be key to whether the others are worth deeply pursuing.

## What do the *B. subtilis* memory observations in this case study mean?

Though we do not yet know whether the memory we have observed is fitness enhancing and evolved, or merely incidental, we can speculate. Looking qualitatively at the three behavioral observables together, we notice that when cells are grown to low density in the less rich GM medium prior to the onset of starvation conditions, they on average grow very fast after resuspension in starvation media, and after a brief lag start turning on their degradative enzyme synthesis and their probabilistic sporulation machinery, even as the population continues to grow. Whereas when cells are grown in richer, nutrient filled LB medium to the same low density prior to the onset of starvation conditions, they take a quite different approach. In this case, cells seem to adopt a wait-and-see strategy, forgoing growth and delaying sporulation and AprE synthesis for many hours.

**A game strategy with memory?:** The most tempting speculation is that *B. subtilis* is playing a memory strategy in an evolutionary game. From a game perspective, one could take these observations as a sign that after transitioning from a less rich medium to starvation, *B. subtilis* uses its memory of past nutrient-limited growth in the context of an implicit internal model of environmental dynamics to ‘predict’ how long starvation conditions will last. If the cells expect starvation to last a long time, a rational course of action might be to create as many spores as possible, as fast as possible, to maximize the spore count that will lie dormant until the next period of nutritional plenty. On the other hand, if growth in a rich environment prior to starvation in the context of this internal model produces a prediction of a short period of starvation, the rational action might be to *delay* sporulation, thereby decreasing the chances of having committed irreversibly to an unnecessary, costly 8 hour developmental program during which conditions could improve and the cells could be growing. Viewed in this way, *B. subtilis*’s cell-history dependent behavior might constitute an evolved probabilistic memory strategy in its game of survival. Such a strategy would trump diversification strategies without memory [28-30,35,66,67], and be analogous to adaptive model-based bet hedging over a diversified portfolio in the stock market [28].

If the above scenario is true, one would expect sporulation initiation *delay* to be a likely feature of the sporulation regulation strategy of *B. subtilis* to exhibit memory. Within our experimental compendium, the delay in turning on the sporulation machinery, as estimated by the amount of time it takes for GFP to start being noticeably expressed from the SpoIIE promoter by a population (normalized GFP intensity > 0.035, after which GFP rapidly increases), ranges from a relatively short 1.5 hours to a much longer nearly 8 or more hours after the onset of starvation (Figs. 5b and S2 in Supplemental Information). Calculating the mutual information between GFP expression delays and cell history, we see that most (86%) of the short-term memory in the sporulation reporter can be recapitulated by reducing the trajectories to this single number ( $I(M; \text{Initiation Delay})/I_{\text{trans}}(M; Y; t_{\text{trans}}=11\text{hrs}) = 1.685/1.96 \approx 0.86$ ). This calculation does not prove that the history-dependence we have observed is an evolved and fitness enhancing memory strategy in a game, but it is suggestive.

**...or an artifact of metabolism?:** Then again, the explanation could have little to do with evolutionary games. It could be that differences in metabolic stores, housekeeping apparatus, or metabolic state induced by the different media and different biomass of the culture simply represent initial conditions from which entry into sporulation and other stress responses is more or less easy [1]. For example there might be more ribosomes after growth in LB than there are after growth in GM, forcing cells coming from the latter to stop growth and initiate sporulation sooner. Or it could be that growth in GM, a medium that while not nutrient-limited is lacking the excess of simple carbon and nitrogen sources and readily available amino acids found in LB, activates metabolic pathways that can facilitate growth and spore formation in stress conditions. Then, when transferred to starvation conditions, cells might be able to use this metabolic machinery (and perhaps some form of intracellular nutrient storage) to scavenge whatever scarce nutrients are to be found in the new medium in order to grow and turn on their sporulation and degradative enzyme pathways nearly immediately. Whereas with a history of growth in rich, complex LB medium, cells might enter starvation conditions of SM without enzymatic or other reserves necessary for a near-immediate response to severely limited conditions, and thus require a delay while the cells construct the necessary metabolic machinery to acclimate to their environment.

These possibilities are not mutually exclusive; history-dependent behaviors could stem from some combination of evolved diversification game strategy and artifactual adaptive metabolic processes. Experiments comparing the fitness of wildtype bacteria to mutants with disrupted memory mechanisms coupled to a game theoretic analysis will be necessary to distinguish among the possibilities, and would identify the mechanistic source of memory behaviors in the process. In any case, we hope that this conceptual framework and analytical approach to quantifying memory in cellular behaviors will be a useful point of departure for studying a new set of questions about cellular regulation and evolutionary strategy in microbes.

## Materials and Methods

**Strains and culture media.** *Bacillus subtilis* 168 *trpC* (Bacillus Genetic Stock center) was used as the wild-type strain. *Escherichia coli* strain DH5 $\alpha$  was used for all plasmid amplifications and isolations. *Escherichia coli* was grown at 37°C in LB supplemented, when necessary, with ampicillin at a final concentration of 100  $\mu$ g/ml. *B. subtilis* was cultured at 37°C in either LB, growth medium (GM) or sporulation medium (SM). GM and SM media are commonly used in the ‘induction of sporulation by resuspension protocol’ described by Harwood and Cutting [58] and were supplemented with 50 $\mu$ g/ml and 20  $\mu$ g/ml L-tryptophan respectively. Antibiotics were added, with the following final concentrations: chloramphenicol, 5  $\mu$ g/ml; spectinomycin, 100  $\mu$ g/ml.

**DNA isolation and manipulation.** Total genomic DNA from *B. subtilis* 168 was isolated with DNeasy Blood & Tissue Kit (Qiagen) following manufacturer’s protocol for Gram positive bacteria. Plasmid DNA was extracted from *E. coli* with the QIAprep kit (Qiagen). DNA restriction and cloning were performed according to standard procedures [68]. Restriction enzymes and T4 DNA ligase were obtained from New England BioLabs and used according to the manufacturer’s instructions. DNA fragments were purified from agarose gels with the QIAquick gel purification kit (Qiagen). *Vent* DNA polymerase (New England Biolabs) was used for PCRs.

***B. subtilis* reporter strain construction.** Strains and plasmids are listed in Table S2 in Supplemental Information. To integrate the fluorescent reporter fusions in the *B. subtilis* genome the pLFKEE integration vector was constructed as followed. The GFP variant GFPmut2 [69] was excised from pMF19 [70] by digestion with *Bam*HI/*Eco*RI enzymes and ligated into pEA18 (a gift from Antje Hofmeister) digested with the same enzymes, to give pLF22. The plasmid pEA18 (*cmp*, *spc*) is a vector [71] allowing integration by double cross-over at the *amyE* locus, with a chloramphenicol selection. The *spoIIIE* promoter ( $P_{spoIIIE}$ ) was amplified by PCR from *B. subtilis* 168 genomic DNA using primers PspoIIIE-D/*Eco*RI (atcacggaattcaaatcggtttctcttcgagaagccg) and PspoIIIE-M-R/*Hind*III (atacaaagcttttatctgtgctgtcattatagcg), and digested with *Eco*RI and *Hind*III, then ligated 5’ of *gfpmut2* on pLF22 that had been digested with the same enzymes to give pLF25 ( $P_{spoIIIE}$ -*gfp*, *cmp*). The transcriptional profile of the *spoIIIE* gene was verified by total RNA dot blot before and after induction of sporulation to confirm its early and specific expression induction at the onset of sporulation (see Figure S3).

To obtain the  $P_{aprE}$ -*dsred* fusion, the *dsredexpress* coding sequence was amplified by PCR from pDsRed-Express (Clontech) using primers DsRed-D/*Fse*I (tacggccggcctaaggaggaactacaatggcgagcagtgaggacatcatcaagg) and DsRed-X/*Eco*RV (agatcatgatcagatctacaggaacaggtggtggcg). The PCR fragment obtained was digested with *Fse*I and *Eco*RV. A modified version of the *aprE* promoter ( $P_{aprE}$ ) (developed and tested in [72]) was amplified by PCR from pSG-TTGACA [72] using primers PaprESG-D/*Age*I (tgaaccggtgtcaaacatgagaattcagcg) and PaprE-R/*Fse*I (caaggccggccaaattcagagtagactactaaaagac). The resulting PCR fragment was digested with *Age*I and *Fse*I and ligated with *Fse*I/*Eco*RV-digested *dsredexpress* into *Age*I/*Eco*RV-digested pLF25 in a three-point ligation to give pLFKEE ( $P_{spoIIIE}$ -*gfp*,  $P_{aprE}$ -*dsred*, *cmp* *spc*). Selection of plasmid constructions in *E. coli* clones was done by adding ampicillin as described above and correct fusions were verified by sequencing.

To construct *B. subtilis* KEE, pLFKEE was transformed into *B. subtilis* 168 competent cells as previously described [58] and selected on LB solid medium supplemented with chloramphenicol. Integration clones were screened for their *amyE* phenotype on LB + 1% starch solid medium [58]. The inability of the clones obtained to grow on spectinomycin was checked to eliminate single cross-over plasmid integration events. Correct integration of the fusion at the *amyE* locus was verified by PCR analysis.

**Medium-shift experimental protocol.** Before each experiment, cells were streaked from  $-80^{\circ}\text{C}$  glycerol stocks on LB plates with chloramphenicol and grown overnight. One colony was picked and inoculated in 5 ml liquid LB medium with chloramphenicol in a series of dilution tubes and grown overnight at  $37^{\circ}\text{C}$ . The culture the closest to  $\text{OD}_{600}$  of 1.0 was used to inoculate 60 ml of LB or GM in 250-ml flasks to a final  $\text{OD}_{600}$  of 0.05 (flask D) after elimination of the culture medium by centrifugation of the cells ( $6,000 \times g$ , 3 min). The culture was split in two, and successive dilutions of 1:2 were performed to a total of 5 flasks of 30 ml culture (flask D and dilution flasks: -1, -2, -3, -4). Cells in all four flasks were grown simultaneously at  $37^{\circ}\text{C}$ , 200 rpm, until the most concentrated culture grew to an  $\text{OD}_{600}$  of 1.0 (Flask D). Then, 25 ml of each culture were harvested by centrifugation ( $8,000 \times g$ , 5 min) and resuspended in a pre-warmed SM medium volume calculated to obtain a final OD of 0.5 (medium density). Three aliquots of 200  $\mu\text{l}$  from each flask were transferred to a sterile Costar 96-well black plate with flat clear bottom (Corning). Cells in the plate were grown in a Tecan Safire microplate spectrophotometer at  $37^{\circ}\text{C}$  medium linear shaking setting (395 rpm). Culture turbidity ( $\text{OD}_{600}$ ) and fluorescence were measured at 15 minutes intervals for a total time of 24 hours. GFPmut2 was read at wavelengths of 481 nm (excitation) and 507 nm (emission), and DsRedexpress was read at 557 nm (excitation) and 579 nm (emission).

**Memory and mutual information analysis.** There are a number of ways to translate the memory quantification definitions in Appendix S1 into an analysis algorithm. We took a simple fixed-interval, clustering-based approach executed as a five-step algorithm implemented the MATLAB® (<http://www.mathworks.com/>) analysis environment, as follows (see Supplemental Information Section S2 for schematic):

**(step 0 – select time intervals):** The first step in analyzing the data is to select time intervals to analyze. We parsed the time series data (30 trajectories measured over 24 hours for each of three observables) into a ‘short-term’ set taken well before steady-state is reached (first 11 hours after  $t_0$ , the onset of starvation – though we could have taken any endpoint between 8 and 19 hours and obtained the same result (see panel (b) in Section S3)) and an ‘long-term’ set. For our purposes, we take as our ‘proxy’ for long-term, asymptotic behavior the last three hours of our measurements, from 21 to 24 hours after  $t_0$ , because by then all signals have remained flat for several hours. Experiments run for longer periods of time indicate that these signals remain flat for as long as we have measured them (36 hours, data not shown). However, we view this long-term data set as only a proxy for asymptotic behavior because though these signals remain constant for at least 36 hours, cells are forming spores and might be physiologically changing in other respects during this period and beyond.

**(step 1 – cluster data):** We used the Matlab script in S2.2 to hierarchically cluster the 30 short-term and 30 long-term trajectories of each observable (10 cell histories  $\times$  3 replicates) and to select ‘optimal’ clustering partitions for each. The assumption here is that the behavior of the observable (e.g., GFP intensity) falls into distinct classes, for example, increasing or decreasing. This script a) constructs a Euclidean distance matrix with the Matlab function `pdist.m`, b) constructs dendrograms using ward and average linkage with the function `dendrogram.m`, c) performs silhouette analysis on all tree cuts of both trees from (b) with the Matlab function `silhouette.m` [73], and d) ‘optimizes’ data clustering by selecting the partition that maximizes the mean silhouette, a measure of the compactness and separation of the clusters in the partition [73]. This step produced six  $30 \times 1$  cluster vectors, one short-term and one long-term cluster vector for each of the three observables (i.e., `ClustSPO_short`, `ClustSPO_long`, `ClustAprE_short`, `ClustAprE_long`, `ClustOD_short`, `ClustOD_long`).

**(step 2 – estimate memory):** Next we estimated the short-term and long-term memory in bits of each individual observable with the Matlab program `Entropy_MutualInfo.m` in S2.1. This program accepts two input vectors, A and B, and calculates from them individual informational entropies  $H(A)$  and  $H(B)$ , the entropy of the pair  $H(A,B)$ , and the mutual information between A and B,  $I(A;B) = H(A)+H(B)-H(A,B)$ .  $H(X)$  is defined in Supplementary Information (Appendix S1), and  $H(X,Y)$  is calculated by first calculating the joint probability distribution over  $(X,Y)$  and then calculating the entropy  $H$

over this joint distribution. Thus, memory is estimated to be the mutual information between cell history and cell behavior and calculated by calling `Entropy_MutualInfo.m` with input vectors  $A=M=[1\ 1\ 1\ 2\ 2\ 2\ \dots\ 10\ 10\ 10]$ , the cell history vector, and  $B$  equal to one of the six cluster vectors from step 1. To calculate memory fidelities, we normalized these memory estimates by dividing by  $H(M)=3.32$ , the entropy of the cell history space.

**(step 3 – estimate memory in higher dimensions):** The third step of the algorithm is to estimate the short and long-term

memory exhibited by the combined activities of pairs of observables and by the triple of observables. To do this, we first used the script in S2.3 to combine cluster vectors from multiple read-outs. This script takes as its input two cluster vectors `Clust1` and `Clust2` and outputs a combined cluster vector `Clust3` (e.g., if `Clust1=ClustSpo_short`; and `Clust2=ClustAprE_short`; then the output `Clust3` is a vector capturing all combined short-term behaviors of `Spo` and `AprE`, for example  $(Spo, AprE)=(\text{increasing, decreasing}), (\text{increasing, increasing})$  or  $(\text{decreasing, decreasing})$ ). Next, by calling `Entropy_MutualInfo.m` with inputs  $A=(\text{the cell history vector } M)$ , and  $B=(\text{the combined cluster vector } Clust3)$ , we calculate the mutual information between cell history and cell behavior, and thus the memory exhibited by the combined activity of the vector of observables contributing to `Clust3`. After computing short- and long-term memory for all four possible vector combinations of the observables, these estimates were divided by  $H(M)$  to estimate memory fidelities and normalized according to Definition (6) in Methods to estimate memory orthogonalities. Finally, we (step 4) calculated the mutual information between all pairs of observables using the cluster vectors from (step 1) as inputs to `Entropy_MutualInfo.m`.

We took this fixed-interval, clustering-based approach because of our desire to focus on how different cell histories can lead to qualitatively different stress response behaviors, and because a much larger data set would be required to use algorithms such as that suggested by Swinney to estimate mutual information at measurement intervals short enough to avoid excessive ‘blurring’ of the time series dynamics [74,75]. See Section S3 in Supplemental Information for a detailed discussion of alternative approaches and why we chose the one we did, and Section S2 for Matlab scripts and programs, including a note on a bootstrap method for calculating confidence intervals that one could apply to data sets with a sufficient number of replicates (not present in this data set).

#### **Acknowledgement:**

We thank our undergraduate laboratory assistant Tina Tipawan Mahatdejkul, Antje Hofmeister for strains and vectors, and the Arkin and Keasling labs. This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

#### **Bibliography**

1. Robinson TP, Ocio MJ, Kaloti A, Mackey BM (1998) The effect of the growth environment on the lag phase of *Listeria monocytogenes*. *Int J Food Microbiol* 44: 83-92.
2. Gawande PV, Griffiths MW (2005) Growth history influences starvation-induced expression of *uspA*, *grpE*, and *rpoS* and subsequent cryotolerance in *Escherichia coli* O157:H7. *J Food Prot* 68: 1154-1158.
3. Rozen Y, Belkin S (2001) Survival of enteric bacteria in seawater. *FEMS Microbiol Rev* 25: 513-529.
4. Casadesus J, D'Ari R (2002) Memory in bacteria and phage. *Bioessays* 24: 512-518.
5. Hoffer SM, Westerhoff HV, Hellingwerf KJ, Postma PW, Tommassen J (2001) Autoamplification of a twocomponent regulatory system results in "learning" behavior. *J Bacteriol* 183: 4914-4917.
6. Hellingwerf KJ (2005) Bacterial observations: a rudimentary form of intelligence? *Trends Microbiol* 13: 152-158.
7. Koshland DE, Jr. (1977) A response regulator model in a simple sensory system. *Science* 196: 1055-1063.
8. Hallet B (2001) Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. *Curr Opin Microbiol* 4: 570-581.
9. Holden NJ, Gally DL (2004) Switches, cross-talk and memory in *Escherichia coli* adherence. *J Med Microbiol* 53: 585-593.
10. Dubnau D, Losick R (2006) Bistability in bacteria. *Mol Microbiol* 61: 564-572.

11. Igoshin OA, Price CW, Savageau MA (2006) Signalling network with a bistable hysteretic switch controls developmental activation of the sigma transcription factor in *Bacillus subtilis*. *Mol Microbiol* 61: 165-184.
12. Wolf DM, Arkin AP (2003) Motifs, modules and games in bacteria. *Curr Opin Microbiol* 6: 125-134.
13. Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149: 1633-1648.
14. Chickarmane V, Troein C, Nuber UA, Sauro HM, Peterson C (2006) Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol* 2: e123.
15. Bagowski CP, Ferrell JE, Jr. (2001) Bistability in the JNK cascade. *Curr Biol* 11: 1176-1182.
16. Ferrell JE, Jr. (2002) Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr Opin Cell Biol* 14: 140-148.
17. Tyson JJ, Chen K, Novak B (2001) Network dynamics and cell physiology. *Nat Rev Mol Cell Biol* 2: 908-916.
18. Novak B, Csikasz-Nagy A, Gyorffy B, Chen K, Tyson JJ (1998) Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophys Chem* 72: 185-200.
19. Kashiwagi A, Urabe I, Kaneko K, Yomo T (2006) Adaptive response of a gene network to environmental changes by fitness-induced attractor selection. *PLoS ONE* 1: e49.
20. Becskei A, Seraphin B, Serrano L (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *Embo J* 20: 2528-2535.
21. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339-342.
22. Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EP, Landgraf D, et al. (2007) Rational design of memory in eukaryotic cells. *Genes Dev* 21: 2271-2276.
23. Greber D, Fussenegger M (2007) Mammalian synthetic biology: engineering of sophisticated gene networks. *J Biotechnol* 130: 329-345.
24. Song H, Smolen P, Av-Ron E, Baxter DA, Byrne JH (2006) Bifurcation and singularity analysis of a molecular network for the induction of long-term memory. *Biophys J* 90: 2309-2325.
25. Morimoto BH, Koshland DE, Jr. (1991) Short-term and long-term memory in single cells. *Faseb J* 5: 2061-2067.
26. Acar M, Becskei A, van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435: 228-232.
27. Maynard Smith J (1982) *Evolution and the theory of games*. Cambridge ; New York: Cambridge University Press. viii, 224 p.
28. Wolf DM, Vazirani VJ, Arkin AP (2005) Diversity in times of adversity: probabilistic strategies in microbial survival games. *J Theor Biol* in press.
29. Wolf DM, Vazirani VJ, Arkin AP (2005) A microbial modified Prisoner's Dilemma game: how frequency-dependent selection can lead to random phase variation. *J Theor Biol* in press.
30. Kussell E, Leibler S (2005) Phenotypic diversity, population growth, and information in fluctuating environments. *Science* 309: 2075-2078.
31. Aumann R, Maschler M (1995) *Repeated Games with Incomplete Information*. Cambridge: MIT Press.
32. Basar, T, Olsder GJ (1995) *Dynamic Noncooperative Game Theory*. London and San Diego: Academic Press.
33. Nowak M, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 364: 56-58.
34. Axelrod R, Hamilton WD (1981) The evolution of cooperation. *Science* 211: 1390-1396.
35. Nowak M, Sigmund K (1992) Tit for tat in heterogeneous populations. *Nature* 355: 250-252.
36. Veening JW, Hamoen LW, Kuipers OP (2005) Phosphatases modulate the bistable sporulation gene expression pattern in *Bacillus subtilis*. *Mol Microbiol* 56: 1481-1494.
37. Grossman AD (1995) Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annu Rev Genet* 29: 477-508.
38. Msadek T (1999) When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends Microbiol* 7: 201-207.

39. Voigt CA, Wolf D, Arkin AP (2004) The *B. subtilis* SIN Operon: An Evolvable Network Motif. *Genetics*.
40. Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440: 545-550.
41. Bergman NH, Anderson EC, Swenson EE, Niemeyer MM, Miyoshi AD, et al. (2006) Transcriptional profiling of the *Bacillus anthracis* life cycle in vitro and an implied model for regulation of spore formation. *J Bacteriol* 188: 6092-6100.
42. Kreuzer-Martin HW, Chesson LA, Lott MJ, Dorigan JV, Ehleringer JR (2004) Stable isotope ratios as a tool in microbial forensics--Part 2. Isotopic variation among different growth media as a tool for sourcing origins of bacterial cells or spores. *J Forensic Sci* 49: 961-967.
43. Kreuzer-Martin HW, Chesson LA, Lott MJ, Dorigan JV, Ehleringer JR (2004) Stable isotope ratios as a tool in microbial forensics--Part 1. Microbial isotopic composition as a function of growth medium. *J Forensic Sci* 49: 954-960.
44. Lim HN, van Oudenaarden A (2007) A multistep epigenetic switch enables the stable inheritance of DNA methylation states. *Nat Genet* 39: 269-275.
45. Ewing JA (1881) The effects of stress on the thermoelectric quality of metals. *Proceedings of the Royal Society of London* 32: 399-402.
46. Krasnosel'skij MA, Lifshits JA, Sobolev AV (1989) *Systems with Hysteresis*. Berlin: Springer Verlag.
47. Sha W, Moore J, Chen K, Lassaletta AD, Yi CS, et al. (2003) Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proc Natl Acad Sci U S A* 100: 975-980.
48. Shannon CE (1948) A mathematical theory of communication. *Bell Systems Technical Journal* 27: 379-423 and 623-656.
49. Landauer TK (1986) How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science* 10: 477-493.
50. Borst A, Theunissen FE (1999) Information theory and neural coding. *Nat Neurosci* 2: 947-957.
51. DeWeese MR, Meister M (1999) How to measure the information gained from one symbol. *Network* 10: 325-340.
52. Butts DA (2003) How much information is associated with a particular stimulus? *Network* 14: 177-187.
53. Sharpee T, Rust NC, Bialek W (2004) Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16: 223-250.
54. Ash RB (1990) *Information Theory*: Dover Publications.
55. Arigoni F, Duncan L, Alper S, Losick R, Stragier P (1996) SpoIIE governs the phosphorylation state of a protein regulating transcription factor sigma F during sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* 93: 3238-3242.
56. Stragier P, Losick R (1996) Molecular genetics of sporulation in *Bacillus subtilis*. *Annu Rev Genet* 30: 297-241.
57. Stahl ML, Ferrari E (1984) Replacement of the *Bacillus subtilis* subtilisin structural gene with an In vitro-derived deletion mutation. *J Bacteriol* 158: 411-418.
58. Harwood C, Cutting SM (1990) *Molecular Biology Methods for Bacillus*. Wiley-Intersciences.
59. Collins S, Bell G (2004) Phenotypic consequences of 1,000 generations of selection at elevated CO<sub>2</sub> in a green alga. *Nature* 431: 566-569.
60. Ancel Meyers L, Bull JJ (2002) Fighting change with change: adaptive variation in an uncertain world. *Trends in Ecology & Evolution* 17: 551-557.
61. Gazzaniga MS (2004) *The Cognitive Neurosciences III*: MIT Press.
62. Willeboordse FH, Kaneko K (2005) Externally controlled attractor selection in a high-dimensional system. *Phys Rev E Stat Nonlin Soft Matter Phys* 72: 026207.
63. Segel LA (1998) Multiple attractors in immunology: theory and experiment. *Biophys Chem* 72: 223-230.
64. Huang S, Eichler G, Bar-Yam Y, Ingber DE (2005) Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* 94: 128701.
65. Bengtsson G (1977) Output regulation and Internal models: a frequency domain approach. *Automatica* 13: 333-345.
66. Lande R, Orzack SH (1988) Extinction dynamics of age-structured populations in a fluctuating environment. *Proc Natl Acad Sci U S A* 85: 7418-7421.

67. Thattai M, van Oudenaarden A (2004) Stochastic gene expression in fluctuating environments. *Genetics* 167: 523-530.
68. Sambrook J, Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. NY: Cold Spring Harbor Laboratory Press.
69. Cormack BP, Valdivia RH, Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* 173: 33-38.
70. Fujita M, Losick R (2002) An investigation into the compartmentalization of the sporulation transcription factor sigmaE in *Bacillus subtilis*. *Mol Microbiol* 43: 27-38.
71. Quisel JD, Lin DC, Grossman AD (1999) Control of development by altered localization of a transcription factor in *B. subtilis*. *Mol Cell* 4: 665-672.
72. Jan J, Valle F, Bolivar F, Merino E (2001) Construction of protein overproducer strains in *Bacillus subtilis* by an integrative approach. *Appl Microbiol Biotechnol* 55: 69-75.
73. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20: 53-65.
74. Vastano JA, Swinney HL (1988) Information transport in spatiotemporal systems. *Physical Review Letters* 60: 1773-1776.
75. Samoilov M, Arkin A, Ross J (2001) On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos* 11: 108-114.
76. Chen I, Dubnau D (2003) DNA transport during transformation. *Front Biosci* 8: s544-556.
77. Kunst F, Msadek T, Bignon J, Rapoport G (1994) The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in *Bacillus subtilis*. *Res Microbiol* 145: 393-402.
78. Garrity LF, Ordal GW (1995) Chemotaxis in *Bacillus subtilis*: how bacteria monitor environmental signals. *Pharmacol Ther* 68: 87-104.
79. Marahiel MA, Nakano MM, Zuber P (1993) Regulation of peptide antibiotic production in *Bacillus*. *Mol Microbiol* 7: 631-636.
80. Entian KD, de Vos WM (1996) Genetics of subtilin and nisin biosyntheses: biosynthesis of lantibiotics. *Antonie Van Leeuwenhoek* 69: 109-117.
81. Kearns DB, Chu F, Branda SS, Kolter R, Losick R (2005) A master regulator for biofilm formation by *Bacillus subtilis*. *Mol Microbiol* 55: 739-749.
82. Sanchez A, Olmos J (2004) *Bacillus subtilis* transcriptional regulators interaction. *Biotechnol Lett* 26: 403-407.

## Figure Captions

**Figure 1:** The *B. subtilis* stress response meta-network, where each oval represents both a stress response and the regulatory network of 100 or so interacting molecular species that regulates it. Among the many ingenious genetic and biochemical programs employed by *B. subtilis* to cope with environmentally adverse conditions are its ability to take up extracellular DNA, competence [40,76]; differentiate into an inert heat-, chemical, and UV-resistant spore [37]; secrete degradative enzymes to identify and digest new food sources [77]; become motile and chemotax toward possibly better surroundings [78]; synthesize antibiotics to eliminate competitors in the same ecosystem [79,80]; turn on alternative metabolic pathways, and form biofilms ([81], not shown), just to name a few [38]. The cross-repressive feedback between sporulation and competence, and the many positive feedback loops within each large ‘individual’ stress response pathway [10,36-40,82], are suggestive of switches and other elements that could potentially encode memory. The two stress response pathways monitored in our experiments, sporulation and synthesis of the degradative enzyme subtilisin, are denoted by bold-faced ovals. The fluorescent reporters (GFP and DsRed) fused to the respective promoters *PspoIIE* and *PaprE* are indicated (see Materials and Methods).

**Figure 2:** An ‘adaptive’ memory experiment. In an adaptive memory experiment, each (identical) sample of a biological system is subjected to one of several conditions prior to time  $t_0$ , and then observed in a common condition after  $t_0$ . If different past histories lead to different short-term behaviors in current conditions, the system can be said to exhibit short-term memory. If different past histories lead to different long-term behaviors, the system can be said to exhibit long-term memory.

**Figure 3:** Information-based conceptual schema for measuring memory in microbes. In communication theory (top), the informational entropy of the signal space  $H(X)$  captures the number of different messages  $X$  that can be communicated and their probabilistic dispersal; the mutual information  $I(X,Y)$  between transmitted and received signals quantifies the amount of information actually communicated. A memory experiment, in contrast, involves subjecting cells to distinct treatments  $M$  prior to time  $t_0$ , followed by an identical treatment  $S$  after time  $t_0$ , with cell behavior from  $t_0$  on monitored through temporal sampling of one or more observable variables  $Y$ . As applied to bacterial memory (bottom), the informational entropy of the cell history space  $H(M)$  captures the number of different cell histories prior to time  $t_0$  tested by the experimental compendium and their probabilistic dispersal; the mutual information  $I_{\text{trans}}(M;Y;t_{\text{trans}})$  between the transient response of the observable variable  $Y$  after time  $t_0$  and the cell history prior to time  $t_0$  captures the short-term memory of cell history exhibited by  $Y$  over the cell history space in response to treatment  $S$ . Likewise, the mutual information  $I_{\text{asym}}(M,Y)$  between the long-term response of  $Y$  and cell history prior to  $t_0$  captures the long-term memory of cell history exhibited by  $Y$ .

**Figure 4:** Different types of history-dependent behavior one might observe. a) Short-term deterministic memory. State trajectories ‘remember’ their initial condition for some time, and then converge to a common asymptotic behavior. b) Long-term deterministic memory. State trajectories of multi-stable systems ‘remember’ which basin of attraction their initial condition started in indefinitely (the basin containing  $X_{01}$  vs. the basin containing  $X_{02}$  and  $X_{03}$ ), but retain a memory of the exact initial condition within a basin of attraction only transiently ( $X_{02}$  vs.  $X_{03}$ ). c) Short-term and Long-term memory in a system with unobservable states. The state space of the cell is two dimensional ( $X,Y$ ), but only one of the two dimensions,  $X$ , is observed. Though all four initial conditions are distinct in the larger space, the unobserved  $Y$  component renders them identical to the observer. Thus the trajectories appear to diverge from a common starting point and approach one of two asymptotic states. This gives the observer the impression of first an increase in information and memory and then a decrease as the trajectories approach their long-term values. d,e) If measurements are made on single cells rather than on averaged populations (as we did in this paper), history-dependent distributions may be observed. d) Short-term stochastic memory. State trajectories are probabilistic in individual cells, with a distribution over the population that initially retains a ‘memory’ of the initial condition of the population. In the long-term, this memory degrades as the distribution approaches a global attractor. e) Long-term stochastic memory. The distribution over the population retains a ‘memory’ of the initial condition indefinitely, or at least over the time-horizon of the experiment.

**Figure 5:** *B. subtilis* memory data compendium. These plots show the dynamics of the sporulation initiation reporter *PspoIIE-gfp* expression (a), the degradative enzyme synthesis reporter *PaprE-dsred* expression (b), and cell growth (c) of *B. subtilis* KEE after the onset of starvation (resuspension in SM) as a function of cell history prior to starvation, as measured by fluorescence (GFP, and DsRed) and  $OD_{600}$  time series measurements taken every 15 minutes for 24 hours, respectively.

The 10 cell histories tested consisted of growth in either rich LB medium or poorer GM medium to one of five densities  $D$ ,  $-1$ ,  $-2$ ,  $-3$ ,  $-4$ , (see experimental overview section and Materials and Methods for details). Fluorescent intensities in (a-b) were divided by  $OD_{600}$  (c) and then normalized to a  $[0\ 1]$  scale by dividing by the maximum. The error bars show standard deviation over replicates at each time point.

**Figure 6:** The map from cell history to *B. subtilis* stress response clusters. The transient dynamics and long-term levels of the sporulation initiation (*PspoIIE-gfp* expression), AprE synthesis (*PaprE-dsred* expression), and growth ( $OD_{600}$ ) signals were clustered using the automatic method in Materials and Methods. This figure shows the heat maps for each signal in Figure 5 (dark red indicates maximum, and dark blue minimum), the number of behavioral classes for each signal, and which subset of the ten cell histories in our test set corresponds to each cluster. For example, the asymptotic sporulation initiation signal from *PspoIIE-gfp* fusion clustered into two classes, one (top, 1) corresponding to a history of growth in rich LB medium to the three highest densities,  $D$ ,  $-1$ , and  $-2$ , and the other class (bottom, 2) corresponding to all other cell histories.

**Figure 7:** Estimates of cell-history memory and mutual information in *B. subtilis*. The upper left bar plot shows our estimate of long-term (blue bars) and short-term (first 11 hours, red bars) memory fidelity (% of the maximum recoverable information about cell history) exhibited in starvation medium SM by sporulation initiation (*PspoIIE-gfp* expression), degradative enzyme synthesis (*PaprE-dsred* expression), and growth dynamics ( $OD_{600}$ ), and over all vector pairs of observable read-outs and the vector triple, with respect to the cell history space tested by our compendium. The lower right bar plot shows our estimate of the number of bits of mutual information shared by all pairs of short-term (red bars) and long-term (blue bars) observable signals in our memory data compendium. The surrounding flow diagram circuit illustrates the experimental and analytical scenario.

## Table

n	Cell history	Cell History Description
1	LB: D	Grown in LB (rich medium) to density D ( $OD_{600} = 1$ )
2	LB: -1	Grown in LB to density -1 ( $OD_{600} \approx 0.65$ )
3	LB: -2	Grown in LB to density -2 ( $OD_{600} \approx 0.4$ )
4	LB: -3	Grown in LB to density -3 ( $OD_{600} \approx 0.2$ )
5	LB: -4	Grown in LB to density -4 ( $OD_{600} \approx 0.1$ )
6	GM: D	Grown in GM (less rich medium) to density D
7	GM: -1	Grown in GM to density -1
8	GM: -2	Grown in GM to density -2
9	GM: -3	Grown in GM to density -3
10	GM: -4	Grown in GM to density -4

**Table 1:** Cell history table. The cell history space  $M$  consists of 10 cell histories  $M=(MediumI, DensityI)$ : growth in either rich Luria Broth medium (LB) or a less rich growth medium (GM) [58] to one of five cell densities, D, -1, -2, -3, -4.

# Figures

Figure 1

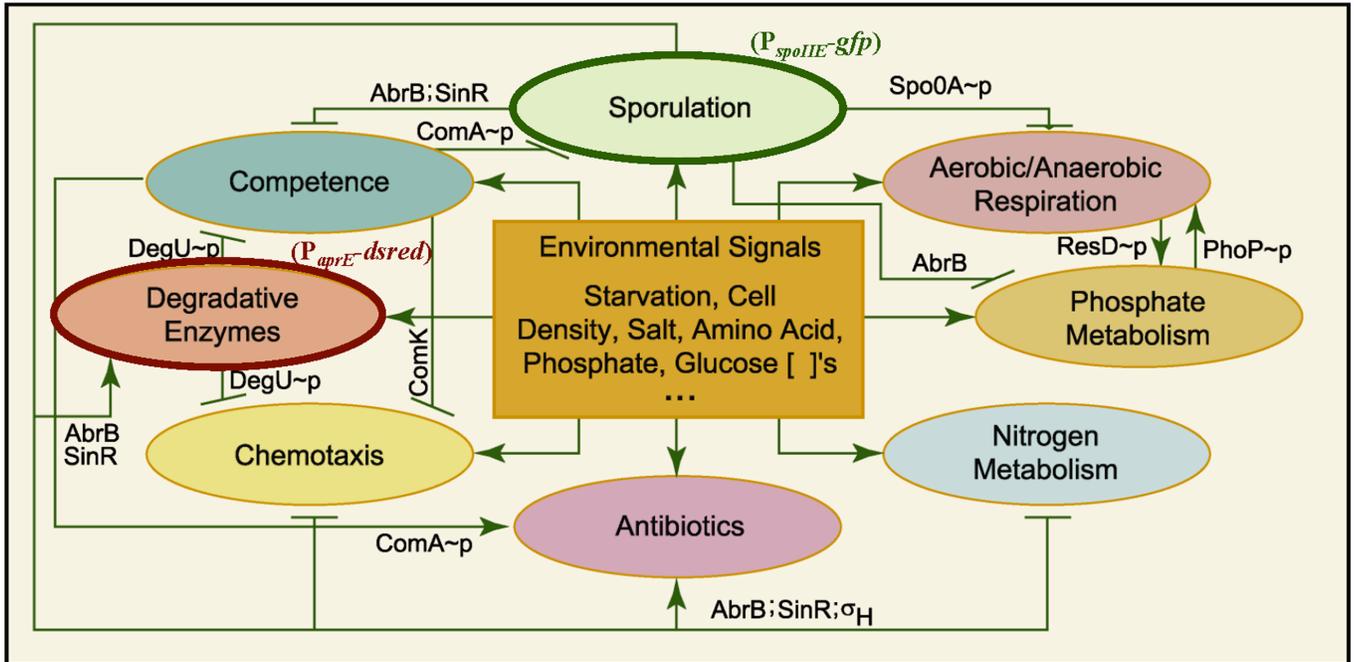
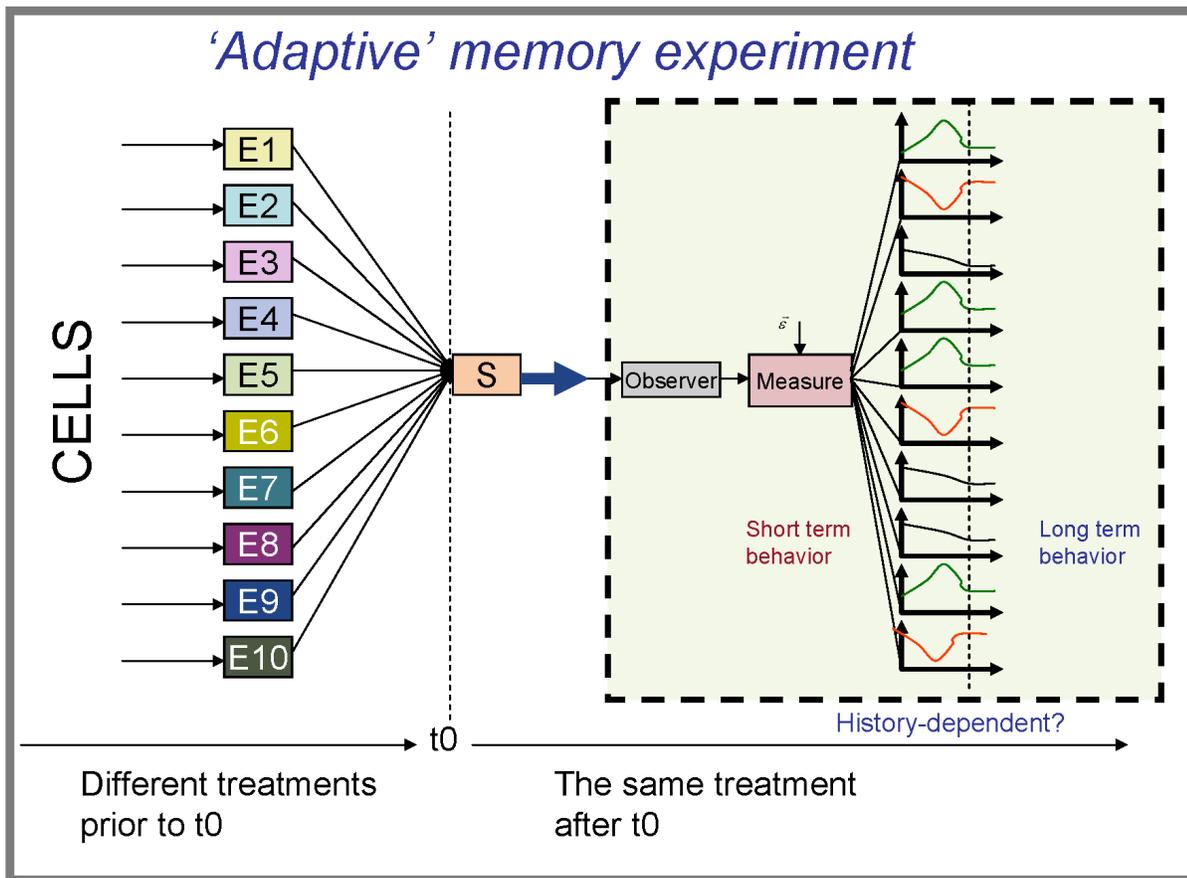


Figure 2



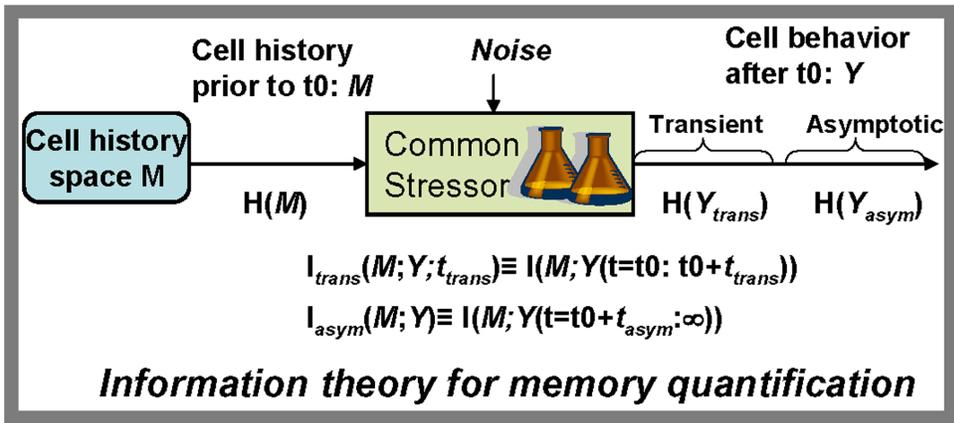
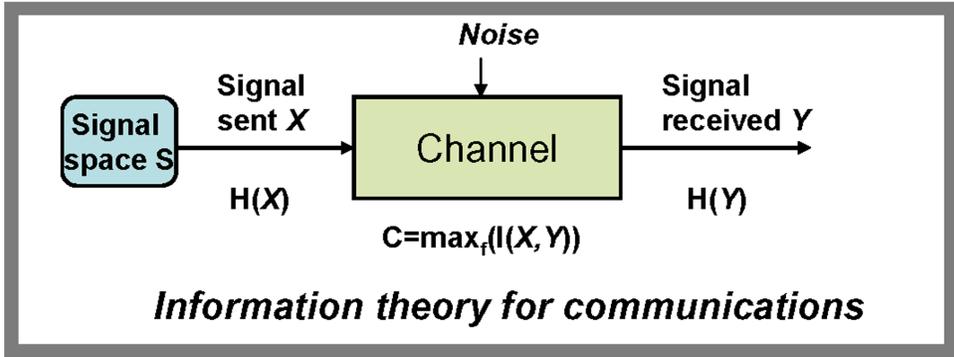
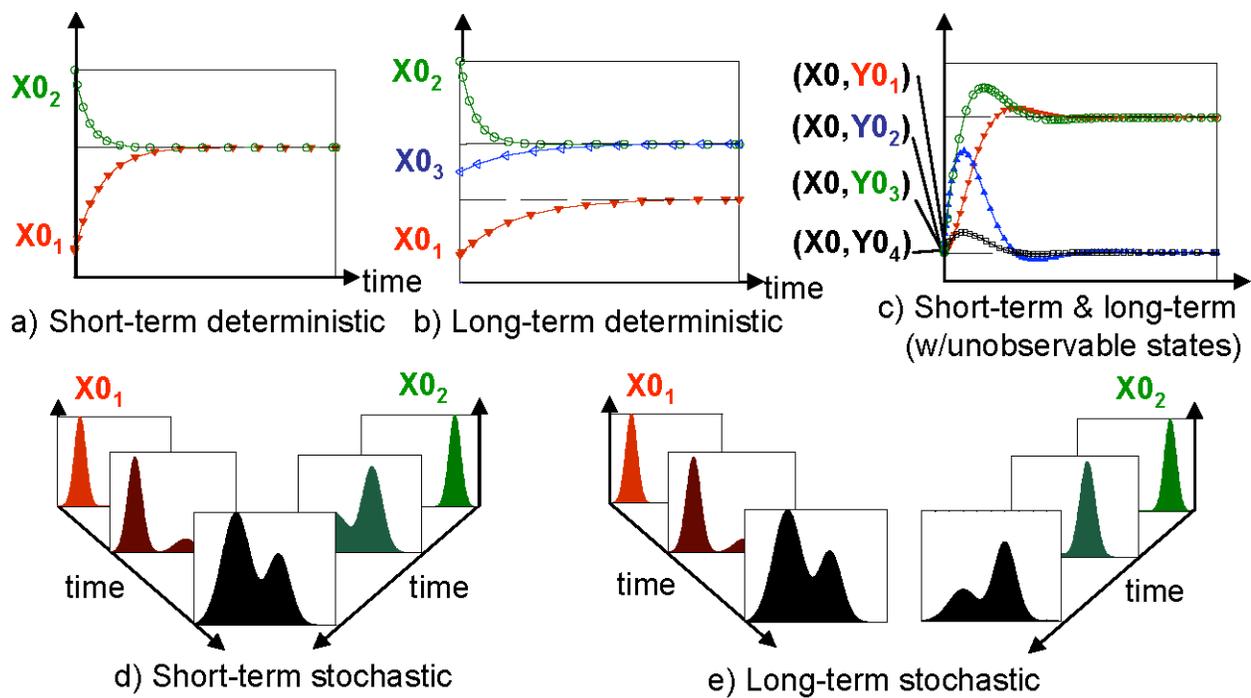


Figure 3

Figure 4



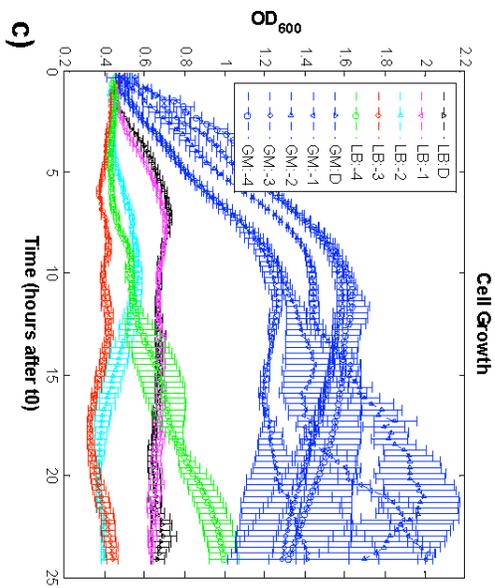
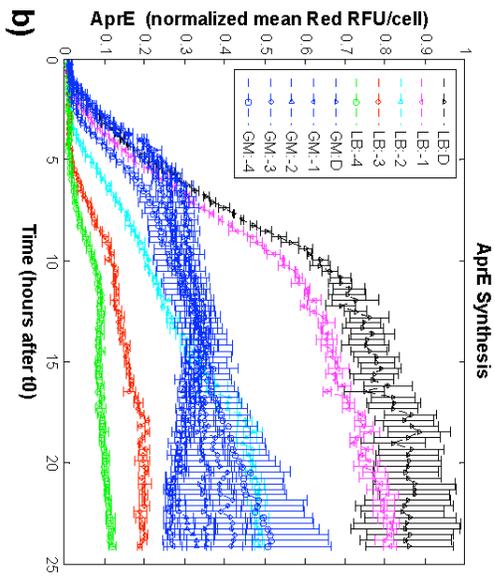
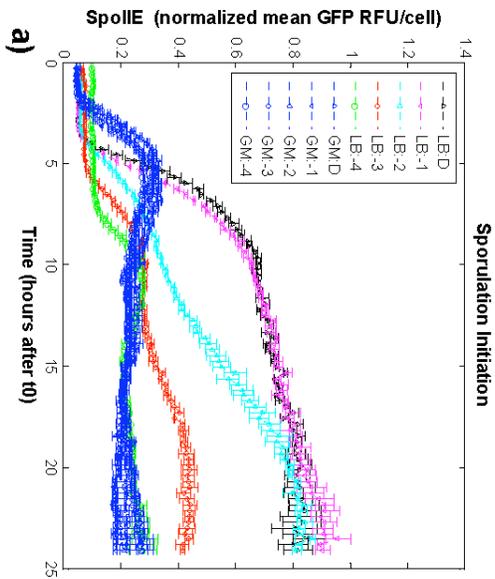


Figure 5

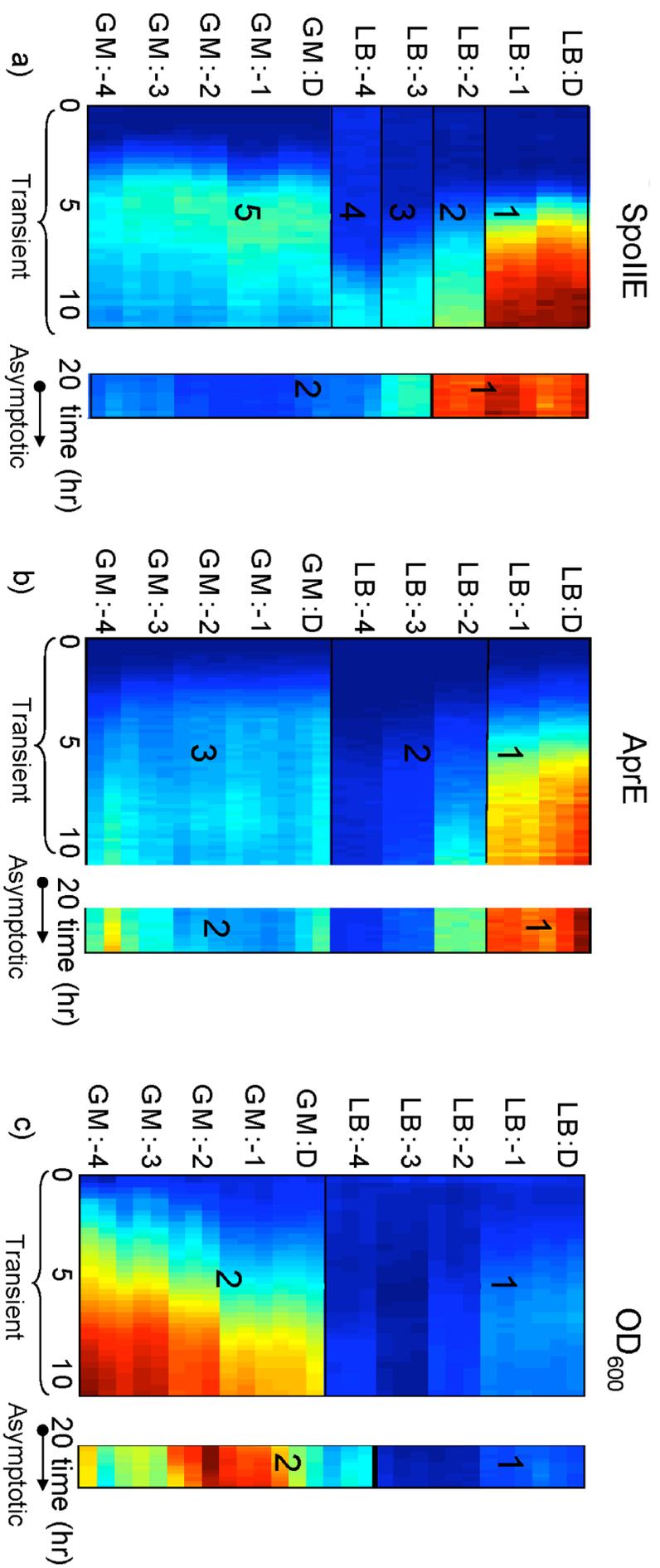


Figure 6

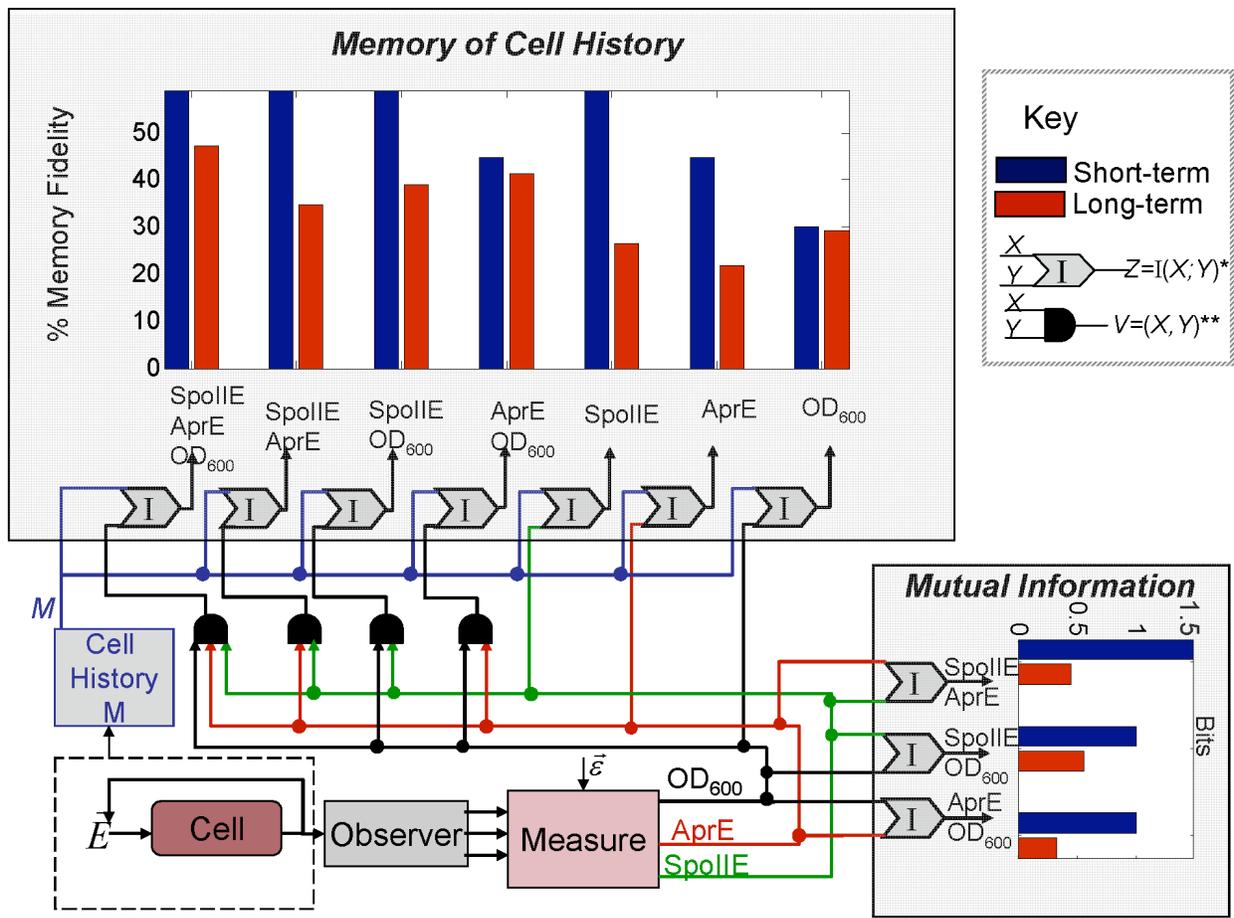


Figure 7

## Supplementary Information S1

For “Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium”, by Denise M. Wolf, Lisa Fontaine-Bodin, Ilka Bischofs, Gavin Price, Jay Keasling, and Adam P Arkin. PLoS ONE 2008

### Appendix S1. Memory quantification definitions

1. **(Adaptive) memory experiment compendium:** A set of experiments in which populations of one or more cells are subjected to distinct treatments  $m$  prior to time  $t_0$ , and an identical treatment  $S$  following  $t_0$ , with at least one observable read-out of behavior  $Y$  sampled over a time series following  $t_0$ . The set of all data collected from the memory experiment compendium, viewed as a measurement-transformed sampling of the map from past cell history to future cellular response, is called the *memory data compendium*.
2. **Entropy of cell history space tested:** Within the memory experiment compendium, cell history  $M$  is considered a random variable. If  $M$  is the set of all experimentally tested cell treatments/histories  $m$  prior to  $t_0$  that  $M$  could be, and  $p(m) = \Pr(M = m)$ , then  $M$  has  $H(M) = - \sum_{m \in M} p(m) \log_2(p(m))$  bits of informational entropy [1].
3. **Long term memory:** Let  $t_{asym}$  be the time it takes for the observable response  $Y$  to approach steady state. Then within  $M$ , the cells under study exhibit  $I_{asym}(M; Y) \equiv I(M; Y(t=t_0+t_{asym}:\infty))$  bits of *long term memory* in the observable response  $Y$  to stress condition  $S$ .
4. **Short term memory:** Let  $t_{trans} < t_{asym}$ . Though short-term, transient behavior (and memory) may be measured over any interval or at any time point between  $t_0$  and  $t_{asym}$  (see the discussion in SI.2) we use the interval  $[t_0 t_{trans}]$  in our memory calculations as follows: within  $M$  the cells under study exhibit  $I_{trans}(M; Y; t_{trans}) \equiv I(M; Y(t = t_0:t_0+ t_{trans}))$  bits of *short term memory* over  $(t_0: t_0+ t_{trans})$  in the observable response  $Y$  to stress condition  $S$ .
5. **Memory fidelity:** The *short-term memory fidelity* exhibited over  $(t_0: t_0+ t_{trans})$  in response  $Y$  of the cells to stressor  $S$  given  $M$  is  $P_{trans}(M; Y; t_{trans}) \equiv I(M; Y(t = t_0:t_0+t_{trans}))/H(M)$ . The *long-term memory fidelity* exhibited in response  $Y$  of the cells to stressor  $S$  given the cell history space  $M$  is  $P_{asym}(M; Y) \equiv I(M; Y(t = t_0+ t_{asym}:\infty))/H(M)$ . This normalized mutual information metric, a measure between 0 and 1 of the fraction of uncertainty about the past conditions tested that is reduced by knowledge of future cellular response, has also been called the coefficient of constraint [2].
6. **Memory orthogonality:** The *memory orthogonality* between two pathway responses  $Y1$  and  $Y2$  of cells subjected to stress condition  $S$  given cell histories  $M$  is:  $Mem_{orth}(M; (Y1, Y2)) \equiv (I(M; (Y1, Y2)) - \max(I(M; Y1), I(M; Y2))) / \min(I(M; Y1), I(M; Y2))$ .  $Mem_{orth}$  equals 1 if the two variables combined as a vector yield the upper bound of memory under these conditions, and 0 if the two variables in combination yield the lower bound (a consequence of the inequality

$\min(I(M;Y1),I(M;Y2)) \leq I(M;(Y1,Y2)) \leq I(M;Y1) + I(M;Y2)$  [2]). This definition extends naturally to output triples and higher order combinations, as well.

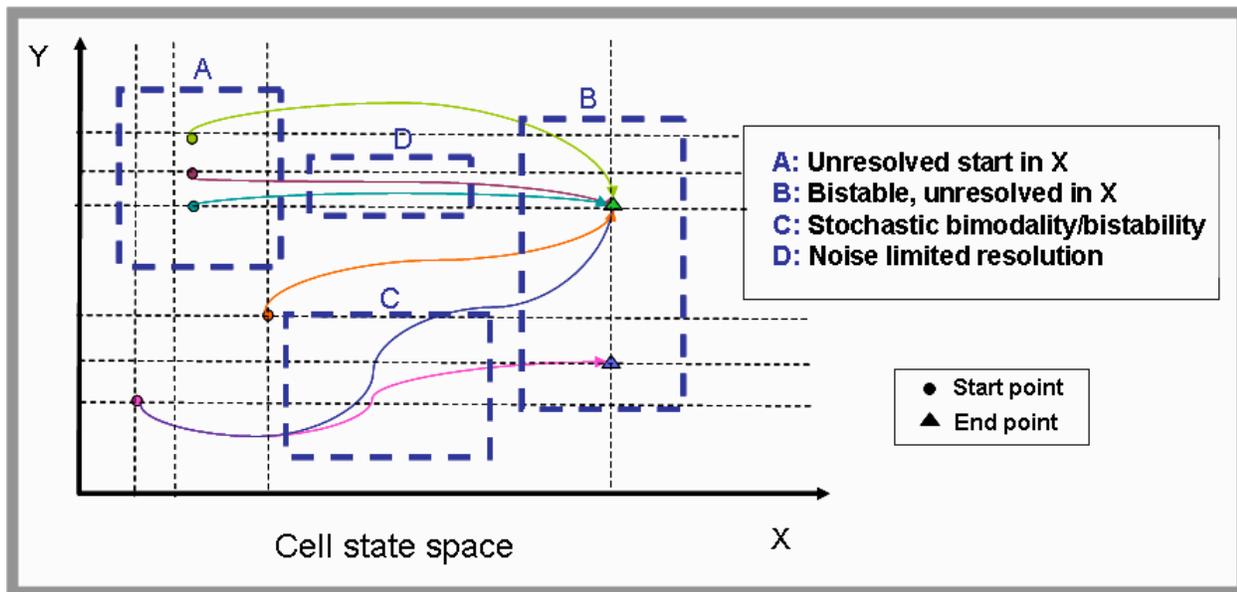
## References

1. Shannon CE (1948) A mathematical theory of communication. Bell Systems Technical Journal 27: 379–423 and 623–656.
2. Ash RB (1990) Information Theory: Dover Publications.

## Supplementary Information S2

For “Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium”, by Denise M. Wolf, Lisa Fontaine-Bodin, Ilka Bischofs, Gavin Price, Jay Keasling, and Adam P Arkin. PLoS ONE 2008

### S2. Fundamental limitations of memory experiments



**Figure S2.** Any attempt to quantify or classify memory is fundamentally limited by the possibility of unobservable states, uncontrolled and unobservable inputs, and measurement errors and distortions. This figure shows a variety of initial conditions (start points, circles) in a toy cell state space, each corresponding to a different cell history up to time  $t_0$ . After  $t_0$ , the system evolves over time to representative asymptotic behaviors (end points, triangles), including a deterministic steady state (green triangle in box B), bistability (both triangles in box B), and stochastic bistability (boxes C, B). Box A shows distinct initial conditions that are observable if both X and Y are observed or only Y is observed, but not if only X is observed. Thus, the amount of history-dependence in the system would be underestimated if only X were observed. Likewise, box B shows distinct asymptotic states that are not observable if only X is measured. The three trajectories starting in box A also demonstrate how memory of initial conditions can decrease over time, possibly even with no long term memory at all. The trajectories passing through box D are resolvable only until measurement noise overwhelms the signal. If measurements are averaged over the population, stochastic or multimodal features of the trajectory (boxes C, B) become invisible, whereas if measurements are taken over the population in the form of histograms (flow cytometry), a population consisting of cells that stochastically alternate between ‘off’ and ‘on’ states is indistinguishable from a population of cells that switch into one of the two states and then remain in that state for the duration of the experiment..

*Unobservable states:* In a perfect world, one would take direct, noiseless, high-resolution measurements of every state variable in the cell and its environment, and exercise perfect control over input perturbations. Real world experiments suffer from limitations due to imperfect observation and input control, lack of knowledge of the space of meaningful biological inputs, and finite time and

resources. These limitations necessarily constrain our ability to detect, classify, and quantify memory. One major limitation derives from the presence of unobserved cell state variables. A real-world observer selects a small subset of state variables to observe and measure, thereby collapsing the high-dimensional state space of a cell or population into a low-dimensional projection. Collapsed dynamics can lead to misclassifying and misquantifying memory.

Though physical first principles predict that chemistry within cells should behave like Markov processes, the presence of unobserved states can make a Markov memory appear to be non-Markov. If, for example, two initial conditions differ only in the y axis but have the same value along the x axis (Figure S2, box A), an observation of only the x-component of the two diverging trajectories would suggest to a naïve observer that the system behaviour depends not only on the initial condition, but also on the path leading to that initial condition. For similar reasons, unobserved states can also lead to underestimates of transient and asymptotic memory. Two system trajectories responding to two distinct cell histories may appear to be identical from the perspective of the low-dimensional observations, yet be perfectly distinct in the higher-dimensional native state space (Figure S2, boxes A,B).

Uncontrolled, unobserved system inputs can also lead to misclassifying and misquantifying memory, though the errors are different. If one were to repeat an experiment where a stimulus is applied and the system response is measured, and there are uncontrolled, unobserved, randomly varying input components that the observer is not aware of, a deterministic system could be misclassified as stochastic. If the uncontrolled, unobserved input correlates with the controlled portion of the input stimulus, one could overestimate the ability of the system to ‘remember’ the specific observed stimulus, though not the overall memory of the system.

*Measurement distortions, information loss and errors:* The next layer of challenge in quantifying cellular memory derives from imperfect and distorting measurement of cellular response. Distortions and information loss arise from proxy measurements, instrument error, experimental error, time discretization/integration, and from indirect measurement modalities. Fluorescent transcriptional fusions that supposedly measure promoter activity introduce their own protein transcription, translation, folding, and degradation dynamics, and can cause toxicity and therefore perturb the system being measured. Errors from proxy measurements combine with those from instrument error, experimental error, and information loss from time discretization/integration to limit the resolution and dynamic range of the behavior we can observe. This in turn can lead to underestimates of the amount of memory in a system (quantification errors). The cell may have distinct trajectories and attractor states associated with distinct cell histories, but if the trajectories and states cannot be resolved due to the ‘noise’ in these various types of measurement error (Figure S2, box D), an underestimation of the memory capacity of the system will result. Classification errors can result too, for example a stochastic response that appears deterministic because the distribution lies within measurement noise bounds.

Measurement modalities operating on bulk, averaged populations are especially prone to classification ambiguities. With bulk measurements as one obtains with fluorimetry, for example, averaged measurements over populations of cells conflate deterministic behavior (all cells express some GFP) and stochastic bistability (some cells express a high level of GFP and others express none) (Figure S2, boxes B, C). This can lead to a memory classification error if one assumes the average measurement reflects individual cell behavior. A memory quantification error (underestimate) can also result, if the average response to distinct cell histories is the same but the single-cell response distributions differ.

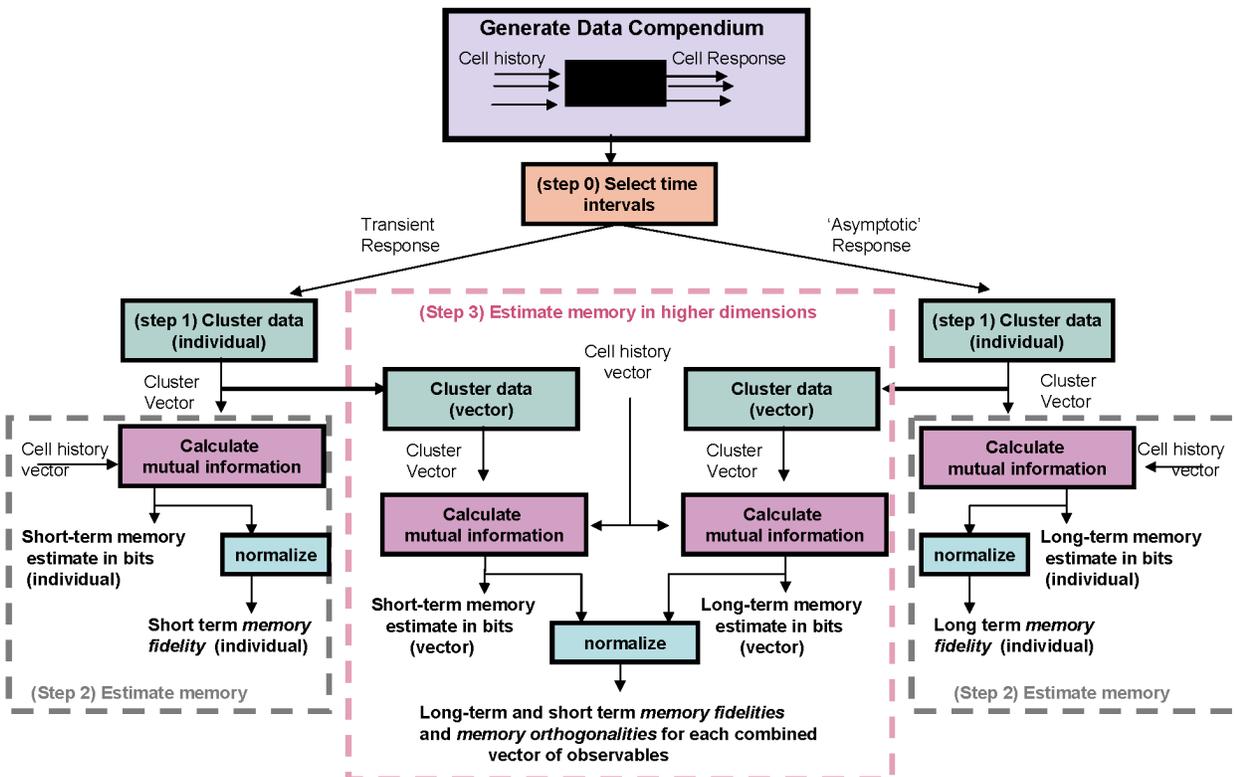
Population distribution measurements as one obtains through flow cytometry are less prone to classification errors, as they can distinguish between deterministic behavior and stochastic behavior, but they still suffer ambiguities discriminating between stochastic bistability associated with rapid transitioning between states bistability with little if any transitioning between states on experimental time scales. They also lose information due to the limited dynamic range of the machine, finite binning of fluorescence levels, different sized cells, differential cell growth and death rates among cell subpopulations in different states, and the inability to follow single cell fates. Even single-cell measurements from time-lapse microscopy suffer from sampling errors, image processing errors, errors due to fluctuating illumination and drift, and difficulty identifying rare events. All of these errors distort trajectories and can lead to underestimations of memory.

Because these limitations are inherent in the estimation of memory processes and most likely result in underestimates of the ability of the system to ‘remember’ the cell histories tested by the experimental compendium, we interpret quantifications of memory within our compendium as lower bound estimates.

## Supplementary Information S3

For “Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium”, by Denise M. Wolf, Lisa Fontaine-Bodin, Ilka Bischofs, Gavin Price, Jay Keasling, and Adam P Arkin. PLoS ONE 2008

### S3. Matlab codes for memory analysis



**Figure S3.** Schematic of a fixed-interval, information-based approach to quantifying memory in bacteria (A3 in Fig. S4). As described in Materials and Methods, in our implementation of a fixed-interval approach to memory quantification in bacteria we first (step 0) parse the time series data into a transient set and a long-term ‘asymptotic set’. Then, for the transient and asymptotic data sets, respectively, we: (step 1) use the Matlab scripts in S3.2 to hierarchically cluster the trajectories and select an ‘optimal’ partition using the Silhouette criterion, a clustering validation and optimization technique based on maximizing the compactness and separation of the clusters in a partition [1]. This step produced six cluster vectors, one transient and one ‘asymptotic’ cluster vector for each of the three observables (i.e., ClustSPO\_trans, ClustSPO\_asym, ClustAprE\_trans, ClustAprE\_asym, ClustOD\_trans, ClustOD\_asym); (step 2) calculate memory in bits as the mutual information between cell history and cell behavior cluster for each observable on each time-scale using the Matlab program in S3.1 with input vectors  $\text{ClustA}=\text{M}=[1\ 1\ 1\ 2\ 2\ 2\ \dots\ 10\ 10\ 10]$  as the cell history vector and ClustB equal to one of the six cluster vectors from step 1. To calculate memory fidelities, we normalized the memory estimates by dividing by  $H(\text{M})=3.32$ , the entropy of the cell history space; (step 3) calculate memory in bits exhibited by pairs of observables and by the triple of observables by (a) using the script in S3.3 to combine cluster vectors from multiple read-outs (e.g., Clust1=GFP cluster vector; Clust2=DsRed cluster vector; Clust3=combined (GFP,DsRed) cluster vector) and (b) using as inputs to the program in S3.1, ClustA=(the cell history vector M) and ClustB=(the combined cluster vector Clust3), calculate the mutual information between cell history and cell behavior in the higher dimensional spaces. These estimates are normalized by  $H(\text{M})$  to estimate *memory fidelities* and by Equation (1) to estimate *memory orthogonalities*. Finally, we (step 4, not shown) calculate the mutual information between the observables using the program in S3.1. Though we did

not have enough replicates to meaningfully do so in this paper, with enough replicates one could estimate confidence intervals for all memory and mutual information bit counts by performing a statistical (non-parametric) bootstrap analysis. Such an analysis would involve random sampling with replacement from the replicate sets, followed by clustering and mutual information calculations for each derived data set to generate a distribution of memory estimates. From this distribution, confidence intervals could easily be determined.

### **S3.1) Matlab program to calculate memory and mutual information (Steps 2-4 in the data analysis algorithm in Materials and Methods).**

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Function: Entropy_MutualInfo.m
%
%Description: This MATLAB program accepts two input vectors, A and B,
%and calculates from them individual entropies H(A) and H(B), the
%entropy of the pair H(A,B), and the mutual information between A and
%B:  $I(A;B) = H(A)+H(B)-H(A,B)$ .
%
%Inputs:
%   ClustA='optimal' clustering vector for A (each entry an integer)
%   ClustB='optimal' clustering vector for B (each entry an integer)
%Outputs:
%   HA=entropy of A in bits
%   HB=entropy of B in bits
%   HAB=entropy of the vector (A,B) in bits
%   IAB=mutual information of A and B in bits
%
%Interpretation: IF (A=cell history) and (B=Response)
%                THEN memory = IAB. IF (A=response 1) and (B=Response2)
%                THEN mutual information = IAB and memory in combined
%                response vector = HAB (assuming cell history
%                space uniformly distributed).
%
%Author: Denise Wolf dnmwolf@lbl.gov
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [HA,HB,HAB,IAB]=Entropy_MutualInfo(ClustA,ClustB)

no_clustersA=max(ClustA);
no_clustersB=max(ClustB);

%H=-sum(log2(pi)pi) = informational entropy

HA=0;

for i=1:no_clustersA
    size_clustiA=length(find(ClustA==i));
    piA(i)=size_clustiA/length(ClustA);
    HA=HA-log2(piA(i))*piA(i);
end

```

```

HB=0;

for i=1:no_clustersB
    size_clustiB(i)=length(find(ClustB==i));
    piB(i)= size_clustiB(i)/length(ClustB);
    HB=HB-log2(piB(i))*piB(i);
end

%%I(X,Y)=H(X)+H(Y)-H(X,Y)= mutual information of X,Y
%%H(X,Y)=sum_x(sum_y(pi_j*log2(pi_j))= entropy of vector (X,Y)

PAB=zeros(no_clustersA,no_clustersB); %PAB is the probability distribution of A,B
HAB=0;
for j=1:no_clustersB
    for i=1:no_clustersA

PAB(i,j)=length(intersect(find((ClustA==i)),find((ClustB==j))))/length(ClustA);
        if PAB(i,j)>0
            HAB=HAB-PAB(i,j)*log2(PAB(i,j));
        end
    end
end
end

IAB=HA+HB-HAB;

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

### **S3.2) Matlab scripts for visualizing and clustering data (Step 1).**

```

%Input: Data = m x n time series matrix, where m = the number of
% trajectories and n= the number of time points.
% History_Labels = m x 1 text vector with cell history labels.
%
% Directions: Run this script twice, first with the ward linkage and
% second with the average linkage, and select the partition
% (Tx; x∈[1,7]) that maximizes the mean silhouette. We run this script
% twice because the two linkage functions can produce
% different trees, and thus different partitions. The 'best' partition of
% the 14 calculated, according the silhouette criterion
% (maximum mean silhouette over all clusters, for seven cutoffs each from %
% the trees constructed using average and ward linkages) is then selected
% as the 'optimal' clustering output of the procedure.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%Calculate distance matrix
Data_dist=pdist(Data,'euclidean');

%Calculate linkage function
Zfunction_ward=linkage(Data_dist,'ward');
Zfunction_avg=linkage(Data_dist, 'average');
Zfunction=Zfunction_ward; %or avg

```

```

%Visualize clustering over all cell histories
figure;[H,T,perm] = dendrogram(Zfunction,0, 'colorthreshold','default',
'orientation','left','labels',History_Labels,'colorthreshold',4.5);

figure %Calculate cluster vectors T for the first 7 levels of tree-cut
[H,T2,perm] = dendrogram(Zfunction,2, 'colorthreshold','default',
'orientation','left','labels',History_Labels);
[H,T3,perm] = dendrogram(Zfunction,3, 'colorthreshold','default',
'orientation','left','labels', History_Labels);
[H,T4,perm] = dendrogram(Zfunction,4, 'colorthreshold','default',
'orientation','left','labels', History_Labels);
[H,T5,perm] = dendrogram(Zfunction,5, 'colorthreshold','default',
'orientation','left','labels', History_Labels);
[H,T6,perm] = dendrogram(Zfunction,6, 'colorthreshold','default',
'orientation','left','labels', History_Labels);
[H,T7,perm] = dendrogram(Zfunction,7, 'colorthreshold','default',
'orientation','left','labels', History_Labels);

figure;
subplot(3,3,1) %3x3 plot, first figure= visualize 2-d projection
[Y,eigvals] = cmdscale(Dfunction);
plot(Y(:,1),Y(:,2),'.','MarkerSize',15);
text(Y(:,1),Y(:,2),History_Labels,'FontSize',7);
xlabel('MDS scaled distance')
ylabel('MDS scaled distance')

subplot(3,3,2) %Second figure = silhouettes for 2-level tree cut
[s2,h2]= silhouette(Data,T2,'euclid');

subplot(3,3,3) %Third figure = silhouettes for 3-level tree cut
[s3,h3]=silhouette(Data,T3,'euclid');

subplot(3,3,4) %Fourth figure = silhouettes for 4-level tree cut
[s4,h4]=silhouette(Data,T4,'euclid');

subplot(3,3,5) %Fifth figure = silhouettes for 5-level tree cut
[s5,h5]=silhouette(Data,T5,'euclid');

subplot(3,3,6) %Sixth figure = silhouettes for 6-level tree cut
[s6,h6]=silhouette(Data,T6,'euclid');

subplot(3,3,7) %Seventh figure = silhouettes for 7-level tree cut
[s7,h7]=silhouette(Data,T7,'euclid');

subplot(3,3,8) %Eighth figure = mean silhouette for each partition
bar([2:1:7],[mean(s2) mean(s3) mean(s4) mean(s5) mean(s6) mean(s7)])
xlabel('Number of clusters in partition')
ylabel('Mean Silhouette')

end

```

Caveat: Silhouette is a popular cluster validation/selection measure that scores a partition based on both the compactness of its clusters and their separation [1]. People like it because it is systematic, and because it takes into account local and global properties of a partition. However, like all clustering validation/selection measures, Silhouette has its biases. It is a conservative

criterion in that there is a tendency for the method to swamp out detail and under-report the number of clusters if there are multiple widely-spaced scales of cluster separation.

For the data set analyzed in this paper, this is most evident in the clustering of the transient data set for growth ( $OD_{600}$ ). By eye it appears that there might be some structure in the data set at early time points ( $OD_{600}$  of cells with a history of growth in GM, but to different densities, figs. 4 and 5c) that gets subsumed into two large clusters because of the large separation between the set faster growing trajectories with a history of growth in GM and the much slower growing set of trajectories with a history of growth in LB. Despite this conservatism, we used Silhouette because of all the methods we considered, it seemed to do the best job selecting tree cuts (mostly because the majority of our data is structured on a single level of resolution). However, given the difficulty in selecting a good clustering validation technique, we suggest that the reader take the structure of their data sets into account when selecting a clustering criterion so as to get the best 'match' between the bias of the method and the data to be analyzed.

### S3.3) Matlab script for combinatorially combining cluster vectors (Step 2-3).

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Input:   Clust1 - the cluster vector from one observable (integers)
%         Clust2 - the cluster vector from another observable (integers)
%
%Output:  Clust3 - the cluster vector obtained by combinatorially combining
%           the cluster vectors ClustA and ClustB.
%
%Example: Say Clust1 = [ 1 1 2 2 2] and Clust2 = [ 1 1 1 2 2]. Then Clust3
%         =[1 1 2 3 3] to show that there are three possible combinations:
%         (Clust1,Clust2)=(1,1); (Clust1,Clust2)=(2,1) and (Clust1,Clust2)=(2,2)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```
D=[Clust1 Clust2];
```

**Clust3=clusterdata(D,0.1);** %clusterdata.m is a function in the  
**Statistics**

**Matlab.** % **Toolbox of**

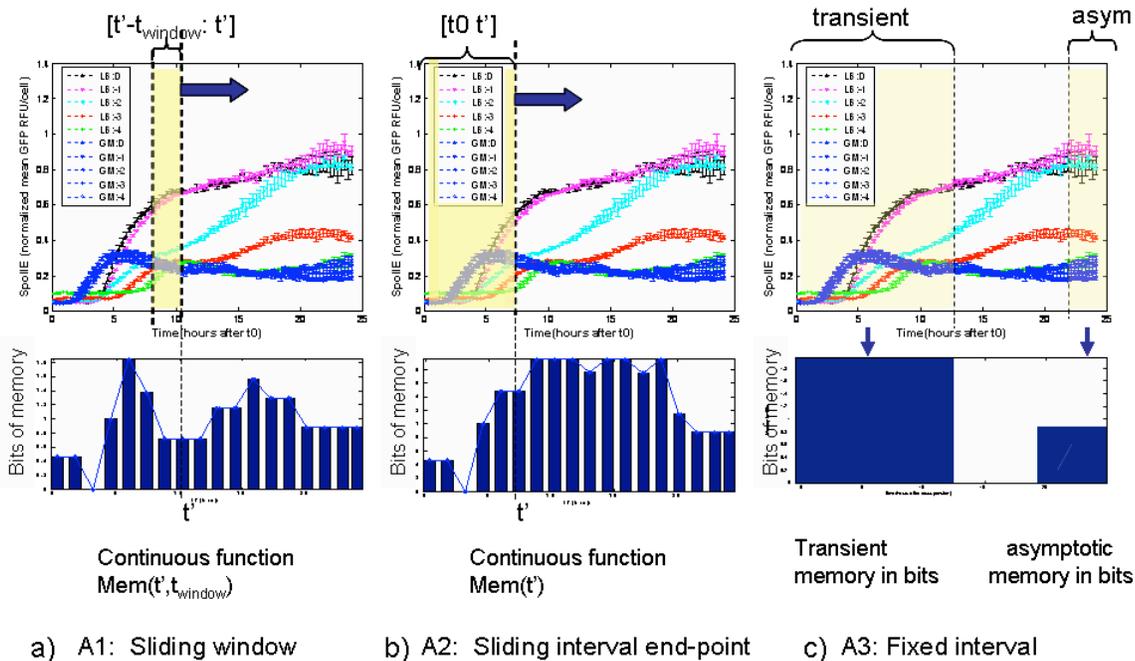
### References

1. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20: 53-65.

## Supplementary Information S4

For “Memory in Microbes: Quantifying History-Dependent Behavior in a Bacterium”, by Denise M. Wolf, Lisa Fontaine-Bodin, Ilka Bischofs, Gavin Price, Jay Keasling, and Adam P Arkin. PLoS ONE 2008

S4. Alternative strategies in time-series interval sampling for memory calculations:



**Figure S4.** Three possible approaches to interval sampling for memory calculations: a) sliding window (A1), b) sliding interval end-point (A2), and c) fixed interval (A3).

Memory calculations for adaptive memory experiments involve calculating the mutual information between cell history before time  $t_0$  and cell behavior after time  $t_0$ . If cellular behavior is measured at a single time point, say at 24 hours after  $t_0$ , it would be clear how to proceed, at least in terms of temporal sampling of the response: one would calculate the mutual information between the cellular response measurements at that single time point and the cell histories corresponding to each measurement. However, because in our experiments behavior after time  $t_0$  is measured as a time series – every 15 minutes for 24 hours – a choice must be made about which time points or intervals to analyze.

On one end of the spectrum is a sliding window approach (which we will call A1). If the window size is 1, admitting just a single time point, this approach amounts to calculating the mutual information between cell history and cell response at each sampled time point after  $t_0$ . Or one could select a window size greater than 1, and ‘slide’ the window along the time axis, producing a calculation for

mutual information between cell history and response from  $t' - t_{\text{window}}$  to  $t'$ , as  $t'$  ranges from  $t_0 + t_{\text{window}}$  to the final measurement at  $t_0 + 24$  hours. The sliding window approach produces a (sampled) continuous time-varying memory measurement function, as shown in Figure S4a above for a window size of 5 sample points ( $\approx 1.5$  hours) applied to the sporulation initiation time series data. (Though Swinney's information analysis algorithm is in general appropriate for a sliding window approach, we used the clustering and silhouette analysis approach described in Supplementary Information Section S3 for this example because a much larger data set than ours would be required to use Swinney's analysis algorithm at measurement intervals short enough to avoid excessive 'blurring' of the time series dynamics [1] [2]).

On the other end of the spectrum is a fixed-interval approach (approach A3). This approach involves extracting transient and asymptotic segments of the temporal cellular response. The mutual information between the transient response and cell history is then used to estimate transient memory, whereas the mutual information between the asymptotic response and cell history is used to estimate asymptotic memory. This approach is discrete, in that it maps a time series onto just two numbers – the transient and asymptotic bit counts, as shown in Figure S4c for the sporulation initiation time series data.

Between these two ends of the spectrum is the sliding interval end-point approach (approach A2). In this approach all analyzed measurement intervals start at time  $t_0$ , but the end point is variable, or sliding, starting from  $t_0$  and ending at the final time point,  $t_0 + 24$ hr. Figure S4b shows the result of this approach applied to the sporulation initiation time series data.

Each of the three approaches has its pros and cons. Both the sliding window (A1) and the sliding interval end-point (A2) approaches are appealing in that the measures are continuous, and as the final time gets very large relative to time it takes to converge to asymptotic behavior, both measurements converge to the asymptotic value. Another interesting aspect of these approaches is that they communicate the temporal increase and decrease of information and mutual information from the perspective of the observer. The increase captures, among other (conflated) dynamics, the rate at which the history-dependent states of the cells become observable in the read-out, whereas the decrease captures (also among other conflated dynamics) the decay rate of memory as it relaxes from the maximum amount of transient memory to the lesser long-term memory. This notion of memory decay time could also be mapped onto a single measurement like a half-life - the time after  $t_0$  after which the amount of memory remains at less than half of the difference between the maximum transient memory and the long term memory. One potential down-side, however, especially for the sliding window method, is the emphasis on the moment experience of an outside observer of as increases and decreases in information and memory due to 'curve crossing' resulting from unobservable states are traced (see Figure S4a). This emphasis is a departure from our goal of quantifying alternative pathway control conditioned by cell history. A naive interpretation of memory curves derived from these approaches could also be misleading. For instance, looking at Figure S4a,b, one might conclude that the cells mostly do not 'remember' their history for the first few hours, and then start to remember their history when the curves diverge. However, since cell history prior to  $t_0$  is most likely to impact cell behavior after  $t_0$  largely through an initial condition of the cell as a whole at time  $t_0$ , a more likely interpretation is that though the cells encode memory of their past experiences maximally at time  $t_0$ , the manifestation of this memory in the observable does not become apparent for several hours.

Consequently, given our interest in capturing substantial qualitative differences in transient and long-term patterns of behavior of the pathways as conditioned on prior history, we chose the discrete, fixed-

interval approach (A3). Though this approach is not as visually interesting as the continuous information and memory curves derived from A1 and A2, we believe it is more consistent with our goals of calling out history-dependent behavior patterns and is appealing in its relative simplicity.

One issue with the fixed-interval method we chose is that one must choose time intervals to analyze. This may be done ad hoc, according to the interests of the analyst, or more systematically. One way to approach interval selection systematically is to perform a sliding interval end-point analysis, as shown in (b) above for the sporulation time series, to determine the information dynamics of the curves as a function of interval. If the goal is to pick a transient interval that maximizes information & memory, the end point can be selected by choosing a time point at the maximum on the curve. For our data, intervals that start at  $t_0$  and end anywhere from 8 to 19 hours after  $t_0$  are roughly equivalent, informationally speaking. We selected 11 hours after  $t_0$  to be the end point of the transient data interval, but would have obtained the same results had we chosen nearly any point between 8 and 19 hours after  $t_0$ .

## References

1. Vastano JA, Swinney HL (1988) Information transport in spatiotemporal systems. *Physical Review Letters* 60: 1773-1776.
2. Samoilov M, Arkin A, Ross J (2001) On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos* 11: 108-114.

S5. Complete set of memory and mutual information calculations

**Mutual information across observable output pairs in bits**

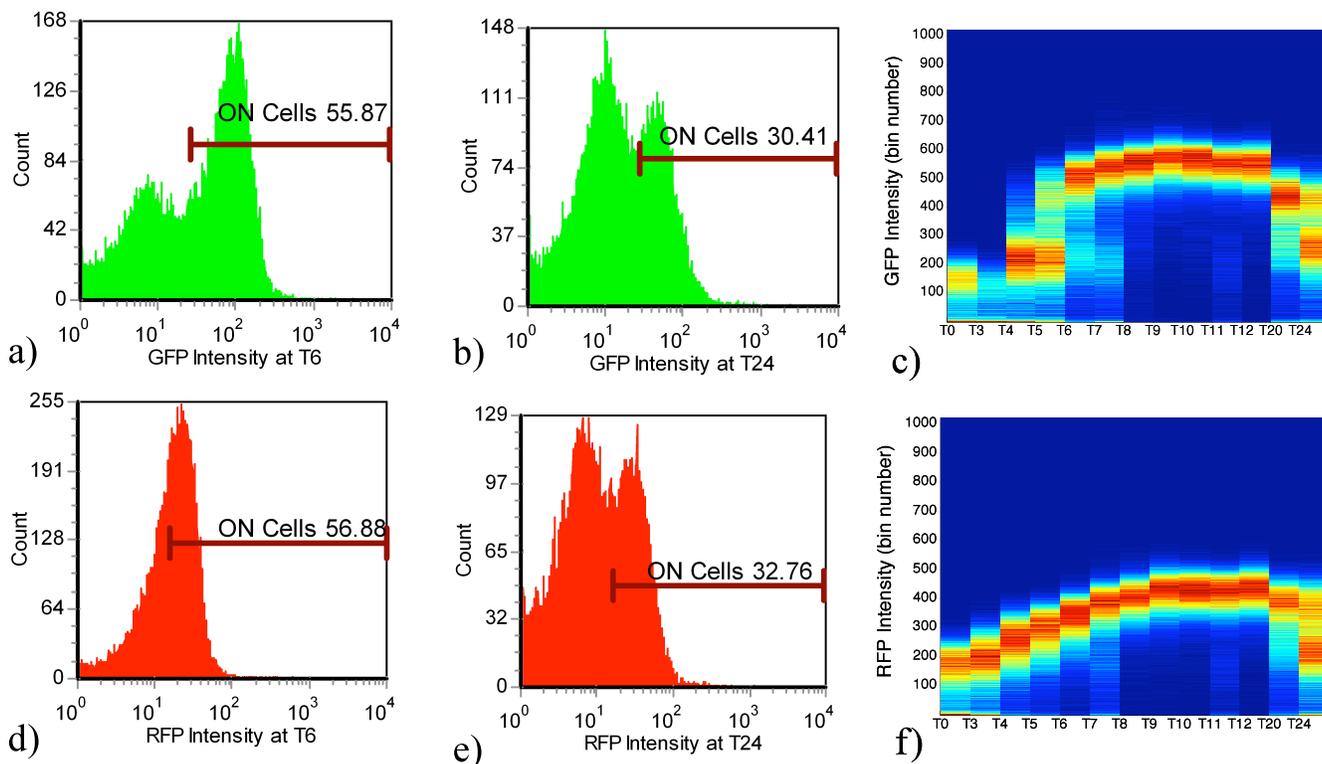
Memory of cell history in bits

	AprE asym	SpoIIE asym	OD <sub>600</sub> asym	AprE trans	SpoIIE trans	OD <sub>600</sub> trans
AprE asym	<b>0.7219</b>	0.4464	0.3219	0.7219	0.7219	0.2365
SpoIIE asym	1.1568	<b>0.8813</b>	0.5568	0.6058	0.8813	0.3958
OD <sub>600</sub> asym	1.371	1.2955	<b>0.971</b>	0.6955	0.971	0.61
AprE trans	1.4855	1.761	1.761	<b>1.4855</b>	1.4855	1
SpoIIE trans	1.961	1.961	1.961	1.961	<b>1.961</b>	1
OD <sub>600</sub> trans	1.4855	1.4855	1.361	1.4855	1.961	<b>1</b>

**Table S5:** Estimated lower bounds on cell-history memory (red, lower left triangle) and mutual information (blue, upper right triangle) in informational entropy bits for each individual *B. subtilis* stress response read-out and each vector pair of read-outs, with respect to the cell histories tested by our compendium. Read-outs include both transient and ‘asymptotic’ sporulation initiation (*PspoIIE-gfp* expression), AprE synthesis (*PaprE-dsred* expression), and growth (OD<sub>600</sub>) signals. For example, the mutual information between transient sporulation and AprE signals can be found in the 4<sup>th</sup> row and 5<sup>th</sup> column of the matrix (1.4855 bits) whereas the memory encoded in the combined transient sporulation and AprE signals can be found in the 5<sup>th</sup> row and 4<sup>th</sup> column of the matrix (1.961 bits). Notice that the upper right 3x3 off-diagonal block contains estimates of the mutual information between all combinations of transient and long-term signals, which provides an estimate for how much transient memory carries over into the long term response. See Materials and Methods for details on the calculations.

## S6. The distribution over the population of GFP and DsRedExpress expression levels for selected cell histories and time points.

The histograms below derive from a flow cytometric analysis of the *B. subtilis* reporter strain KEE at selected time points and for selected cell histories.



**Figure S6.** Histograms from flow cytometry analysis of *B. subtilis* cells with a history of growth in rich LB medium to a density of  $-2$  ( $OD_{500}=1.016$ ) prior to resuspension in starvation (SM) medium at time  $t_0$  at an initial density of  $OD_{600}=0.5$ . a) *PspolIE-gfp* expression (GFP intensity) at 6 hours after  $t_0$ ; b) *PspolIE-gfp* expression (GFP intensity) at 24hrs after  $t_0$ ; c) normalized heatmap of *PspolIE-gfp* expression histograms (GFP intensity) sampled at selected time points from  $t_0$  to 24 hours after  $t_0$ ; d) *PaprE-dsred* expression (RFP intensity) at 6 hours after  $t_0$ ; e) *PaprE-dsred* expression (RFP intensity) at 24hrs after  $t_0$ ; f) normalized heatmap of *PaprE-dsred* expression (RFP intensity) sampled at selected time points from  $t_0$  to 24 hours after  $t_0$ . In the heatmaps shown in (c,f), the distance along the y-axis is fluorescent intensity channel number (logarithmic scale), and the color along the y-axis corresponds to the (relative) number of cells (red corresponds to the maximum, yellow to an intermediate number of cells and dark blue to no cells).

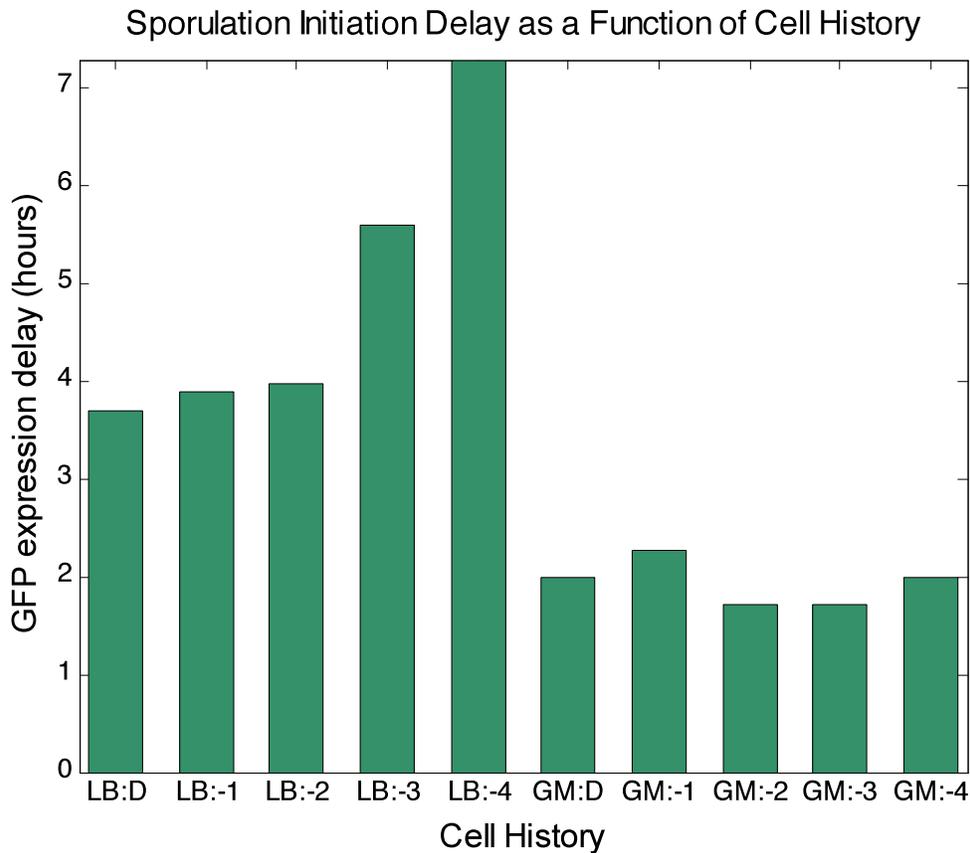
### Flow cytometry protocol:

Cells sampled at 10h and 20h after resuspension in SM medium were washed twice in one volume of  $0.2 \mu\text{M}$  filtered PBS (phosphate-buffered saline, pH 7.4), 100x diluted in the same buffer and directly measured on a Partec CyFlow space flow cytometer (Partec GmbH, Münster Germany) operating an argon laser (488 nm). For each sample, at least 50,000 cells were analyzed. Data containing the green fluorescent signals were collected by a 520 nm BP filter and the red fluorescent signals were collected by a 590 nm BP filter. Data was captured using Flomax 2.4f software (May 23 2006) and further analyzed using the commercial software package FCSEXPRESS (<http://www.denovosoftware.com>). Background fluorescence was analyzed with parental strain *B. subtilis* 168 with each flow cytometric

experiment to discriminate background from GFP and DsRed specific fluorescence. The heatmaps were constructed using the Matlab© functions `surface.m`, `axis.m`, and `set.m`.

## S7. Sporulation initiation delay as a function of cell history

The plot below shows sporulation initiation delay as a function of cell history as manifested in GFP expression delay in our *B. subtilis* reporter strain KEE.



**Figure S7.** Sporulation initiation delay as a function of cell history. Cells were subjected to one of 10 cell histories prior to time  $t_0$  and then to a common stress condition, resuspension in SM, a sporulation ‘starvation’ medium, after  $t_0$  (see Table 1 for key to cell history labels). In theory at least, sporulation initiation *delay* is a likely feature of the sporulation regulation strategy of *B. subtilis* to exhibit memory, because of its potential to have a large impact on fitness. By calculating the mutual information between sporulation initiation delay times and total transient sporulation dynamics, we find that 85.95% of the transient memory exhibited by the sporulation initiation pathway is explained by cell-history modulation of sporulation initiation delay.

## S8. Bacillus strains and plasmids table

The experiments in this paper made use of the following bacterial strains and plasmids:

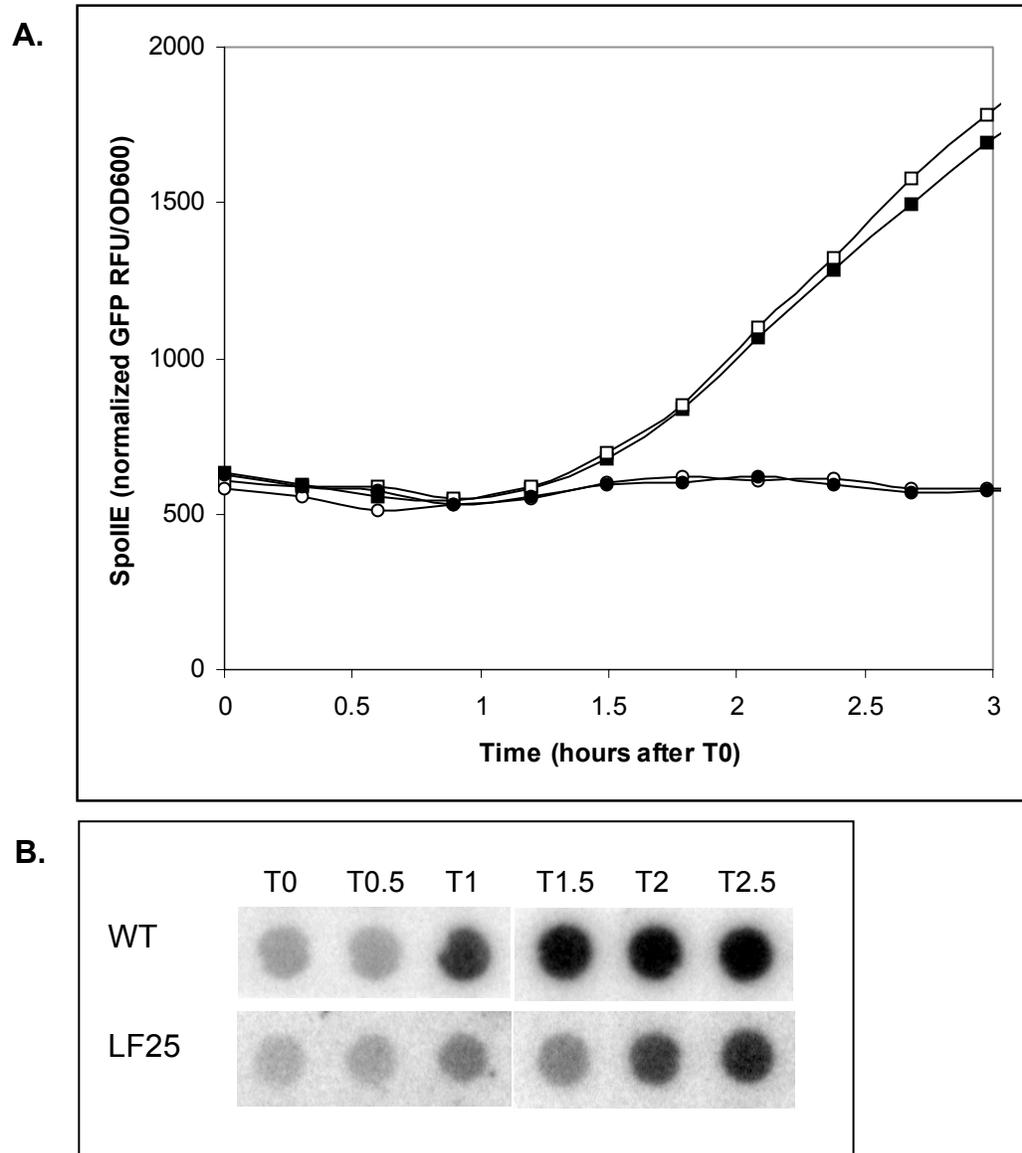
Strains and plasmids	Genotype	Reference
<b>Bacillus subtilis</b>		
168	<i>trpC2</i>	[1]
LF25	168, <i>amyE::P<sub>spoIIIE</sub>-gfp</i> , <i>cmp</i>	This study
KEE	168, <i>amyE::P<sub>spoIIIE</sub>-gfp</i> , <i>P<sub>aprE</sub>-dsred</i> , <i>cmp</i>	This study
<b>Plasmids</b>		
pMF19	<i>P<sub>spoIIIG</sub>-gfp</i> , <i>spc</i>	[2]
pEA18	<i>P<sub>xyI</sub>-gfp</i> , <i>cmp spc</i>	[3]
pLF22	<i>P<sub>spoIIIG</sub>-gfp</i> , <i>cmp spc</i>	This study
pLF25	<i>P<sub>spoIIIE</sub>-gfp</i> , <i>cmp spc</i>	This study
pSG-TTGACA	<i>P<sub>aprE</sub>-lacZ</i> , <i>cmp</i>	[4]
pDsRed-Express	<i>P<sub>lac</sub>-DsRed-Express</i> , <i>amp</i>	Clontech
pLFKEE	<i>P<sub>spoIIIE</sub>-gfp</i> , <i>P<sub>aprE</sub>-dsred</i> , <i>cmp spc</i>	This study

## References

1. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249-256.
2. Fujita M, Losick R (2002) An investigation into the compartmentalization of the sporulation transcription factor sigmaE in *Bacillus subtilis*. *Mol Microbiol* 43: 27-38.
3. Quisel JD, Lin DC, Grossman AD (1999) Control of development by altered localization of a transcription factor in *B. subtilis*. *Mol Cell* 4: 665-672.
4. Jan J, Valle F, Bolivar F, Merino E (2001) Construction of protein overproducer strains in *Bacillus subtilis* by an integrative approach. *Appl Microbiol Biotechnol* 55: 69-75.

## S9. *PspoII*E-*gfp* fusion activity compared to the *spoII*E gene expression profile

The data below shows that our GFP reporter for sporulation initiation in *B. subtilis* faithfully tracks expression of the stage II sporulation gene *spoII*E.



**Figure S9.** A. GFP activity in strain LF25 (*amyE::PspoII*E-*gfp* *cmp*) and *spoII*E transcriptional profile. Wild type and LF25 strains were grown in parallel in GM medium to an OD<sub>600</sub> of 0.6 and 0.7 respectively. Cells were resuspended in SM medium to induce sporulation (as described in Materials and Methods) and two 200  $\mu$ l aliquots were transferred into a microplate for time course measurement in a Safire spectrofluorimeter (TECAN inc.) with shaking, at 37°C. Fluorescence (481 nm absorption and 507 nm emission) and OD<sub>600</sub> were measured every 15 minutes. A. LF25 strain time points are shown with squares, whereas wild type *Bacillus subtilis* strain time points are shown with circles. Relative fluorescence (RFU) was normalized by the OD<sub>600</sub>.

B. The transcriptional profile of the *spoII*E gene was verified by total RNA dot-blot. Total RNA was extracted from *B. subtilis* cultures after induction of sporulation as previously described for *C. acetobutylicum*, [1]. RNA quality and quantity were checked by capillary electrophoresis using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, California). Total RNA samples (8 $\mu$ g each) taken just after resuspension (T0) and at 30 minutes

intervals (T0.5, T1, T1.5, T2, T2.5) were spotted on positively charged Nylon membranes (Roche) using a dot blot manifold (Perkin Elmer). Denaturation, fixation on membrane and hybridization were performed as previously described for Northern blots [Fontaine, 2001 #15]. A radioactively labeled probe was PCR amplified within the *spoIIE* gene using primers *spoIIE-D* (cgtcggtaccATGGAAAAGCAGAAAGAAGAG) and *spoIIE-R* (cctcggatccaccTGAAATTTCTTGTTTGTGTTTGA) on *B. subtilis* 168 genomic DNA as template. The resulting 689-bp fragment was radiolabeled as previously described [8].

The data shown in (B) confirmed *spoIIE* early and specific expression induction at the onset of sporulation, starting clearly from T1. The GFP activity of the *PspoIIE-gfp* fusion (squares) in (A) showed a clear increase starting from T1.5 after the resuspension event.

## References

1. Fontaine L, Even S, Soucaille P, Lindley ND, Cocaign-Bousquet M (2001) Transcript quantification based on chemical labeling of RNA associated with fluorescent detection. *Anal Biochem* 298: 246-252