

## **Fitness cost associated with cell phenotypic switching drives population diversification dynamics and controllability**

**Lucas Henrion<sup>1</sup>, Juan Andres Martinez<sup>1</sup>, Vincent Vandembroucke<sup>1</sup>, Mathéo Delvenne<sup>1</sup>, Samuel Telek<sup>1</sup>, Andrew Zicler<sup>1</sup>, Alexander Grünberger<sup>2</sup>, Frank Delvigne<sup>1\*</sup>**

<sup>1</sup>Terra research and teaching centre, Microbial Processes and Interactions (MiPI), Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

<sup>2</sup>Microsystems in Bioprocess Engineering, Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology, Karlsruhe, Germany

\*Corresponding author: [F.Delvigne@uliege.be](mailto:F.Delvigne@uliege.be)

### **Abstract**

Isogenic cell populations can cope with stress conditions by switching to alternative phenotypes. Even if it can lead to increased fitness in a natural context, this feature is typically unwanted for a range of applications (e.g., bioproduction, synthetic biology, biomedicine...) where it tends to decrease the controllability of the cellular response. However, little is known about the diversification profiles that can be adopted by a cell population. We characterized the diversification dynamics for various systems (bacteria and yeast) and for different phenotypes (utilization of alternative carbon sources, general stress response and more complex development patterns). Interestingly, our results suggest that the diversification dynamics and the fitness cost associated with cell switching are coupled. For quantifying the contribution of the switching cost on population dynamics, we built a stochastic model that allowed us to reproduce the dynamics observed experimentally and identified three diversification regimes, i.e., constrained (at low switching cost), dispersed (at medium and high switching cost), and bursty (for very high switching cost). Furthermore, we used a cell-machine interface that we call the Segregostat to demonstrate that different levels of control can be applied to these diversification regimes, enabling applications involving more precise cellular responses.

**Keywords:** phenotypic switching, population control, automated flow cytometry, biological noise, bet hedging, information theory

## 1. Introduction

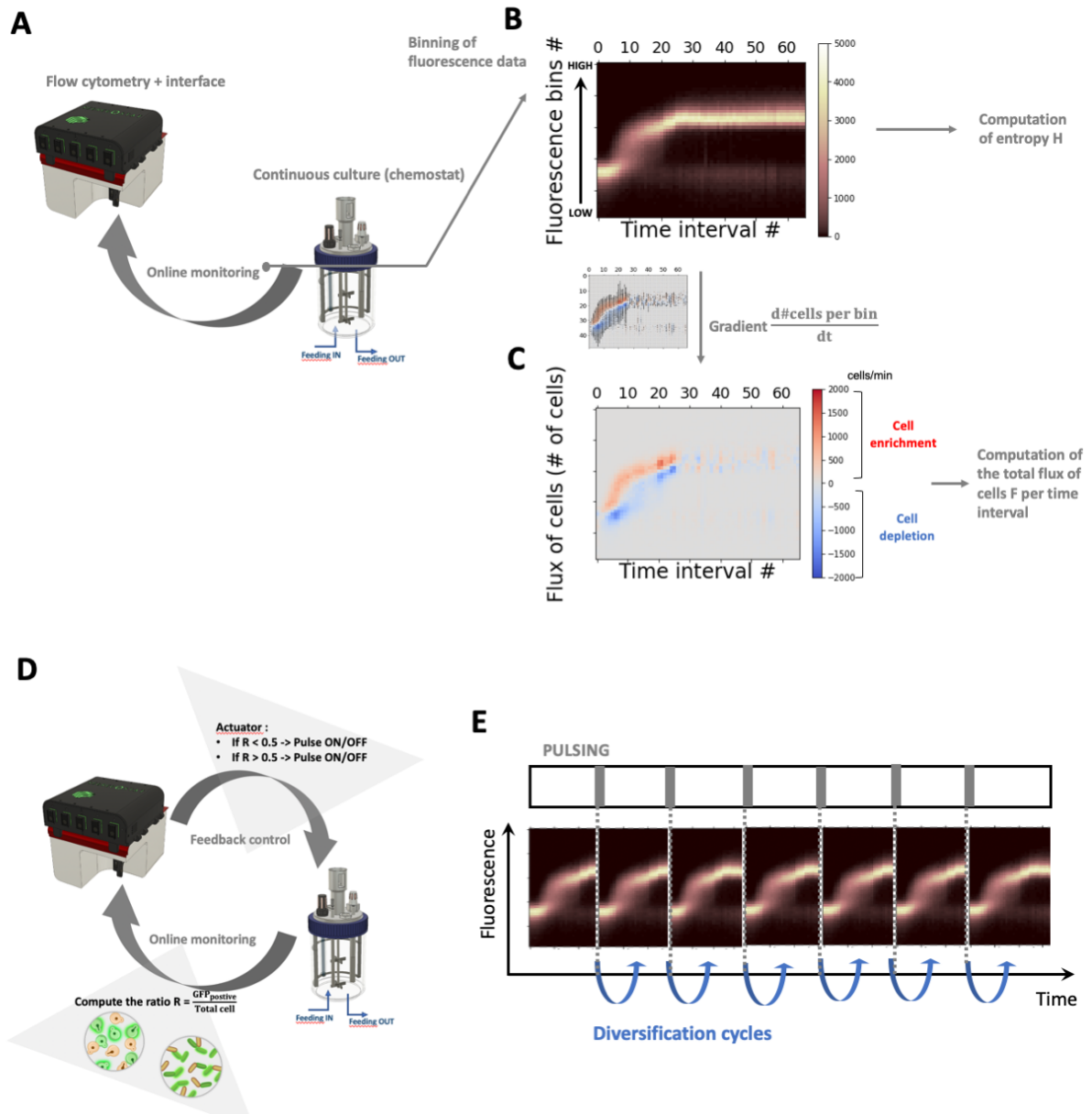
Cell populations can respond to environmental changes, and to the frequency of these changes, by adjusting phenotypes resulting from the activation of dedicated gene circuits<sup>1-3</sup>. This phenotypic plasticity has a lot of importance in microbial ecology, where the fitness of a cell population depends on a cost-benefit ratio between the sensing machinery needed for its activation and deactivation<sup>4</sup> and the activity of a given gene circuit. Thus, controlling the phenotype of cells has a lot of importance in various fields of research, such as bioproduction and synthetic biology, where coordinated gene expression is typically wanted<sup>5-10</sup>. Generating and controlling cell collective behavior is considered as a hallmark of synthetic biology<sup>9,11,12</sup>, and is now enabled by the parallel advances made at the level of cell cultivation procedures (i.e., microfluidics<sup>13</sup> and cell-machine interfaces<sup>14</sup>), as well as the manipulation of synthetic gene circuits<sup>15-17</sup>. Effective control of gene expressions and their underlying cellular functions can be achieved in cell populations<sup>5,18,19</sup> or individual cells within a population<sup>20,21</sup>. Different approaches can be used to coordinate/synchronize gene expression in cell populations. On the one hand, specific gene circuits can be designed in order to generate natural oscillations<sup>12,22</sup>. On the other hand, external forcing can be used for coordinating cellular responses<sup>6,23,24</sup>. According to this last approach, a given stimulus (e.g., chemical inducer<sup>20</sup>, light<sup>18,25</sup>...) is repeatedly applied at a given frequency and amplitude in order to entrain gene expression within a cell population. In this case, the effective transfer of information from the extracellular environment to the effector sites within cellular systems is of critical importance and can be corrupted by biological noise<sup>26-28</sup>. In silico experiments pointed out that specific environmental fluctuation frequencies could significantly reduce stochasticity in cell switching, giving rise to corresponding population diversification regimes<sup>1,28</sup>. However, the main factors affecting these diversification regimes are not known. This feature will be experimentally investigated in this work by looking at the temporal diversification profile of different types of cell populations in continuous culture. For this purpose, chemostat runs will be complemented by experiments conducted in Segregostat. The Segregostat relies on a cell-machine interface to generate environmental perturbations that are compatible with the diversification rate of the considered cell population (**Figure 1D**)<sup>19</sup>. This “rational” environmental forcing allows for the observation of several diversification cycles in one experimental run (**Figure 1E**). We applied this technology to look at the dynamics of cell populations with cellular functions leading to different fitness costs, i.e., utilization of alternative carbon sources (*E. coli*), general stress response (*E. coli* and *S. cerevisiae*), sporulation (*B. subtilis*) and activation of a T7-based expression system (*E. coli*). We determined that, based on the fitness cost associated to the cell switching mechanism (referred as switching cost or fitness cost in this study), three different population diversification regimes, with different level of sensitivity to environmental perturbations, can be observed.

## 2. Results

### 2.1. Environmental forcing triggered by cell switching dynamics leads to coordinated gene expression for diverse biological systems

Temporal diversification of cell populations has been followed based on chemostat cultivation of GFP reporter bearing strains and automated flow cytometry (FC) (**Figure 1A**). In order to

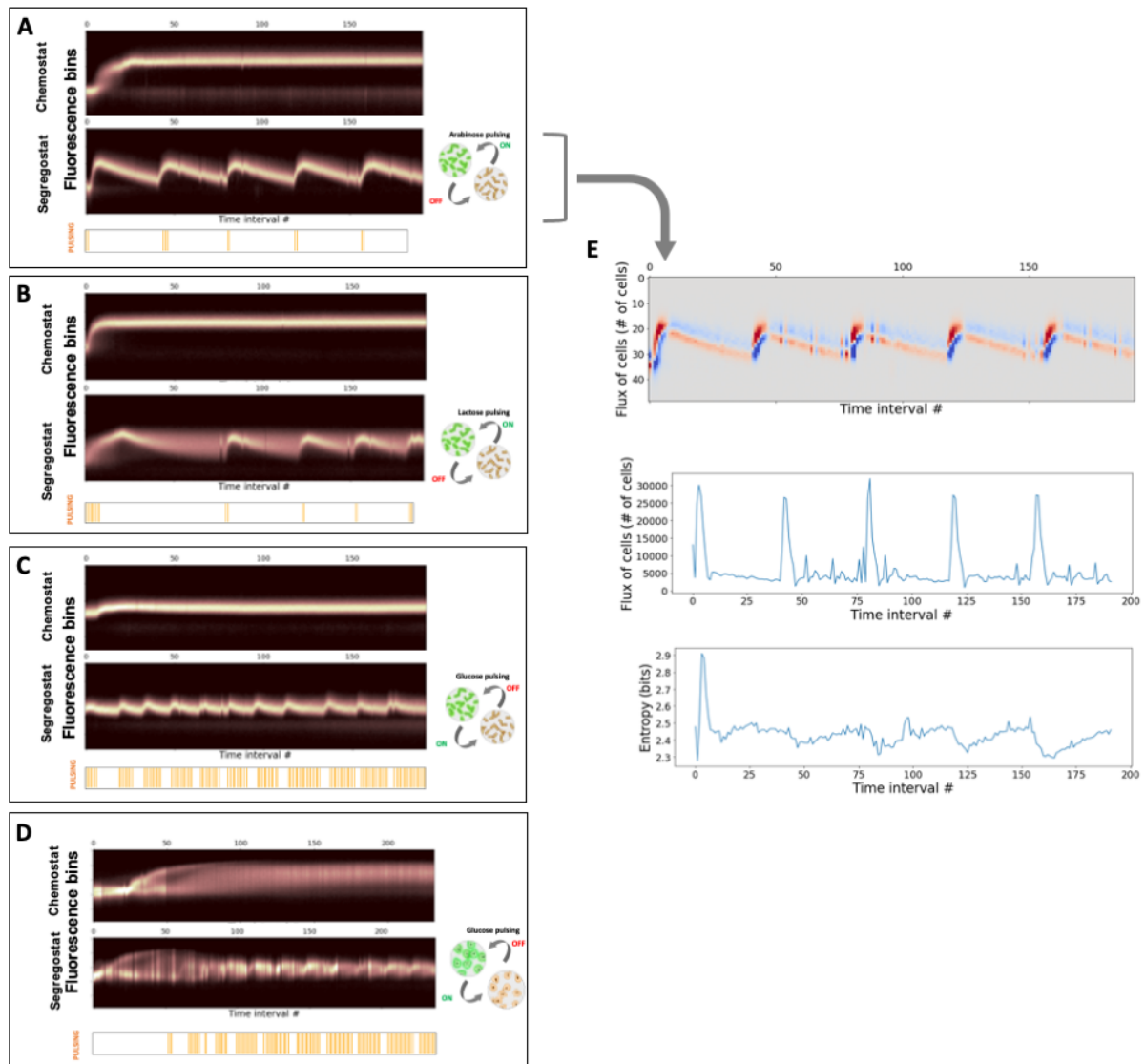
characterize population dynamics based on snapshot data, we developed a methodology to compute the fluxes of cells from a phenotype to another and the resulting degree of heterogeneity of the population, i.e., in our case, based on the measurement of information entropy (**Figure 1B&C**). This entropy is a measure derived from information theory allowing to compute the degree of heterogeneity of a population<sup>29</sup>. Briefly, GFP distributions obtained from automated FC measurements are binned and these bins are used to compute the cumulative probabilities of occurrence (**Supplementary file, Supplementary note 1, figure S1**). The benefit of this proxy is its independence from the mean of the distribution, by contrast with other noise proxies (e.g., Fano factor) that are known to be overestimated when the mean value increases<sup>30,31</sup>. Using entropy to analyze chemostat experiments, however, provides limited information about diversification dynamics. The main diversification process takes place during the transition between the batch and continuous phases of the culture. Therefore, we used a cell-machine interface allowing to produce several diversification cycles in a single experiment. This device is called the Segregostat and comprises a continuous cultivation device connected to an in-house online flow cytometry (FC) platform<sup>19</sup> (**Figure 1D**). This device lets us observe several diversification cycles per experiment, leading to a better characterization of the population switching dynamics (**Figure 1E**). Practically, the cells analyzed based on automated FC are clustered into a GFP negative and a GFP positive group. Depending on the gene circuits used, a pulse of inducer is applied when the ratio between the two clusters (i.e., either 50% or 10% of the total amount of cells in the desired state, depending on the cellular system considered) is not met.



**Figure 1:** scheme of the device used for characterizing cell population diversification dynamics. **A** Chemostat culture is monitored based on automated FC. **B** The fluorescence distribution acquired by FC are assembled into a time scatter plot. This time scatter plot is then further reordered into 45 fluorescence bins in order to compute the evolution of the entropy H of the population (**Supplementary File, Supplementary note 1**). **C** The binned data are further processed by applying a gradient in order to compute the fluxes of cells into the phenotypic space, leading to the quantification of the total fluxes of cells per time interval F. **D** Scheme of the Segregostat set-up. Pulses of nutrient are added in function of the ratio between GFP negative and GFP positive cells, as recorded by automated FC. **E** Expected evolution of a Segregostat experiment where, upon controlled environmental forcing, several diversification cycles can be generated.

This methodology was applied to map the diversification of cell populations upon chemostat and Segregostat cultivations. We began our analysis by considering two gene circuits involved in simple cellular processes in *E. coli*, i.e., the activation of the arabinose (**Figure 2A**) and lactose (**Figure 2B**) operons respectively. This type of cellular process is quite simple since it involves two inputs, i.e., the absence or limitation in glucose and the presence of either lactose or arabinose as an alternative carbon source<sup>4</sup>. Since these cultures were carried out in continuous mode, it was quite easy to ensure glucose limitation. Furthermore, the gene

circuitry behind the activation/deactivation of these two operons is well documented in the literature<sup>4,32</sup>.



**Figure 2:** time scatter plots binned into 50 cell clusters (fluorescence bins) for cultivations made in chemostat and Segregostat for **A** the *P<sub>araB</sub>:GFP* system in *E. coli*. **B** the *P<sub>lacZ</sub>:GFP* system in *E. coli*. **C** the *P<sub>bolA</sub>:GFP* system in *E. coli*. **D** the *P<sub>glc3</sub>:GFP* system in *S. cerevisiae*. Animated movies for the time evolution of the FC raw data for each system are available (**Supplementary Movies M1 to M8**). For the Segregostat experiments, environmental forcing has been performed based on nutrient pulsing (type of nutrient shown in the drawings for cell switching). **E** Computation of the flux of cells and the entropy for the *P<sub>araB</sub>:GFP* system cultivated in Segregostat mode (similar computation have been done for the other systems and can be found in **Supplementary information, Figure S2**). All the data are displayed in function of time intervals of 12 minutes, as the population snapshots have been acquired by automated FC.

A feature of these two systems is that, upon activation, these circuits do not result in a reduction of fitness. More often than not, that is not the case in natural gene circuits. We thus decided to investigate other gene circuits involved in more complex cellular processes, and known to lead to a higher switching cost (growth reduction). We first chose to consider the general stress response in *E. coli* and selected the promoter of the *bolA* gene as a representative  $\sigma^S$ -dependent system<sup>33,34</sup> (**Figure 2C**). In order to extend our analysis to

another biological system, we also selected a promoter involved in the accumulation of glycogen in yeast<sup>35</sup>, i.e.,  $P_{glc3}$ , as a representative reporter of bet-hedging in *S. cerevisiae*<sup>36,37</sup> (**Figure 2D**). Both genes are involved in very complex regulons, making it difficult to find an external trigger. However, both reporter systems are known to share common features in the sense that the growth of individual cells is anticorrelated with the level of expression of these general stress response reporters, making them very useful for analyzing cell collective behavior such as bet-hedging<sup>34,38</sup>. We then decided to use the external glucose concentration as the main actuator for these two systems. Glucose-limited chemostats were then run as reference conditions. For the Segregostat experiments, glucose was pulsed instead of lactose or arabinose, allowing to generate feast to famine environmental transitions.

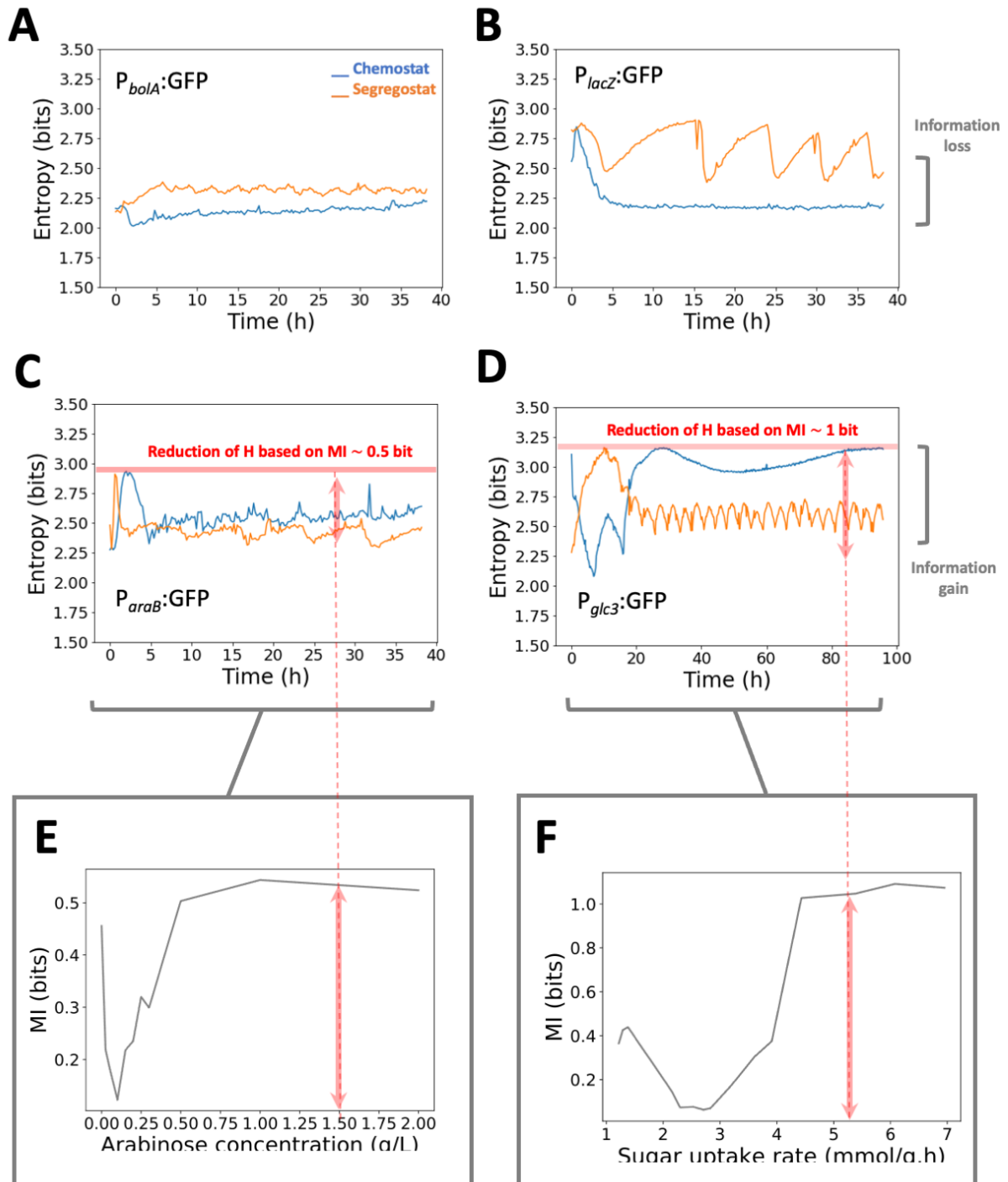
For all four cellular systems investigated, Segregostat cultivation led to entrainment and sustained oscillation of gene expression (**Figure 2A-D**). Based on the analysis of the entropy  $H(t)$  and the flux of cells  $F(t)$  over time, we observed that for all systems, entrainment phases were accompanied by an increased flux of cells switching to the alternative phenotype and a corresponding decrease of entropy  $H(t)$  at the time of pulsing (**Figure 2E** where the analysis is shown for the  $P_{araB}:GFP$  system; analyses for the other systems can be found in **Supplementary file, Supplementary note 1, figure S2**). However, the entropy increases during the relaxation phase (GFP dilution upon cell division). At this stage, the question is whether the different population profiles recorded based on automated FC can lead to different functionalities. We then decided to conduct an in-depth characterization of the heterogeneity of cell populations.

## 2.2. Coordinated gene expression doesn't necessarily lead to a more homogeneous cell population

We then compared the entropies recorded for each systems considered cultivated either in chemostat or Segregostat modes (**Figure 3A-D**). For the  $P_{araB}:GFP$  and  $P_{bolA}:GFP$  only slight differences of  $H$  were observed between chemostat and Segregostat conditions. However, for the two other systems, noticeable differences were observed. For the  $P_{lacZ}:GFP$ , the average entropy in the Segregostat is always higher than in chemostat. This is probably due to the leakiness of the promoter during the relaxation phase of the diversification cycles. This causes unwanted induction even when little to no inducer is present. On the other hand, we observed a significant reduction in entropy for the  $P_{glc3}:GFP$  system when cultivated in Segregostat, suggesting that cell entrainment can have an effect on the structure of the population. In order to understand why we do observe such differences in behavior between the cellular systems, we computed the mutual information (MI) between the environmental conditions and the activation of the target gene circuit for the  $P_{araB}:GFP$  and  $P_{glc3}:GFP$  systems. MI is a proxy derived from information theory<sup>26,27</sup> and involves the computation of the entropy of the cell population, as defined in the previous section. In short, MI tells us how much we can learn about the input (i.e., in our case the environmental stimulus used for entraining the cell population) from the output (i.e., in our case the distribution of GFP in cell population, the dispersion being quantified based on the entropy). Thus, in our case, MI is a proxy for information transfer efficiency between the inducer concentration and the cell population induction. The total entropy for each system was evaluated from the conditional probabilities obtained by exposing cell populations to different cultivation conditions (**Figure 3E&F**). Detailed description of the experiments done for determining the conditional probability distributions can be found in **Supplementary file, Supplementary note 2, Figures S3-S6**. MI

was obtained by subtracting the time-dependent entropies to the total entropies recorded for each system. For the  $P_{araB}$ :GFP system, MI is already relatively high in the chemostat leaving little room for improvement in the Segregostat (**Figure 3C&E**). It means that a reduction in entropy between the two cultivation modes has to be expected when the amount of information conveyed in chemostat condition is low (e.g., when cells cultivated in chemostat do not sense the inducer and, accordingly, do not activate the corresponding gene circuit). This is exactly what happened during the chemostat culture of the  $P_{glc3}$ :GFP system (**Figure 3D**). MI analysis had pointed out that there was still room for additional reduction in entropy (**Figure 3D&F**), and this was observed in Segregostat where glucose pulses reduced the average entropy (**Figure 3D**). This indicates that  $P_{glc3}$ :GFP is the only studied system where entrainment leads to a more homogeneous population. The next section investigates why such a phenomenon occurs.





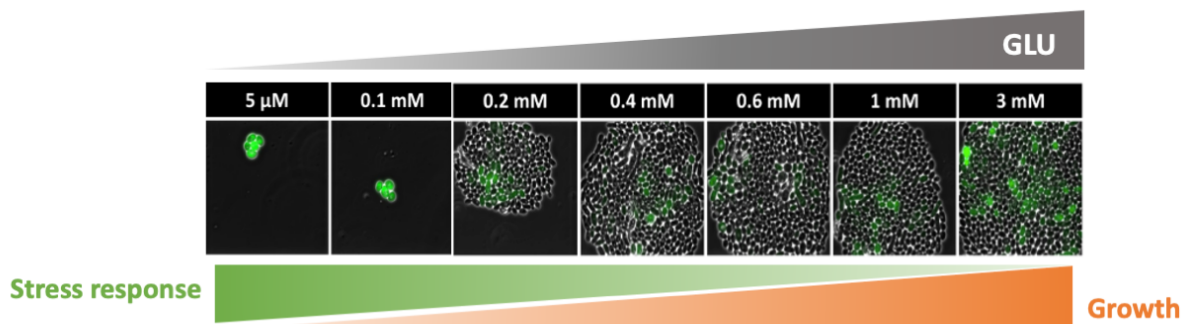
**Figure 3:** Comparative analysis of the entropy computed from the binned population data for four different cell systems. Comparative analysis of the entropy profile for the **A**  $P_{bolA}:GFP$  system in *E. coli*. **B**  $P_{lacZ}:GFP$  system in *E. coli*. **C**  $P_{araB}:GFP$  system in *E. coli*. **D**  $P_{glc3}:GFP$  system in *S. cerevisiae*. For the  $P_{araB}:GFP$  and  $P_{glc3}:GFP$  systems, the conditional probabilities, i.e., the GFP distribution of the population exposed at different environmental conditions, have been experimentally determined (**Supplementary information, Supplementary note 2**), allowing the computation of the mutual information (MI). **E** MI for the  $P_{araB}:GFP$  system exposed to different arabinose concentrations. The MI distribution for the  $P_{araB}:GFP$  system suggests that, at high arabinose concentration a gain of information of approximately 0.5 bit has to be expected (the value has been reported by a red line on **Figure 3C**). **F** MI for the  $P_{glc3}:GFP$  system exposed to different sugar uptake rates in an accelerostat cultivation device (**Supplementary information, Figure S4**). The MI distribution for the  $P_{glc3}:GFP$  system suggests that, at high glucose concentration a gain of information of approximately 1 bit has to be expected (the value has been reported by a red line on **Figure 3D**).



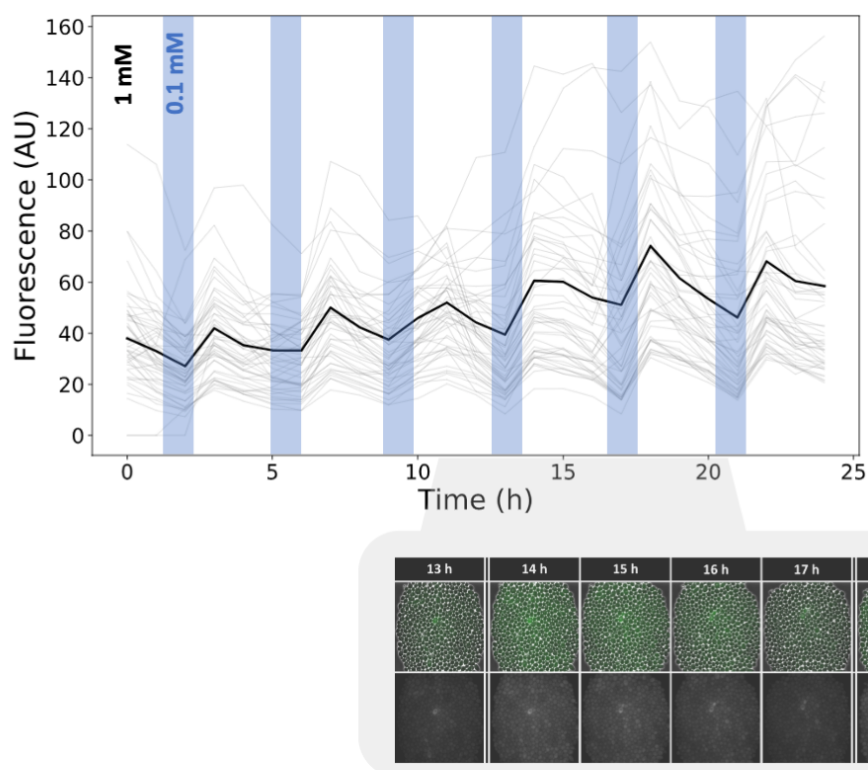
### 2.3. Fitness cost associated to cell phenotypic switching leads to a higher entropy at the population level that can be reduced upon environmental forcing

Stress response pathways in yeast are known to be involved in bet-hedging strategies, leading to a trade-off between growth and expression of stress-related genes<sup>37,38</sup>. The *glc3* gene belongs to this category. Accordingly, cells activating *glc3* exhibit reduced growth. This phenomenon has been characterized based on Microfluidics Single-Cell Cultivation (MSCC)<sup>39</sup> experiments allowing to expose yeast cells to tightly defined glucose concentrations (**Figure 4A**). Unlike with FC analyses where only population snapshots are captured, these experiments let us monitor cell traces and thus analyze the fitness cost (growth reduction upon switching) associated to the switching. It can be observed that, at low glucose concentration (< 0.2 mM), a single cell fully activates stress reporter and stops growing (**Supplementary movie M12**). At a higher glucose concentration, growth of the microcolonies is faster, but some stochastic switching events can be clearly observed, with cells suddenly expressing the fluorescence reporter and stopping their growth (**Supplementary movie M13**). In order to confirm the beneficial impact of Segregostat condition on the  $P_{glc3}$ :GFP, we used dynamic microfluidic single-cell cultivation (dMSCC)<sup>40</sup> where we applied environmental fluctuations between 0.1 and 1 mM of glucose, at the frequency recorded in the Segregostat conditions (**Figure 2D**). We observed a very homogenous gene expression pattern with cells turning green in perfect synchrony (**Figure 4B, Supplementary movie M14**). The growth of all cells was comparable, the stress level being kept at a low level thanks to the fluctuating environmental conditions. It seems that when phenotypic switching is associated with a loss of fitness, there is more stochasticity in the diversification pattern followed by the population. However, when the nutrient level is changed at a given frequency, switching can be kept under control, leading to a drastic reduction of the phenotypic heterogeneity of the cell population.

**A**



**B**

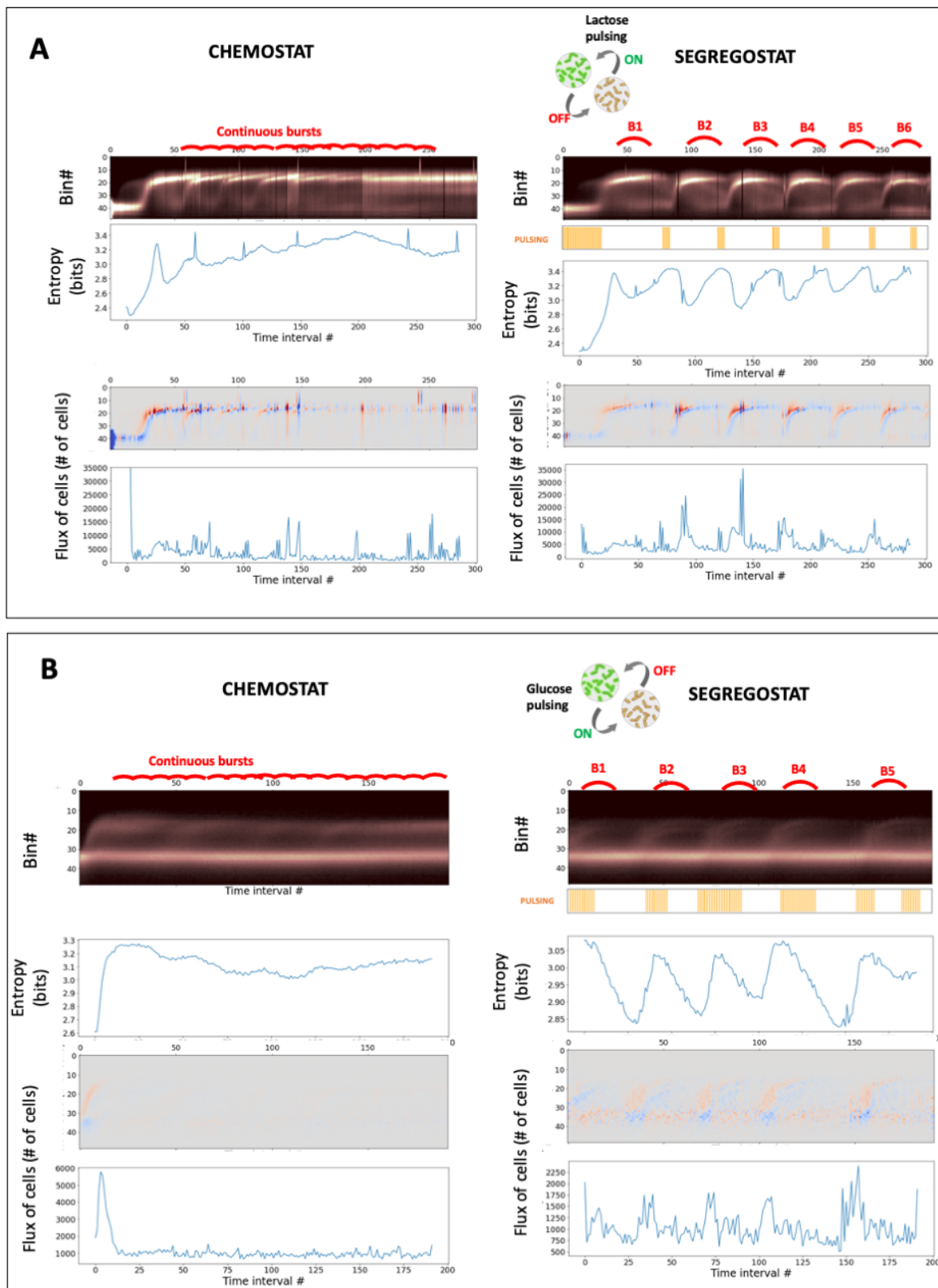


**Figure 4:** **A** Picture of yeast  $P_{glc3}::GFP$  microcolonies cultivated in a MSCC device at different glucose concentrations. Movies of a microcolony growing at a concentration of 0.1 mM and 1mM can be accessed in **Supplementary movie M12** and **M13** respectively). **B** Single cell traces of yeast  $P_{glc3}::GFP$  cells cultivated in a dMSCC device fluctuating between 1 and 0.1 mM of glucose ( $T_{1mM} = 3h$ ;  $T_{0.1mM} = 0.8h$ ). Between 10 and 40 cells have been tracked in four different cultivation chambers over two biological replicates (mean fluorescence is shown in bold). Pictures of a microcolony taken at regular time intervals are shown (**Supplementary movie M14**).

## 2.4. Phenotypic switching associated with extreme fitness cost gives rise to a bursty diversification regime

According to the results exposed in the previous section, a lower entropy in Segregostat has to be expected when the fitness cost associated with phenotypic switching is high. To validate this hypothesis, we investigated the diversification dynamics of two other systems known for their high impact on cellular fitness; the T7-based expression system in *E. coli* BL21 – a typical heterologous protein production platform – and the sporulation regulon in *B. subtilis* ( $P_{spoIIIE}$ :GFP). Indeed, for both, phenotypic switching leads to a drastic reduction of growth. Surprisingly, even in chemostat, FC profiles reveal bursts of diversification (**Figure 5A&B**). These bursts are the result of a subpopulation of cells deciding to switch and being washed out from the continuous cultivation device due to the associated fitness cost. During chemostat cultivation, bursts involving small fluxes of cells occur continuously and result in a high entropy at the population level. Then, upon environmental forcing based on Segregostat cultivation, the number of bursts is reduced and the fluxes of cells involved in the process are increased, leading to a substantial but temporary reduction of the entropy for the population (**Figure 5A&B**).

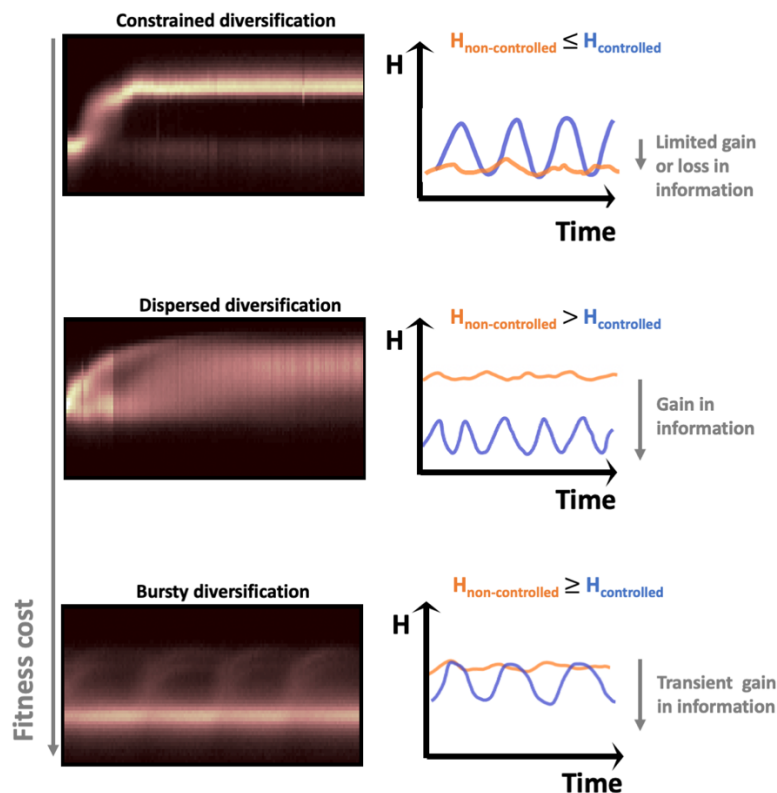
These results point out that the Segregostat reduces the average entropy of gene circuits with a high fitness cost despite very complex dynamics. It has been suggested in the literature that the stochasticity in cell switching is associated to its fitness cost and is important for the survival of the whole population. This feature is well illustrated in this case where a phenotype switch induces a dramatic loss of growth rate, leading to the wash-out of these cells in continuous cultivation conditions. However, this stochasticity can be reduced by applying environmental perturbations at a rate matching the phenotypic switching rate of cells. In the context of the T7 system, this approach led to the maximization of cells in the GFP positive state suggesting that it could be used for mitigating metabolic burden and maximizing productivity in continuous bioprocesses.



**Figure 5:** Temporal diversification profile for the  $P_{T7}$ :GFP system in *E. coli* (A) and the  $P_{sp01E}$ :GFP system (B) cultivated in chemostat and Segregostat modes. Only the continuous phase of the cultivation is shown on the graphs. Bursts of diversification are highlighted in red on the fluorescence bins data (the full diversification profile determined based on automated FC is available as **Supplementary movie M9-M10**). In both cases, the entropy  $H(t)$  and the fluxes of cells in the phenotypic space  $F(t)$  have been computed from the binned fluorescence data. All the data are displayed in function of time intervals, as the population snapshots have been acquitted by automated FC. One time interval corresponds to 12 minutes.

## 2.5. Fitness cost drives the appearance of different dynamical regimes with different levels of controllability

According to the datasets acquired from the different biological systems, we observed three types of diversification regimes named respectively constrained, dispersed and bursty (**Figure 6**). When the fitness cost associated to the switching is low or non-existent, we observe a constrained diversification regime where the population switches upon environmental change and adopts a homogeneous distribution. In that case, the stimulation of the population with controlled environmental pulsing does not homogenize it (since information transmission is already maximal in the non-controlled conditions, **Figure 3A&B**).

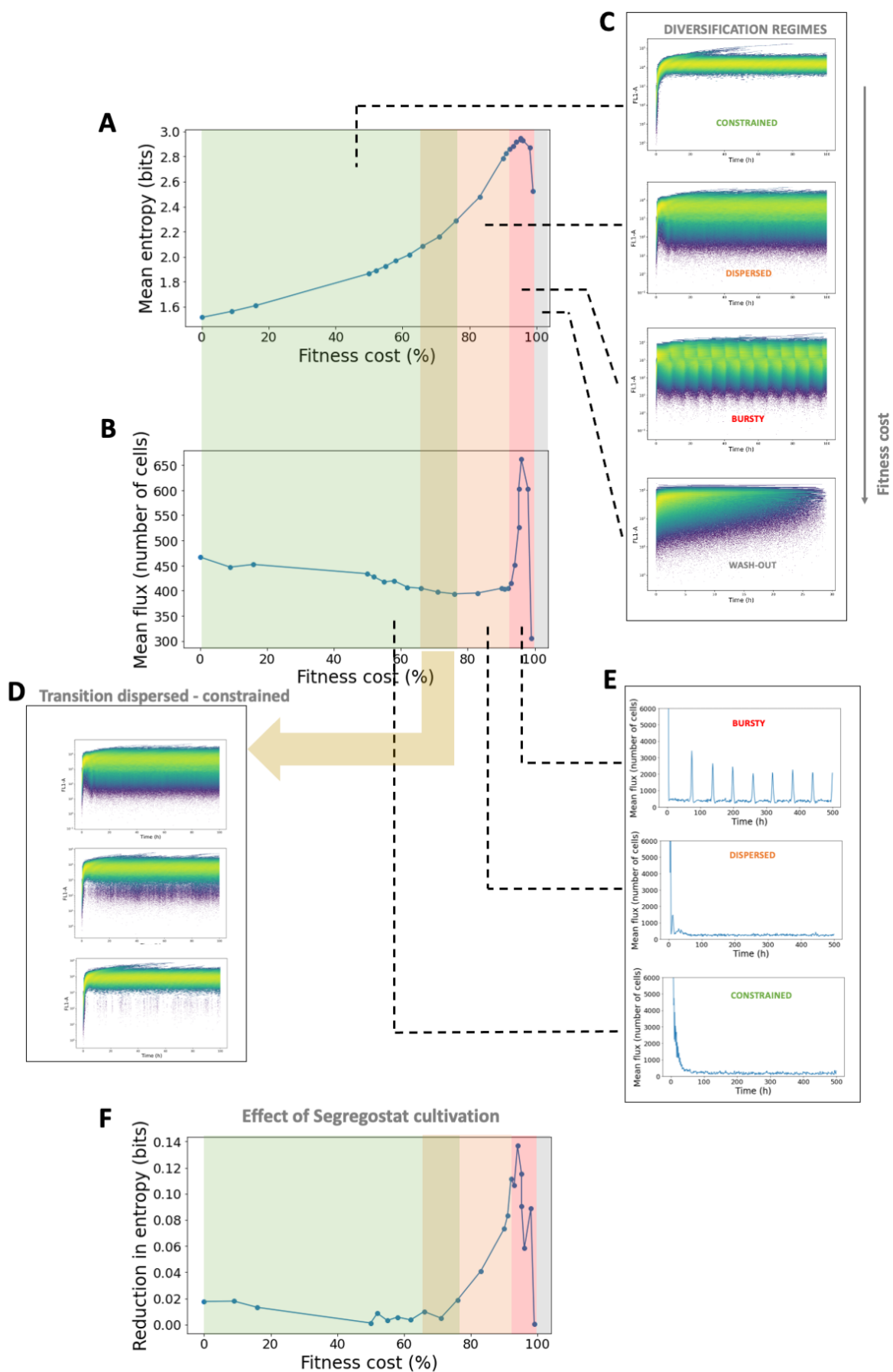


**Figure 6:** illustration of the three diversification regimes observed based on automated FC in function of the fitness cost. The consequences on the control of cell population are also shown based on the time evolution of the entropy of the population between non-controlled (i.e., obtained upon standard chemostat cultivation) and controlled (i.e., obtained upon periodic environmental stimulation of the population in the Segregostat) conditions.

When there is a fitness cost associated to the switch, what we call a dispersed diversification regime can be observed. In that case, cells react to environmental changes but then adopt a more heterogeneous population structure. In this case, the application of controlled

environmental perturbations allowed a substantial reduction in population heterogeneity. For the bursty diversification regime (higher fitness cost), cells switch in bursts, leading to a very heterogenous population structure. The application of controlled environmental perturbations reduces the number of bursts and increases the number of cells involved in these bursts, leading to a transient decrease in population heterogeneity.

It seems that the regime depends on the fitness cost associated with the phenotypic switching event. In order to verify that fitness cost is indeed the driver for the diversification pattern adopted by cell populations, we conducted *in silico* experiments. For this purpose, we considered the kinetic parameters obtained from the inference of the  $P_{glc3}$ :GFP system in yeast and conducted stochastic simulations based FlowStockS (**Supplementary file, Supplementary note 3**). We conducted 32 different chemostat simulations by varying only the value for the fitness cost and computed the entropy  $H$  (**Figure 7A**) and the fluxes of cells ( $F$ ) involved in phenotypic switching (**Figure 7B**). Solely based on the fitness cost associated to the switching, we were able to reproduce the three types of diversification regimes experimentally observed during the experiments (**Figure 3C**). Complete wash-out of the cells was observed for extreme fitness cost (>99% reduction in growth rate). We then wondered if we could observe clear transitions between the different regimes. Such transition was observed between the bursty and the dispersed regime based on the computation of the flux of cells  $F$ . Indeed, the bursty regime is marked by the appearance of a strong variation in flux of cell which is not observed for the other two regimes (**Figure 7E**). The transition between the dispersed and constrained regime is more progressive (**Figure 7D**). FlowStockS was also able to reproduce the behavior of the population under Segregostat cultivation, and the reduction in entropy upon environmental forcing was computed (**Figure 7F**). Again, reduction in entropy depended on the associated fitness cost and thus was observed for the dispersed and bursty regimes, in accordance with our experimental observations.





**Figure 7:** main outputs of the FlowStocKS simulations. **A** Evolution of the mean entropy recorded in chemostat in function of the fitness cost (here expression as the percentage of reduction of the initial growth rate prior phenotypic switching). **B** Evolution of the mean flux of cells recorded during chemostat experiments in function of the fitness cost associated to phenotypic switching (see **Figure 1** for more details about the computation of the flux of cells). **C** Selected simulated time scatter fluorescence plots illustrating the different diversification regimes observed at different fitness costs (the whole simulation dataset can be found as **Supplementary movie M15**). **D** Selected simulated time scatter fluorescence plots illustrating the progressive transition between the dispersed and constrained diversification regimes. **E** Selected simulated time evolution for the fluxes of cells recorded for different values of fitness cost. The bursty regime is characterized by the spontaneous generation of flux of cells (bursts) in chemostat cultivations. **F** Reduction of the entropy upon environmental forcing in function of the fitness cost associated with phenotypic switching. The entropy values have been computed by subtracting the mean entropy value recorded for chemostat experiment to the corresponding ones obtained in Segregostat.

### 3. Discussion

We used the Segregostat to better characterize cell population diversification dynamics by generating successive diversification cycles during the same experimental run. The basic principle behind this technology is to revert the environmental conditions when a fraction of cells (50% or 10% of the total population depending on the system investigated) crossed a predefined fluorescence threshold. This approach allows to maintain a cell population in a dynamic switching state during the experiment. Based on the analysis of the mutual information, i.e., the amount of information transferred from the extracellular conditions to the cell systems<sup>27,29,42</sup>, we determined that for the  $P_{araB}$ :GFP system, the chemostat drives a similar amount of information than the Segregostat. On the other hand, we observed a drastic reduction in entropy when entraining stress-related systems, such as the  $P_{glc3}$ :GFP system in yeast, in the Segregostat. In this case, we determined that the high entropy observed in the chemostat was related to a trade-off between growth and gene expression<sup>36–38,43–45</sup>, which was further confirmed based on microfluidics experiment, and by considering two additional systems where phenotypic switching induced a huge fitness cost i.e., the sporulation system in *B. subtilis* and the T7-based expression system in *E. coli*.

Based on all the data accumulated by automated FC for six different biological systems, we found that cell populations diversified according to three distinct regimes associated to increasing fitness costs, i.e., constrained, dispersed and bursty. The most noticeable difference between these regimes is the level of entropy of the cell population, the entropy being a measure derived from information theory giving a robust estimate of population dispersion<sup>29,41</sup>. The lowest entropy values were observed for the constrained regime and the highest ones were observed for dispersed and bursty regimes. The other difference was observed upon the cultivation of cell populations under fluctuating environmental conditions in the Segregostat were a reduction of entropy compared to chemostat cultivation was observed for the dispersed and bursty regimes but not the constrained. Taken altogether, the data suggested that cell population diversification dynamics are mainly driven by the fitness cost associated with the phenotypic switching mechanism, a higher cost giving rise to the dispersed and bursty regimes. All these observations were confirmed based on stochastic simulations (FlowStocKS), suggesting that the proposed diversification framework could be generalized for characterizing diversification dynamics for any kind of cellular system.

Harnessing phenotypic heterogeneity of microbial populations has been the subject of much research, leading to the design of various technologies aiming at homogenizing gene expression in cell populations<sup>46</sup>. We have shown that the level of diversification of microbial

populations cultivated in continuous bioreactors depends mainly on the fitness cost. Since many applications involve gene circuits whose activation leads to substantial burden for the cellular system<sup>48,49</sup>, active diversification processes have to be expected in a number of cases<sup>50</sup>. Our data point out that, under these circumstances, phenotypic diversity could be promoted, instead of being reduced, by using environmental forcing to provide cell populations with robust temporal patterns in gene expression. For example, bursty diversification profiles have been observed for two cellular systems exhibiting high switching cost. According to this regime, marked cycles of diversification can be observed even in chemostat cultures. These cycles are due to the rapid switching (burst) of a fraction of the population that lower the average growth rate of the population and are washed out of the system. These cells are then replaced by the next burst of diversification, starting from a subpopulation of non-diversified cells. This type of temporal profile has been previously observed, but based on spatially organized cells equipped with synthetic circuits<sup>7,51,52</sup>. Here, we show that it is possible to reproduce such a complex but organized diversification profile with cells in suspension in a bioreactor and that the complex dynamics behind phenotypic heterogeneity are linked to the fitness cost associated to the switch.

## 4. Methods

### 4.1. Stains and plasmids

The analyses of alternative carbon source utilization and stress response in *E. coli* were done based on a *E. coli* W3110 backbone and kanamycin resistance bearing plasmids originating from the Zaslaver collection (i.e.,  $P_{araB}::GFPmut2$ ,  $P_{lacZ}::GFPmut2$  and  $P_{bolA}::GFPmut2$ )<sup>53</sup>. For investigating the T7 induction system response, we used *E. coli* BL21 (DE3) carrying pET28:GFP<sup>56</sup>. To observe the starvation response, we used *S. cerevisiae* CEN-PK 117D background with the chromosomal integration of a reporter cassette  $P_{glc3}::eGFP$ <sup>35,54</sup>. The *E. coli* W3110  $\Delta araBAD$  strain used for determining the conditional entropy of the  $P_{araB}::GFP$  system was constructed through CRISPR-cas9 enhanced lambda red phage mediated homologous recombination as described by Jiang *et al.*<sup>55</sup> (the primer sequences are described in **Supplementary File, Table S1**). Finally, we monitored the early stage of the sporulation process in *B. subtilis* 168 based on a chromosomal integration of  $P_{spoilE}::GFPmut2$  (kindly provided by Denise Wolf and Adam Arkin)<sup>57</sup>.

### 4.2. Cultivation conditions and Segregostat procedure

Bacteria (*E. coli* and *B. subtilis*) precultures and cultures have been performed in a defined mineral salt medium containing (in g/l):  $K_2HPO_4$  14.6;  $NaH_2PO_4 \cdot 2H_2O$  3.6;  $Na_2SO_4$  2;  $(NH_4)_2SO_4$  2.47;  $NH_4Cl$  0.5;  $(NH_4)_2$ -H-citrate 1; glucose 5, thiamine 0.01, antibiotic 0.1. Thiamine is sterilized by filtration (0.2 mg/l). The medium is supplemented with 3 ml/l of a trace element solution, 3ml/l of a  $FeCl_3 \cdot 6H_2O$  (16.7 g/l), 3 ml/l of an EDTA (20.1 g/l) and 2ml/l of a  $MgSO_4$  solution (120 g/l). The trace element solution contains (in g/l):  $CoCl_2 \cdot H_2O$  0.74;  $ZnSO_4 \cdot 7H_2O$  0.18;  $MnSO_4 \cdot H_2O$  0.1;  $CuSO_4 \cdot 5H_2O$ ;  $CoSO_4 \cdot 7H_2O$ . Filtered sterilized kanamycin (50 mg/l) was added for plasmid maintenance in *E. coli*. *S. cerevisiae* cultures and precultures have been performed based on a Verduyn mineral medium<sup>49</sup>. The precultures were performed in 1L baffled flask overnight either at 37 °C (bacteria) or 30°C (*S. cerevisiae*) at a shaking speed of

150 rpm and used to start the batch phase in a lab-scale stirred bioreactor (Biostat B-Twin, Sartorius) total volume: 2 l; working volume: 1 l at an initial OD<sub>600</sub> of 0.5. Once the batch phase was over (typically after 5, 8 and 15 h for *E. coli*, *B. subtilis*, *S. cerevisiae* respectively). The operating conditions used for the various cell systems are provided in **Supplementary information, Supplementary note 4, Table S3**.

Data was collected through online flow cytometry during the experiments, and in Segregostat experiments, the actuator was pulsed based on both the observed distribution and a pre-defined set-point. The Segregostat platform has been described earlier<sup>10</sup>. Briefly, every 12 minutes, a sample is automatically taken from the bioreactor, diluted, and analyzed in a flow cytometer (BD Accuri C6, BD Biosciences) with an FSC-H analysis threshold of 20,000 for bacteria and 80,000 for *S. cerevisiae*. During the chemostat experiment, glucose and, if applicable, an alternative carbon source was fed together, while in the Segregostat experiment, glucose was the sole carbon source in the feed. In Segregostat, a feedback control loop, which includes a custom MATLAB script based on FC data, activates a pump to pulse an actuator. For *S. cerevisiae*, *E. coli* stress response, and *B. subtilis* sporulation, the actuator pulses glucose, while for *E. coli* alternative carbon source utilization, the actuator pulses lactose and arabinose. For all systems except for the sporulation control, a control threshold of 50/50 was utilized. However, for the sporulation control, the regulation was triggered once the fluorescence threshold was exceeded by more than 20% of the cells. This decision was made based on the irreversibility of the sporulation process, which required earlier intervention.

### **4.3. Microfluidic cultivations and time lapse microscopy**

All devices and parameters were the same as previously<sup>59</sup>, excepted few modifications. *S. cerevisiae* cells have been cultivated in the dynamic microfluidic single-cell cultivation (dMSCC) chips provided by Alexander Grünberger's lab (reference 24W, chambers size : 80 μm x 80 μm x 850 nm)<sup>40</sup>. Their design enables the simultaneous use of 2 cultivation media. They are separated in three zones: 2 control zones, fed by either one or the other medium, and a switching zone fed in alternance by the two media. Diverse combinations of Verduyn medium with different glucose concentrations have been tested (i.e., 0.4 mM - 0.6 mM, 0.2 mM – 0.8 mM, 0.1 mM - 1 mM and 5 μM – 3 mM). To approach the Segregostat conditions, the duration of the switching zone feeding with the medium containing more (or less) glucose corresponds to the mean period with (or without) pulsing when population is controlled in the segregostat (i.e., 180 min (or 48 min)). High precision pressure pumps (line-up series, Fluigent, Le Kremlin-Bicêtre, France) were used to precisely control medium flow rate. The temperature was set at 30°C. The chambers were inoculated with one or two cells by flushing the device with a cell suspension (OD<sub>600</sub> between 0.4 and 0.5). At least 6 cultivation chambers were selected manually for each zone of the dMSCC chips. Microscopy images were acquired during 72 hours using a Nikon Eclipse Ti2-E inverted automated epifluorescence microscope (Nikon Eclipse Ti2-E, Nikon France, France) equipped with a DS-Qi2 camera (Nikon camera DSQi2, Nikon France, France), a 100× oil objective (CFI P-Apo DM Lambda 100× Oil (Ph3), Nikon France, France). The GFP-3035D cube (excitation filter: 472/30 nm, dichroic mirror: 495 nm, emission filter: 520/35 nm, Nikon France, Nikon) was used to measure GFP. The phase contrast images were recorded with an exposure time of 300 milliseconds and an illuminator's intensity of 30%. The GFP images were recorded with an exposure time of 500 milliseconds and an illuminator's intensity

of 2% (SOLA SE II, Lumencor, USA). During the first 48 hours, GFP and phase contrast images were acquired every hour. During the 24 last hours, phase contrast images were acquired every 6 minutes and GFP images every hour (or every 20 minutes for one experiment with media 5  $\mu$ M – 3 mM glucose). The optical parameters and the time-lapse were managed with the NIS-Elements Imaging Software (Nikon NIS Elements AR software package, Nikon France, France). The single-cell data have been computed for the 24 last hours of the time lapse for at least 3 chambers per condition (i.e., zone of the dMSCC device). The cell-segmentation of the images and the measure of single-cell mean GFP intensity were performed using either the Python GUI<sup>60</sup> or the Matlab algorithm of Wood and Doncic<sup>61</sup>. This last was also used for cell tracking. In this case, seeds for segmentations have been corrected manually and at least 10 cells per microfluidic cultivation chamber (well segmented and tracked all the time lapse long) were selected manually.

#### 4.4. Modelling cell population dynamics based on FlowStoCKS

The aim of this simulation toolbox is to be able to represent with high fidelity the population snapshots and dynamics captured based on automated flow cytometry. Briefly, population dynamics is modelled based on a set of ODEs representing the time evolution of biomass and substrates according to a Monod kinetics in a continuous cultivation device. From the global population, a given number of cells (approx. 10,000, some cells being washed-out during the simulation) are considered for generating a stochastic process. For these cells, phenotypic switching is modelled according to a Markov chain process driving the synthesis and degradation of GFP. Cell growth and division are taken into account for computing the GFP content. Upon switching, cells may encounter a fitness cost depending on an inhibitory kinetics. The data are then fitted to a seven-decade fluorescence scale in order to fit to the automated FC data. Detailed information, including parameters and equations settings, are provided as supplementary information (**Supplementary file, Supplementary note 3**).

#### Data and code availability

The data sets and the supplementary movies are available on gitlab (<https://gitlab.uliege.be/mipi/published-software/mipi-model-and-simulation-database/-/tree/main/Fitness%20cost%20associated%20with%20cell%20phenotypic%20switching%20drives%20population%20diversification%20dynamics%20and%20controllability>). The FlowStoCKS toolbox is on <https://gitlab.uliege.be/mipi/published-software/mbms-toolbox/-/tree/main/FlowStoCKS>.

#### Supplementary information

##### Supplementary Information file

**Supplementary movie M1** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{araB}$ :GFP system in *E. coli* cultivated in chemostat

**Supplementary movie M2** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{lacZ}$ :GFP system in *E. coli* cultivated in chemostat

**Supplementary movie M3** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{bolA}$ :GFP system in *E. coli* cultivated in chemostat

**Supplementary movie M4** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{glc3}$ :GFP system in *S. cerevisiae* cultivated in chemostat

**Supplementary movie M5** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{araB}$ :GFP system in *E. coli* cultivated in Segregostat

**Supplementary movie M6** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{lacZ}$ :GFP system in *E. coli* cultivated in Segregostat

**Supplementary movie M7** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{bolA}$ :GFP system in *E. coli* cultivated in Segregostat

**Supplementary movie M8** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{glc3}$ :GFP system in *S. cerevisiae* cultivated in Segregostat

**Supplementary movie M9** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the  $P_{spoIIIE}$ :GFP system in *B. subtilis* cultivated in chemostat (first 40h), followed by Segregostat cultivation.

**Supplementary movie M10** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the pET28:GFP system in *E. coli* cultivated in chemostat

**Supplementary movie M11** Time evolution of dotplots FSC-A (Forward Scatter signal) in function of FL1-A (GFP fluorescence signal) obtained based on automated FC for the pET28:GFP system in *E. coli* cultivated in Segregostat

**Supplementary movie M12** Cultivation of *S. cerevisiae* carrying a  $P_{glc3}$ :GFP reporter in a MSCC device. Chemically defined medium is constantly perfused into the chamber with a glucose concentration of 0.1mM, leading to a drastic reduction in growth of the colony and full activation of the  $P_{glc3}$ :GFP reporter.



**Supplementary movie M13** Cultivation of *S. cerevisiae* carrying a  $P_{glc3}$ :GFP reporter in a MSCC device. Chemically defined medium is constantly perfused into the chamber with a glucose concentration of 1mM.

**Supplementary movie M14** Cultivation of *S. cerevisiae* carrying a  $P_{glc3}$ :GFP reporter in a dMSCC device. Cultivation conditions are periodically switched between chemically defined media containing glucose at a concentration of 0.1 and 1 mM respectively. The feast-to-famine transitions have been applied with predefined durations i.e.,  $T_{0.1mM} = 0.8$  h and  $T_{1mM} = 3$ h, for mimicking the Segreostat conditions.

**Supplementary movie M15** FlowStockS simulations of chemostat cultivations under variable fitness cost.

## Acknowledgement

LH and VV are supported by a FRIA grant provided by the “Fonds de la Recherche Scientifique” FRS-FNRS, from the Walloon region of Belgium. JAM is supported by a post-doctoral grant provided by the Service Public de Wallonie (SPW) and the H2020 program of the European commission (Era-Cobiotech project Contibio). MD is supported by a PhD grant provided by the FRS-FNRS and the H2020 program of the European commission (Era-Net Aquatic Pollutant project ARENA). FD received funding from a research grant provided by the Service Public de Wallonie (SPW) and the H2020 program of the European commission (Era-Cobiotech project ComRaDes).

## References

1. Thattai, M. & van Oudenaarden, A. Stochastic gene expression in fluctuating environments. *Genetics* **167**, 523–530 (2004).
2. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
3. Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol Rev* **45**, (2021).
4. Alon, U. *An introduction to systems biology*. (CRC Press, 2020).
5. Nguyen, T. M. *et al.* Reducing phenotypic instabilities of a microbial population during continuous cultivation based on cell switching dynamics. *Biotechnol Bioeng* (2021) doi:10.1002/bit.27860.
6. Pérez-García, S. *et al.* Synchronization of gene expression across eukaryotic communities through chemical rhythms. *Nat Commun* **12**, 4017 (2021).
7. Din, M. O. *et al.* Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**, 81–85 (2016).
8. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
9. Mondragón-Palomino, O., Danino, T., Selimkhanov, J., Tsimring, L. & Hasty, J. Entrainment of a population of synthetic genetic oscillators. *Science* **333**, 1315–1319 (2011).
10. Wagemakers, A. *et al.* Entraining synthetic genetic oscillators. *Chaos* **19**, 033139 (2009).
11. Wang, R. & Chen, L. Synchronizing genetic oscillators by signaling molecules. *J Biol Rhythms* **20**, 257–269 (2005).

12. Potvin-Trottier, L., Lord, N. D., Vinnicombe, G. & Paulsson, J. Synchronous long-term oscillations in a synthetic gene circuit. *Nature* **538**, 514–517 (2016).
13. Grunberger, A., Wiechert, W. & Kohlheyer, D. Single-cell microfluidics: opportunity for bioprocess development. *Curr Opin Biotechnol* **29**, 15–23 (2014).
14. Lugagne, J.-B. & Dunlop, M. J. Cell-machine interfaces for characterizing gene regulatory network dynamics. *Curr Opin Syst Biol* **14**, 1–8 (2019).
15. Levine, J. H., Lin, Y. & Elowitz, M. B. Functional roles of pulsing in genetic circuits. *Science* **342**, 1193–200 (2013).
16. Wong, W. W. & Liao, J. C. The design of intracellular oscillators that interact with metabolism. *Cell Mol Life Sci* **63**, 1215–20 (2006).
17. Din, M. O., Martin, A., Razinkov, I., Csicsery, N. & Hasty, J. Interfacing gene circuits with microelectronics through engineered population dynamics. *Sci Adv* **6**, eaaz8344 (2020).
18. Miliadis-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P. & Khammash, M. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat Commun* **7**, 12546 (2016).
19. Sassi, H. *et al.* Segregostat: a novel concept to control phenotypic diversification dynamics on the example of Gram-negative bacteria. *Microbial Biotechnology* **12**, 1064–1075 (2019).
20. Lugagne, J.-B. *et al.* Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nat Commun* **8**, 1671 (2017).
21. Rullan, M., Benzinger, D., Schmidt, G. W., Miliadis-Argeitis, A. & Khammash, M. An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation. *Mol Cell* **70**, 745-756.e6 (2018).
22. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
23. Bertaux, F. *et al.* Enhancing bioreactor arrays for automated measurements and reactive control with ReacSight. *Nat Commun* **13**, 3363 (2022).
24. Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat Commun* **9**, 3521 (2018).
25. Melendez, J. *et al.* Real-time optogenetic control of intracellular protein concentration in microbial cell cultures. *Integr Biol (Camb)* **6**, 366–372 (2014).
26. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).
27. Tostevin, F. & ten Wolde, P. R. Mutual information between input and output trajectories of biochemical networks. *Phys Rev Lett* **102**, 218101 (2009).
28. Tan, C., Reza, F. & You, L. Noise-limited frequency signal transmission in gene circuits. *Biophys J* **93**, 3753–3761 (2007).
29. Bowsher, C. G. & Swain, P. S. Environmental sensing, information transfer, and cellular decision-making. *Curr Opin Biotechnol* **28**, 149–155 (2014).
30. Silander, O. K. *et al.* A genome-wide analysis of promoter-mediated phenotypic noise in *Escherichia coli*. *PLoS Genet* **8**, e1002443 (2012).
31. Sanchez, A. & Golding, I. Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**, 1188–93 (2013).
32. Aidelberg, G. *et al.* Hierarchy of non-glucose sugars in *Escherichia coli*. *BMC Syst Biol* **8**, 133 (2014).
33. Sampaio, N. M. V., Blassick, C. M., Andreani, V., Lugagne, J.-B. & Dunlop, M. J. Dynamic gene expression and growth underlie cell-to-cell heterogeneity in *Escherichia coli* stress response. *Proc Natl Acad Sci U S A* **119**, e2115032119 (2022).
34. Patange, O. *et al.* *Escherichia coli* can survive stress by noisy growth modulation. *Nat Commun* **9**, 5333 (2018).



35. Zid, B. M. & O'Shea, E. K. Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. *Nature* **514**, 117–121 (2014).
36. Ziv, N., Siegal, M. L. & Gresham, D. Genetic and nongenetic determinants of cell growth variation assessed by high-throughput microscopy. *Mol Biol Evol* **30**, 2568–78 (2013).
37. Levy, S. F., Ziv, N. & Siegal, M. L. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol* **10**, e1001325 (2012).
38. Ziv, N., Shuster, B. M., Siegal, M. L. & Gresham, D. Resolving the Complex Genetic Basis of Phenotypic Variation and Variability of Cellular Growth. *Genetics* **206**, 1645–1657 (2017).
39. Grunberger, A. *et al.* A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level. *Lab Chip* **12**, 2060–2068 (2012).
40. Täuber, S., Golze, C., Ho, P., von Lieres, E. & Grünberger, A. dMSCC: a microfluidic platform for microbial single-cell cultivation of *Corynebacterium glutamicum* under dynamic environmental medium conditions. *Lab Chip* **20**, 4442–4455 (2020).
41. Levchenko, A. & Nemenman, I. Cellular noise and information transmission. *Curr Opin Biotechnol* **28**, 156–164 (2014).
42. Henrion, L., Delvenne, M., Bajoul Kakahi, F., Moreno-Avitia, F. & Delvigne, F. Exploiting Information and Control Theory for Directing Gene Expression in Cell Populations. *Front Microbiol* **13**, 869509 (2022).
43. Basan, M. *et al.* A universal trade-off between growth and lag in fluctuating environments. *Nature* **584**, 470–474 (2020).
44. Wolf, D. M. Diversity in times of adversity : probabilistic strategies in microbial survival games. *Journal of theoretical biology* **234**, 227–253 (2005).
45. Perkins, T. J. & Swain, P. S. Strategies for cellular decision-making. *Mol Syst Biol* **5**, 326 (2009).
46. Binder, D. *et al.* Homogenizing bacterial cell factories: Analysis and engineering of phenotypic heterogeneity. *Metabolic Engineering* **42**, 145–156 (2017).
47. Fiore, G., Perrino, G., di Bernardo, M. & di Bernardo, D. In Vivo Real-Time Control of Gene Expression: A Comparative Analysis of Feedback Control Strategies in Yeast. *ACS Synth Biol* **5**, 154–162 (2016).
48. Ceroni, F., Blount, B. A. & Ellis, T. Sensing the Right Time to Be Productive. *Cell Syst* **3**, 116–117 (2016).
49. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat Methods* **15**, 387–393 (2018).
50. Borkowski, O., Ceroni, F., Stan, G.-B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr Opin Microbiol* **33**, 123–130 (2016).
51. Liao, M. J., Din, M. O., Tsimring, L. & Hasty, J. Rock-paper-scissors: Engineered population dynamics increase genetic stability. *Science* **365**, 1045–1049 (2019).
52. Bittihn, P., Didovyk, A., Tsimring, L. S. & Hasty, J. Genetically engineered control of phenotypic structure in microbial colonies. *Nat Microbiol* **5**, 697–705 (2020).
53. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3**, 623–628 (2006).
54. García-Timmermans, C. *et al.* Raman Spectroscopy-Based Measurements of Single-Cell Phenotypic Diversity in Microbial Populations. *mSphere* **5**, (2020).
55. Jiang, Y. *et al.* Multigene editing in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* **81**, 2506–2514 (2015).
56. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci U S A* **110**, 5028–5033 (2013).
57. Wolf, D. M. *et al.* Memory in microbes: quantifying history-dependent behavior in a bacterium. *PLoS One* **3**, e1700 (2008).

58. Verduyn, C., Postma, E., Scheffers, W. A. & Van Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8**, 501–17 (1992).
59. Martinez, J. A. *et al.* Controlling microbial co-culture based on substrate pulsing can lead to stability through differential fitness advantages. *PLoS Comput Biol* **18**, e1010674 (2022).
60. Dietler, N. *et al.* A convolutional neural network segments yeast microscopy images with high accuracy. *Nat Commun* **11**, 5723 (2020).
61. Wood, N. E. & Doncic, A. A fully-automated, robust, and versatile algorithm for long-term budding yeast segmentation and tracking. *PLoS One* **14**, e0206395 (2019).

## Cell populations adopt distinct phenotypic diversification dynamics exhibiting different levels of controllability depending on fitness cost

Lucas Henrion<sup>1</sup>, Juan Andres Martinez<sup>1</sup>, Vincent Vandembroucke<sup>1</sup>, Mathéo Delvenne<sup>1</sup>, Samuel Telek<sup>1</sup>, Andrew Zicler<sup>1</sup>, Alexander Grünberger<sup>2</sup>, Frank Delvigne<sup>1\*</sup>

<sup>1</sup>Terra research and teaching centre, Microbial Processes and Interactions (MiPI), Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

<sup>2</sup>Microsystems in Bioprocess Engineering, Institute of Process Engineering in Life Sciences, Karlsruhe Institute of Technology, Karlsruhe, Germany

\*Corresponding author: [F.Delvigne@uliege.be](mailto:F.Delvigne@uliege.be)

## Supplementary Information

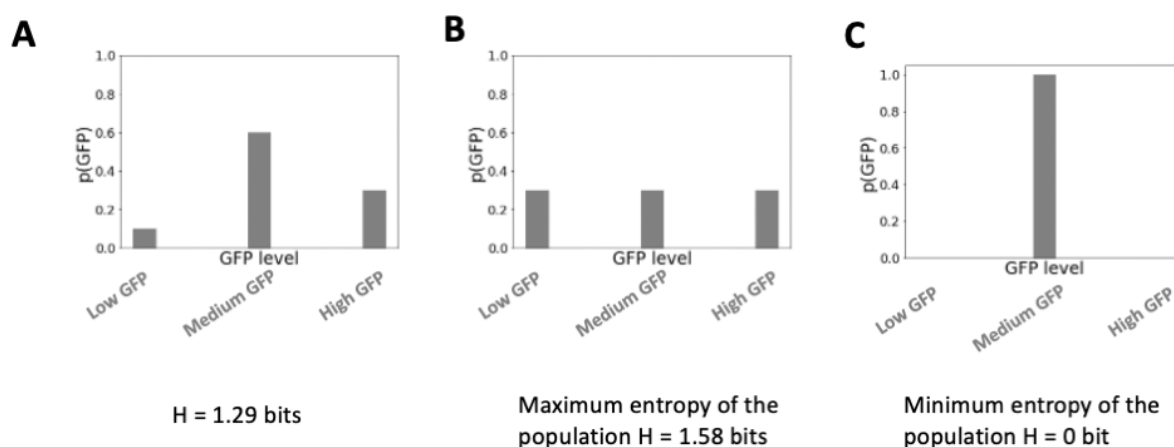
### Supplementary note 1 Determination of the entropy (H) of the population from automated FC data

Information theory has been used in this work for characterizing the response of cell population to environmental perturbations. This theory involves the computation of entropy  $H$ , which can be regarded as a measure of uncertainty about the response of the cell population (output) in function of the environmental stimulation (input). This input-output relationship, which constitute the basis of information theory, will be detailed in **Supplementary note 2**. In this note, we'll concentrate on the description of the entropy  $H$  as a measure of the level of heterogeneity of the population. Entropy can be considered as a measure of uncertainty about the outcome of a draw from a probability distribution<sup>1</sup> e.g., if we pick randomly a cell in our population, how much are we surprised to pick one cell with a given GFP level? In our case, we can measure the entropy based on fluorescence distribution acquired with automated FC based on the following equation:

$$\text{Eqn 1 : } H = - \sum_{i=1}^m p(x_i) \cdot \log_2 p(x_i)$$

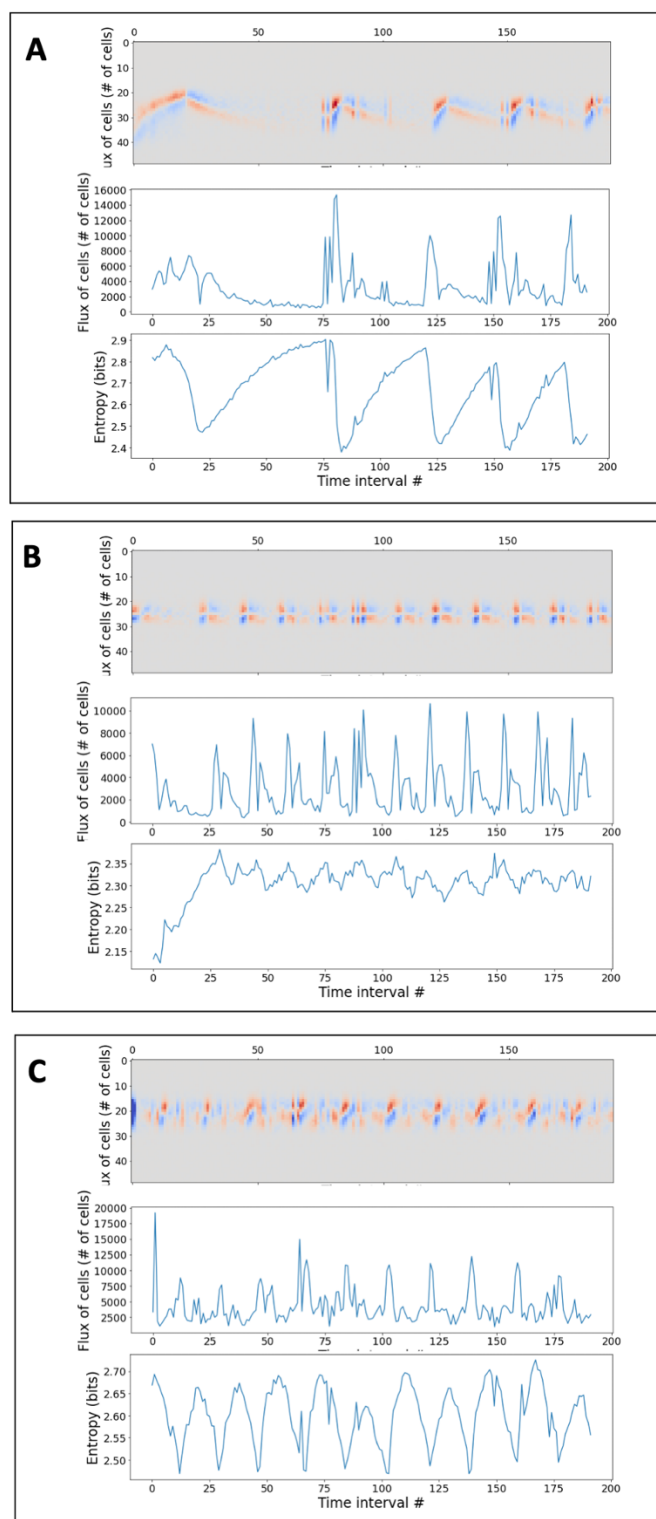
With  $m$  being the number of states observed (i.e., GFP classes in our cases) and  $p$  being the probability to observe this state. The probability to observe different classes of fluorescence can be easily determined based on automated FC. The computation of  $H$  has been exemplified based on fictive population of cells clustered in three fractions according to the level of GFP exhibited by cells (**Figure S1**). As an example, the computation for the first population distribution (**Figure S1A**) is performed as:

$$H = - [(0.1 \cdot \log_2 0.1) + (0.6 \cdot \log_2 0.6) + (0.3 \cdot \log_2 0.3)] = 1.29 \text{ bits}$$



**Figure S1:** computation of  $H$  for cell population clustered in three bins (computation according to equation 1). Each bin corresponds to a subpopulation of cells with a given fluorescence range i.e., low, medium or high.

Based on this first example, the entropy of the population can be either increased (**Figure S1B**), the maximum entropy value being reached when cells are equally distributed into the 3 different clusters. On the opposite,  $H$  can be decreased and set to zero when all the cells exhibit the same fluorescence range (**Figure S1C**). This approach has been applied to automated FC data for computing the evolution of  $H$  for different types of cell population (**Figure S2** for three out of six of the cell systems investigated). In this case, we applied 50 bins for the computation of  $H$  (see Supplementary note 2 for more explanations).



**Figure S2:** computation of the entropy and flux of cells for the **A**  $P_{lacZ}$ :GFP (*E. coli*), **B**  $P_{bolA}$ :GFP (*E. coli*) and **C**  $P_{glc3}$ :GFP (*S. cerevisiae*) systems.

**Supplementary note 2** Experimental determination of the response function for the  $P_{araB}$ :gfp system in *E. coli* and the  $P_{glc3}$ :gfp system in *S. cerevisiae* and computation of the mutual information (MI)



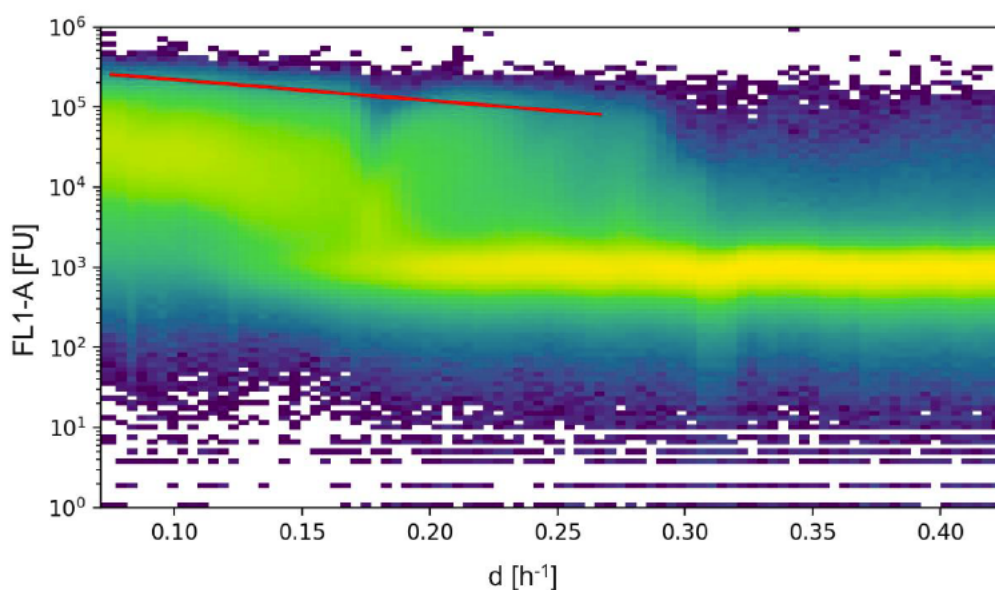
0.1, 0.15, 0.20, 0.25, 0.30, 0.50, 1.00 and 2.00 g/L. The redline represents the fluorescence threshold (i.e., 1000 F.U.) used for computing the GFP positive fraction of cells.

**Table S1**

Primer name	Sequence
Fw_sgRNA_20N_Ara	agctagctcagtcctaggtataataactagtCGCCCGCAGGATATTCGTCGgttttagagcta gaaatagcaagttaaaa
Fw_fragment1	gttaaacgagtatcccggcagca
Rv_fragment1	cgtttactccatccaaaaaacgg
Fw_fragment2	tggagtgaaacgtgactgtataaaaccacagccea
Rv_fragment2	catcggcctcgtagacggtaac

**b. Characterization of the response function for the *S. cerevisiae*  $P_{glc3}$ :GFP system**

The response function of the  $P_{glc3}$ :GFP in *S. cerevisiae* was determined by growing culture at different dilution rates in chemostat (**Figure S4**). For this purpose, the dilution rate of a chemostat was progressively increased in order to release the stress response of the population. This procedure is known as accelerostat (A-stat) and involves the progressive increase of the dilution rate. In our case, the pump flow rate was modified according to step change every 2 hours, resulting in a progressive increase of the dilution rate of  $0.002 \text{ h}^{-1}$  every hour. This incremental range was chosen in order to ensure pseudo steady-state for each increment. The whole process was followed by automated FC for mapping the GFP distribution of the cell population (**Figure S4**).





**Figure S4:** A-stat experiment monitored based on automated FC. The GFP level distribution (FL1-A channel) was determined for each dilution rate (D). The red line highlights the progressive release of the stress response at the population level based on the deactivation of the  $P_{glc3}$ :GFP reporter.

c. Computation of mutual information (MI) based on the response function of a cell population

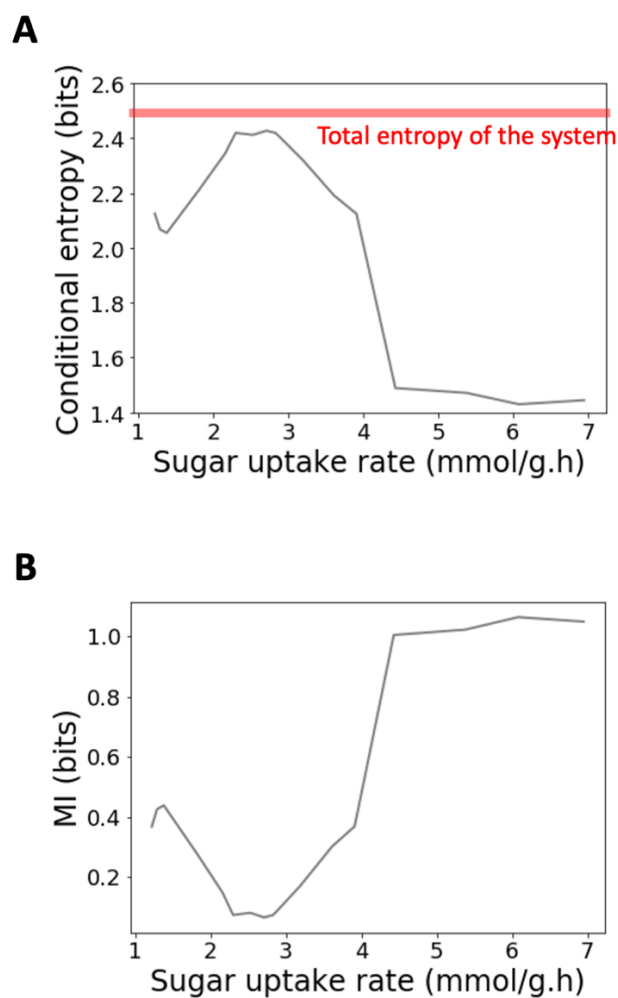
Knowing the response function, and the corresponding conditional GFP distribution, it is possible to compute the MI of a specific cellular system. This computation will be exemplified for the  $P_{glc3}$ :GFP reporter in *S. cerevisiae*. The environmental input for this system is the glucose uptake rate determined based on the value of the dilution rate, as well as based on glucose and biomass measurement. The conditional fluorescence distributions were then acquired for different substrate uptake rate and H was computed accordingly (**Figure S5A**). If all the fluorescence distributions are summed up, the corresponding entropy value is the total entropy of the system. MI is then computed according to (**Figure S5B**):

$$\text{Eqn 2 : } MI(y, x) = H(x) - H(x, y)$$

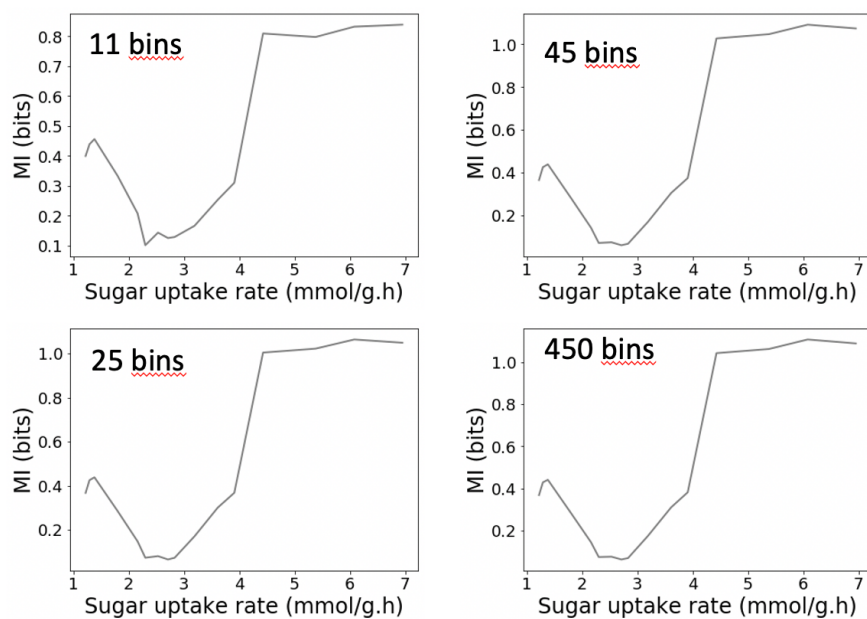
With  $H(x)$  being the total entropy for output  $x$  and  $H(x, y)$  being the conditional entropy computed from the conditional distribution of the output  $x$  ( $x$ , being GFP distribution).

When doing so, it is important to adjust the number of bins used for computing the entropy. In our case, this number was set to 45 bins and leads to a precise computation of MI without increasing the computing power (**Figure S6**).





**Figure S5:** A Evolution of the conditional entropy for the  $P_{glc3}$ :GFP in *S. cerevisiae* exposed to different uptake rates. B MI can be deduced by subtracting the value of the total entropy of the system ( 2.49 bits in this case) by the corresponding conditional entropy.



**Figure S6:** impact of the binning procedure (number of bins considered) on the estimation of MI for the  $P_{glc3}:GFP$  in *S. cerevisiae* (data extracted from A-stat experiments).

### Supplementary note 3 Flow cytometry Stochastic Kinetic Simulator (FlowStockS)

FlowStockS is a computational tool designed to simulate the phenotype distributions of a microbial population under different environmental regimes (Chemostat and Segregostat). Following the model framework classification proposed by Hartmann *et al.*<sup>5</sup>, FlowStockS can be considered as: i) biologically segmented as it considers single cells; ii) abiotically unsegmented as it assumes a homogeneous environment; iii) an unstructured cell model as it does not take intracellular kinetics or metabolic fluxes into consideration. FlowStockS comprises two modules: the growth module and the switch module. The system is resolved using a Markov chain with discrete time, where the growth module is described by a set of ordinary differential equations (ODE). These ODEs detail how single cells grow (**Eqn 3 and Eqn 4**) and consume (**Eqn 5**) their substrate (S), in accordance with the Monod-type equations that include a non-competitive growth inhibition term.

$$\text{Eqn 3: } \mu = \mu_{\max} \frac{[S]}{[S] + K_s} \frac{[I]}{[I] + K_I}$$

$$\text{Eqn 4: } \frac{dX}{dT} = \mu X - DX$$

$$\text{Eqn 5: } \frac{dS}{dT} = D[S_{\text{Feed}}] - Y\mu X - D[S]$$

Where:

$\mu_{\max}$  = Maximal growth rate

$\mu$  = Growth rate

$X$  = Cell biomass

$K_I$  = Growth inhibition constant

$K_s$  = Affinity for the substrate

$Y$  = Substrate to biomass yield

$D$  = Dilution rate

$[S_{\text{Feed}}]$  = Substrate concentration in the feed

$[I]$  = Inhibitor concentration

$[S]$  = Substrate concentration

In the growth module, single cells are simulated to grow until they double in size, at which point they divide into two daughter cells. Additionally, to simulate continuous cultivations, cells are randomly flushed out of the system based on a probability ( $P_{\text{out}}$ ) set by the dilution rate and the time step ( $T_{\text{step}}$ ) used in the simulation (**Eqn 6**).

$$\text{Eqn 6: } P_{\text{out}} = D T_{\text{step}}$$

The growth parameters ( $\mu_{max}$ ,  $KI$ ,  $Y$ ,  $Ks$ ) are given by the switch module, and define the phenotype of each cell. To initiate the switching process, a cell must first cross a time threshold by accumulating  $T_{step}$ , and this commitment process is governed by a switching probability ( $P$ ). This probability is determined by a response function, taking the inducer concentration ( $i$ ) as input. The  $P$  function is a classical sigmoid function characterized by a steepness ( $n$ ) and a 50% switching probability at concentration ( $K$ ).

$$\text{Eqn 7: } P(i) = \frac{[i]^n}{[i]^n + K}$$

The accumulation of  $T_{step}$  is analogous to the build-up of enzymes that are necessary for the expression of a different phenotype, commonly known in many processes e.g., substrate consumption switching and the diauxic shift time<sup>678</sup>. Once the time threshold ( $\tau$ ) is reached, the cell switches phenotype. In our experimental set-up, we use GFP-based reporters to track the phenotype switch. Thus, once the time threshold is reached a GFP is produced as a burst following a production rate set by a gamma distribution ( $I_{(q_{gfp},2)}$ )<sup>9</sup>. Similarly to Taniguchi *et al.*<sup>9</sup>, we assumed that active degradation of GFP or the inhibitor is neglectable and accordingly dilution is determined by cell division.

With this FlowStoCKS toolbox, we conducted simulations using a constant time step ( $T_{step}$ ) of 1 minute, which is two orders of magnitude faster than the fastest process considered ( $T_{step} \ll \mu_{max}$ ). In these simulations, we modeled a population of 10,000 cells, each with an initial mass equal to 1/1000th of the initial biomass. Our simulations were based on the growth characteristics of *Saccharomyces cerevisiae* (**Table S1**). The switch to the stressed phenotype is triggered by glucose limitation resulting in growth inhibition. To simplify our analysis, we assumed that the resulting GFP production is linearly correlated with the inhibitor concentration, and thus its concentration was used as the inhibitor

Finally, by varying inhibition value  $KI$  (lower value means higher growth inhibition), we were able to compute the phenotype distribution both in Chemostat and Segregostat for different fitness cost associated to switching. All FlowStoCKS codes are available on the GitLab page (<https://gitlab.uliege.be/mipi/published-software/mbms-toolbox/-/tree/main/FlowStoCKS>).

**Table S2:** Parameters used for running FlowStoCKS simulations

Parameter	Description	Value	Unit	Source
$\mu_{max\_glucose}$	Maximal growth rate on glucose	0.54	$h^{-1}$	Jain <sup>10</sup>
$Ks\_glucose$	Affinity constant for glucose	0.034	g/l	Jain <sup>10</sup>
$Y\_glucose$	Substrate to biomass yield for glucose yeast	0.5	g/g	Bionumbers ID 109651 <sup>11</sup>
$n$	Hill coefficient	-2		This study <sup>a</sup>
$k$	50 % probability switching concentration	0.05	g/l	This study <sup>a</sup>
$Gfp\_prod$	Mean value of GFP production burst given a gamma distribution of	1e5	fu	This study <sup>b</sup>

scale 2 and growth inhibition

KI	Growth inhibition	From 0 to 1e99	fu	by $\frac{[I]}{[I]+KI}$
delay ( $\tau$ )	Time delay between pulse and GFP production	0.4	h	This study <sup>c</sup>

<sup>a</sup>Approximated from the proportion of induced cells for different glucose concentrations observed in microfluidic cultivation

<sup>b</sup>Set to evaluate the impact of inhibition strength, Gfp\_prod and KI individual values are arbitrary, only the ratio matters as the inhibition strength is set

<sup>c</sup>A delay of 0.4 hours (24 minutes) between sensing and switching was selected based on Segregostat observations, where two consecutive automated FC measurements (24 minutes) separate a pulse and the first increase in fluorescence.

**Supplementary note 4** Operating conditions used for the chemostat and Segregostat experiments

**Table S3**

Strain	Process conditions	Control conditions
<i>E. coli</i> W3110 <i>P<sub>araB</sub>::GFPmut2</i>	D = 0.45 h <sup>-1</sup> pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min <sup>-1</sup> [Glucose feed] = 5 g/L [Arabinose feed] <sub>chemostat</sub> = 1.5 g/L [Arabinose feed] <sub>Segregostat</sub> = 0	Chemostat : none Segregostat : Regulation threshold = 50 % cells bellow 1000 F.U. Regulation = Pulse 0.15 g arabinose
<i>E. coli</i> W3110 <i>P<sub>lacZ</sub>::GFPmut2</i>	D = 0.45 h <sup>-1</sup> pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min <sup>-1</sup> [Glucose feed] = 5 g/L [Lactose feed] <sub>chemostat</sub> = 1.5 g/L [Lactose feed] <sub>Segregostat</sub> = 0	Chemostat : none Segregostat : Regulation threshold = 50 % cells bellow 1000 F.U. Regulation = Pulse 0.15 g lactose
<i>E. coli</i> BL21 pET28:GFP	D = 0.45 h <sup>-1</sup> pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min <sup>-1</sup> [Glucose feed] = 5 g/L [Lactose feed] <sub>chemostat</sub> = 1 g/L [Lactose feed] <sub>Segregostat</sub> = 0	Chemostat : none Segregostat : Regulation threshold = 50 % cells bellow 1000 F.U. Regulation = Pulse 0.5 g lactose
<i>E. coli</i> W3110 <i>P<sub>bolA</sub>::GFPmut2</i>	D = 0.45 h <sup>-1</sup> pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min <sup>-1</sup> [Glucose feed] = 5 g/L	Chemostat : none Segregostat : Regulation threshold = 50 % cells above 2000 F.U. Regulation = Pulse 0.2 g glucose
<i>B. subtilis</i> 168 <i>P<sub>spoII</sub>E::GFPmut2</i>	D = 0.1 h <sup>-1</sup> pH = 7 T = 37 °c Aeration = 1 L/min Agitation = 1000 min <sup>-1</sup> [Glucose feed] = 5 g/L	Chemostat : none Segregostat : Regulation threshold = 20% cells above 1000 F.U. Regulation = Pulse 0.2 g glucose
<i>S. cerevisiae</i> CEN-PK117D <i>P<sub>glc3</sub>::eGFP</i>	D = 0.1 h <sup>-1</sup> pH = 5 T = 30 °c Aeration = 1 L/min Agitation = 1000 min <sup>-1</sup> [Glucose feed] = 5 g/L	Chemostat : none Segregostat : Regulation threshold = 50% cells above 5000 F.U. Regulation = Pulse 0.2 g glucose

## References

1. Harte, J. & Newman, E. A. Maximum information entropy: a foundation for ecological theory. *Trends Ecol Evol* **29**, 384–389 (2014).
2. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).
3. Hansen, A. S. & O’Shea, E. K. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *Elife* **4**, (2015).
4. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl Environ Microbiol* **81**, 2506–2514 (2015).
5. Hartmann, F. S. F., Udugama, I. A., Seibold, G. M., Sugiyama, H. & Gernaey, K. V. Digital models in biotechnology: Towards multi-scale integration and implementation. *Biotechnol Adv* **60**, 108015 (2022).
6. Chu, D. & Barnes, D. J. The lag-phase during diauxic growth is a trade-off between fast adaptation and high growth rate. *Sci Rep* **6**, 25191 (2016).
7. Martinez, J. A. *et al.* Controlling microbial co-culture based on substrate pulsing can lead to stability through differential fitness advantages. *PLoS Comput Biol* **18**, e1010674 (2022).
8. Jones K. D., K. D. S. Cybernetic model of the growth dynamics of Saccharomyces cerevisiae in batch and continuous cultures. *Journal of biotechnology* **71**, 105–131 (1999).
9. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
10. Jain, V. K. Relationship between energy metabolism and growth. I. Glucose dependence of the exponential growth rate of Saccharomyces cerevisiae. *Arch Mikrobiol* **72**, 252–259 (1970).
11. Milo, R. & Phillips, R. *Cell biology by the numbers*. (Garland science, 2016).

