Growing from a few cells: combined effects of initial stochasticity and cell-to-cell variability.

published in: Journal of the Royal Society Interface

SUPPLEMENTARY INFORMATION

A. Barizien[1,2],M.S. Suryateja Jammalamadaka[1,3], G. Amselem[1], Charles. N. Baroud[*1,2]

[1] *LadHyX and Department of Mechanics,*

Ecole Polytechnique, CNRS, 91128, Palaiseau, France

[2] *Physical microfluidics and Bioengineering,*

Department Genomes and Genetics, Institut Pasteur, 75015, Paris, France

[3] *Center for Research and Interdisciplinarity, 75014, Paris, France*

I. NOTATIONS

Throughout this article, the notation $\sim$ is used in its mathematical sense, "is equivalent to", meaning that for two functions f and g :

$$f(t) \sim g(t) \Leftrightarrow \frac{f(t)}{g(t)} \xrightarrow{t \to \infty} 1. \tag{1}$$

The "normalized moments" of a distribution are defined as follows, if $k \geq 2$:

$$\bar{\mu}_k = \left| \mathbb{E}\left[(X - \mu)^k \right] \right|^{1/k}. \tag{2}$$

For $k = 1$, we define the first normalized moment just by the mean of the distribution:

$$\bar{\mu}_1 = \mathbb{E}(X). \tag{3}$$

II. EQUATIONS OF THE BELLMAN-HARRIS MODEL

1. The mean number of cells M(t) is equivalent, for $t \to \infty$, to an exponential function:

$$M_N(t) \sim n_1 e^{\alpha t}, \tag{4}$$

where the value of the growth rate α is given by the unique solution of the equation:

$$\int_0^\infty e^{-\alpha t} \rho(t)\, \mathrm{d}t = \frac{1}{2}, \tag{5}$$

where ρ is the density function related to the division time τ_d. We **can re-write this equation as follows** :

$$\mathbb{E}(X) = \frac{1}{2} \quad \text{with} \quad X = e^{-\alpha \tau_d}. \tag{6}$$

n_1 is a constant whose value is given by

$$n_1 = \frac{1}{4\alpha \int_0^\infty t e^{-\alpha t} g(t)\, \mathrm{d}t}. \tag{7}$$

2. There is a similar equation for the variance of the number of cells Var(t) , for $t \to \infty$:

$$Var(t) \sim \frac{4\mathbb{E}(X^2) - 1}{1 - 2\mathbb{E}(X^2)} M_N(t)^2. \tag{8}$$

The standard deviation is therefore:

$$SD(t) \sim n_2 e^{\alpha t} \quad \text{with} \quad n_2 = n_1 \sqrt{\frac{4\mathbb{E}(X^2) - 1}{1 - 2\mathbb{E}(X^2)}}, \tag{9}$$

and the coefficient of variation of the number of cells in exponential phase:

$$\mathrm{CV}_N \sim \sqrt{\frac{4\mathbb{E}(X^2) - 1}{1 - 2\mathbb{E}(X^2)}}. \tag{10}$$

III. GAUSSIAN DIVISION TIME

In this part, we are going to study the case of a Gaussian division time for the individual bacteria. This means that all bacteria "choose" their division time following a Gaussian law of mean τ_0 and standard deviation σ. We saw that the random variable appearing in the model is $X = e^{-\alpha\tau_0}$. In this case, as τ_d follows a normal law, X follows a rescaled log-normal law and we can easily find the moments of X:

$$\mathbb{E}(X) = e^{-\alpha\tau_0 + \frac{\alpha^2\sigma^2}{2}}, \tag{11}$$

$$\mathbb{E}(X^2) = e^{-2\alpha\tau_0 + 2\alpha^2\sigma^2}. \tag{12}$$

If we plug equation(**??**) into equation (**??**), we get the theoretical value of α:

$$\alpha = \frac{\tau_d - \sqrt{\tau_0^2 - 2\ln(2)\sigma^2}}{\sigma^2}. \tag{13}$$

If the coefficient of variation of the division time $cv = \sigma/\tau_d$ is small ($cv \ll 1$), we get:

$$\alpha \sim \frac{\ln(2)}{\tau_0}\left(1 + \frac{\ln(2)}{2}cv^2\right), \tag{14}$$

$$\mathbb{E}(X^2) \sim 1 + \frac{\ln(2)^2}{4}cv^2. \tag{15}$$

We have for the coefficient of variation of the number of cells in exponential phase:

$$\mathrm{CV}_N \sim \sqrt{2}\ln(2)cv. \tag{16}$$

Which is a linear relation of slope $\sqrt{2}\ln(2) \approx 1$, as observed numerically by [**?**].

We can also easily get the expression of the coefficients of the mean and the standard deviation by plugging a gaussian law into the expression presented above, and we obtain:

$$n_1 = \frac{1}{4\alpha I} \quad \text{where} \quad I = \frac{e^{-\frac{\tau_0^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} + \frac{1}{2}\sqrt{\tau_0^2 - 2\ln(2)\sigma^2}, \tag{17}$$

$$n_2 = n_1\sqrt{\frac{4\mathbb{E}(X^2) - 1}{1 - 2\mathbb{E}(X^2)}} \quad \text{with} \quad \mathbb{E}(X^2) = 0.25\exp\left[\left(\frac{2\ln(2)\mathrm{cv}_\mu}{1 + \sqrt{1 - 2\ln(2)\mathrm{cv}_\mu^2}}\right)^2\right]. \tag{18}$$

IV. NUMBER OF CELLS PER DROPLET LARGER THAN 1

If there are more than one bacteria in a single droplet, the asymptotic behavior of the moments do not change, simply because of the independence of the bacteria in the Bellman-Harris model. But the values of the prefactors do change. Consider a droplet containing initially k bacteria, with $k > 1$. We call $N_i(t)$ the size of the offspring of bacteria i at time t. The total number of bacteria $N_k(t)$ in the droplet at time t is then given by:

$$N_k(t) = \sum_{i=1}^{k} N_i(t). \tag{19}$$

All the $N_k(t)$ are independent and identically distributed, thus we have :

$$\mathbb{E}\left(N_k(t)\right) = \sum_{i=1}^{k} \mathbb{E}\left(N_i(t)\right) = k n_1 e^{\alpha t}, \tag{20}$$

$$Var\left(N_k(t)\right) = \sum_{i=1}^{k} Var\left(N_i(t)\right) = k n_2^2 e^{2\alpha t}. \tag{21}$$

Thus, for the asymptotic coefficient of variation, we get:

$$CV_k(\infty) = \frac{1}{\sqrt{k}} \left(\frac{n_2}{n_2}\right)_{BH}. \tag{22}$$

V. GENERATION-DEPENDENT DIVISION TIME

We can carry out the same analysis for the first three generations taken with different laws for the division times. The principle is the same, except that this time we need to consider the 8 children of the original bacteria in order to apply them the classical Bellman-Harris model, and we finally get :

$$\mathbb{E}(N(t)) \sim 8 n_1 \mathbb{E}(X_1)\mathbb{E}(X_2)\mathbb{E}(X_3) e^{\alpha t}, \tag{23}$$

$$\mathbb{E}(N(t)^2) \sim 8 n_1^2 \mathbb{E}(X_1^2) \left(\mathbb{E}(X_2^2)\left(\frac{\mathbb{E}\left(X_3^2\right)}{1 - 2\mathbb{E}(X^2)} + 2 * \mathbb{E}(X_3)^2\right) + 4\mathbb{E}(X_2)^2 \mathbb{E}(X_3)^2\right) e^{2\alpha t}. \tag{24}$$

VI. COEFFICIENT OF VARIATIONS FOUND IN THE LITERATURE FOR SINGLE CELL DIVISION

Different authors have studied the spreading of division times for single cells for the n first generations, and although their strains and growth conditions differ from ours, we can still compare

our own experimental results with theirs to see if we obtain the same orders of magnitude.

Generation	Experiment	Kousoumanis [**?**]	Pin [**?**] 32deg	Pin [**?**] 25deg
1	34	43	47	31
2	42	49	33	31
3	39	48	38	36
4	31			31

TABLE I: Coefficient of variation(%) for the division times of single cells, for different generations, found in literature.

Taheri et al. [**?**] measured coefficient of variations ranging from 14% to 22% in full exponential phase, for *E. coli*.

VII. RELATION BETWEEN FLUORESCENCE INTENSITY AND NUMBER OF CELLS

We consider that the fluorescence intensity of the droplet is proportional to the number of cells in the droplet, because it is the sum of the fluorescent signals of all cells contained in the droplet, that we consider to be the same for now (the question of its heterogeneity will be discussed later).

$$Fluo_{drop}(t) = a_{fluo} N(t) + Fluo_{noise}. \tag{25}$$

The noise being independent of the number of cells in the droplet, we find the following relations:

$$\mathbb{E}_N(t) = \frac{\mathbb{E}_{Fluo_{drop}}(t) - \mathbb{E}_{Fluo_{noise}}}{a_{fluo}}, \tag{26}$$

$$Var_N(t) = \frac{Var_{Fluo_{drop}}(t) - Var_{Fluo_{noise}}}{a_{fluo}^2}. \tag{27}$$

The coefficient of variation of the number of cells can therefore be estimated by measuring the fluorescence **signal** of the droplets and the **background** noise. The **background** noise is estimated by measuring the fluorescence signal of empty droplets (i.e. droplets that do not contain any bacteria). Note that the coefficient of variation, which is the ratio of the standard deviation over the mean of the number of cells, does not depend on the coefficient of proportionality a_{fluo}. For higher moments of the distribution of cells, estimations can also be carried out, with μ_n the

n^{th} central moment of a distribution. For the third central moment, we simply have, thanks to the independence of the division times of bacteria, and just like for the first and second moments:

$$\mu_3(N) = \frac{\mu_3(Fluo) - \mu_3(Noise)}{a_{fluo}^3}. \tag{28}$$

Then as a general rule, if X and Y are two independent variables, we have, for $n \geq 4$:

$$\mu_n(X+Y) = \mu_n(X) + \mu_n(Y) + \sum_{k=2}^{n-2} \binom{n}{k} \mathbb{E}\left((X - \mathbb{E}(X))^k\right) \mathbb{E}\left((X - \mathbb{E}(X))^{n-k}\right). \tag{29}$$

In our particular case, we have:

$$\mu_4(N) = \frac{\mu_4(Fluo) - \mu_4(Noise) - 6\left(Var(Fluo) - Var(Noise)\right) Var(Noise)}{a_{fluo}^4}. \tag{30}$$

Heterogeneity of the fluorescence signal of the cells

Another source of stochasticity in the experiment can come from the heterogeneity of the fluorescence signal, which is not homogeneous among the cells but varies from one bacteria to another, even if the cells are isogenic[**?**]. In this case, the relation between the fluorescence intensity and the number of cells can be written as:

$$Fluo(t) = \sum_{i=1}^{N(t)} a_f^i(t) \quad \text{where} \quad Fluo(t) = Fluo_{drop}(t) - Fluo_{noise}, \tag{31}$$

where $a_f^i(t)$ represents the fluorescence signal of cell i at time t. To simplify the computations, we will consider that the fluorescence signals of the cells are independent. These is of course a simplifying hypothesis, as the fluorescence signal of two sister cells is probably correlated for instance, but as we average on a lot of cells, this effect is negligible compared to the heterogeneity in itself. We will note σ_f^2 the variance of a_f^i and a_f their mean. The coefficient of variation of the fluorescence signal of the cells is then $cv_f = \sigma_f / a_f$.

Then we have, because we take the sum of independent and identically distributed variables:

$$\mathbb{E}\left(Fluo(t)\right) = a_f \mathbb{E}(N(t)), \tag{32}$$

$$\text{Var}\left(Fluo(t)\right) = \mathbb{E}(N(t))\sigma_f^2 + a_f^2 \text{Var}(N(t)), \tag{33}$$

which yields for the coefficient of variation:

$$CV_{fluo}^2 = CV_N^2 + \frac{cv_f^2}{\mathbb{E}(N(t))}. \tag{34}$$

We can see here that the heterogeneity of the fluorescent signal of the cells adds a corrective term to the coefficient of variation of the fluorescence signal. However, this term evolves as $1/\mathbb{E}(N(t))$. Since $\mathbb{E}(N(t))$ increases exponentially with rate α, the corrective term vanishes rapidly, and we can affirm that in exponential phase, the measured coefficient of variation is not affected by the heterogeneity of the fluorescence signal of the cells.

As discussed in details in [**?**], the number of bacteria per droplet initially follows a Poisson distribution, which parameter can be estimated by counting the number of empty droplets:

$$\lambda = -\ln\left(\frac{\text{Number of empty droplets}}{\text{Total number of droplets}}\right). \tag{35}$$

VIII. SUPPLEMENTARY FIGURES

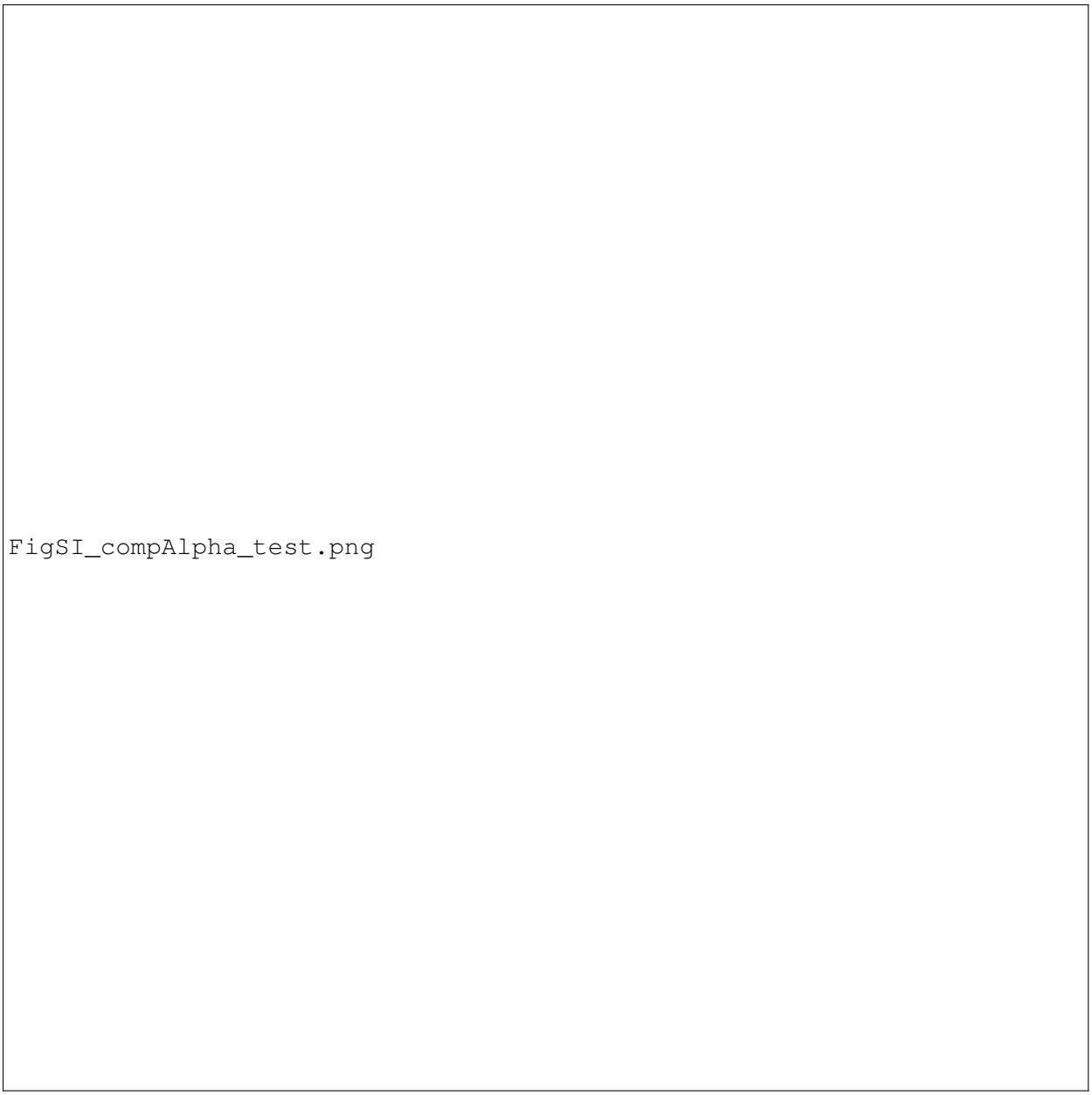

FIG. 1: Principle of the test to compare the three different models at the population scale. First a grid of simulations is built for experimentally possible values of σ_l and σ_t, for all three models. Then each point on the grid is compared to all other points of the grid: the final distributions of number of cells, rescaled by their means, are compared through a two sample Kolmogorov-Smirnov test. As an example, the distribution of the number of cells obtained by an adder model at the position indicated by the red dot is compared to the distributions obtained by the sizer and timer models at the position indicated by the green dot. In this particular example, the results of the adder model at the red position are indistinguishable from that of the timer model at the green position, but distinguishable from that of the sizer ($p < 0.05$). The same method is applied to all points and all models of the grid of simulations. We find that, for each point (σ_l, σ_t) and each model, there is always at least one set of parameters (σ_l, σ_t) of one of the other models leading to similar results with $p < 0.05$.

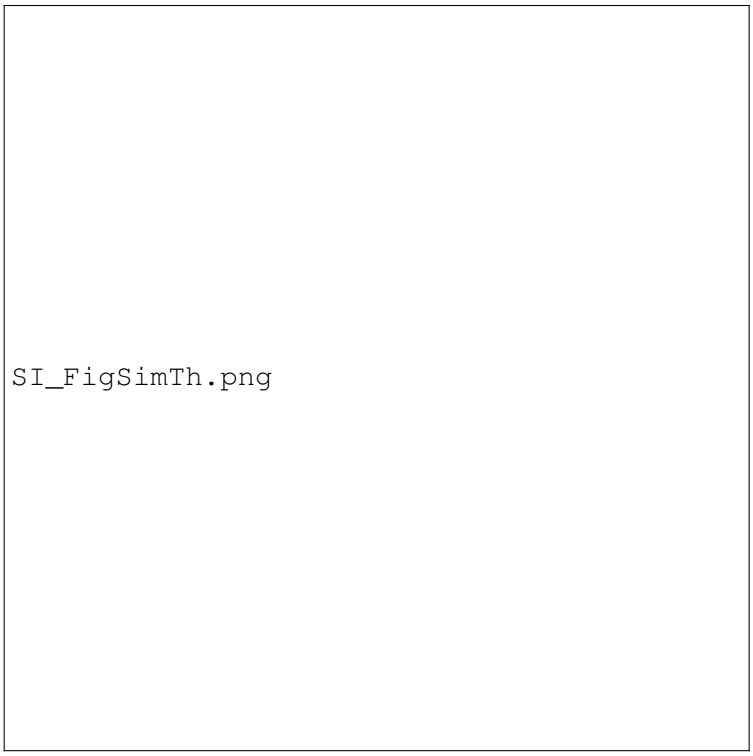

FIG. 2: Comparison between the theoretical predictions (straight lines) of the Bellman-Harris model and Monte-Carlo simulations (stars). The law for the division time of individual bacteria is a Gaussian with mean $\tau_0 = 23$ min and standard deviation $\sigma = 0.25\tau_0$. In (A) and (B) the initial number of cells follows a Poisson distribution of parameter $\lambda = 0.5$. Two thousand independent simulations were performed. (A) Mean (red) and standard deviation (blue) of the number of cells as a function of time, and number of cells (green) as a function of time for each simulation. (B) Evolution of the coefficient of variation of the number of cells. (C) $n_1(\lambda)$ and $n_2(\lambda)$ as a function of the Poisson parameter λ (2000 simulations for each λ). (D) Effect of an evolving distribution of division times. The first bacterial generation divides with a mean time $\tau_1 = 2\tau_0$, and a standard deviation σ_1. Subsequent division times follow a normal law of mean τ_0 and standard deviation σ_0. n_1 and n_2 depend on σ_1/τ_1. The initial number of cells follows a Poisson distribution with a parameter $\lambda = 0.5$. Result of 2000 independent simulations for each σ_1.

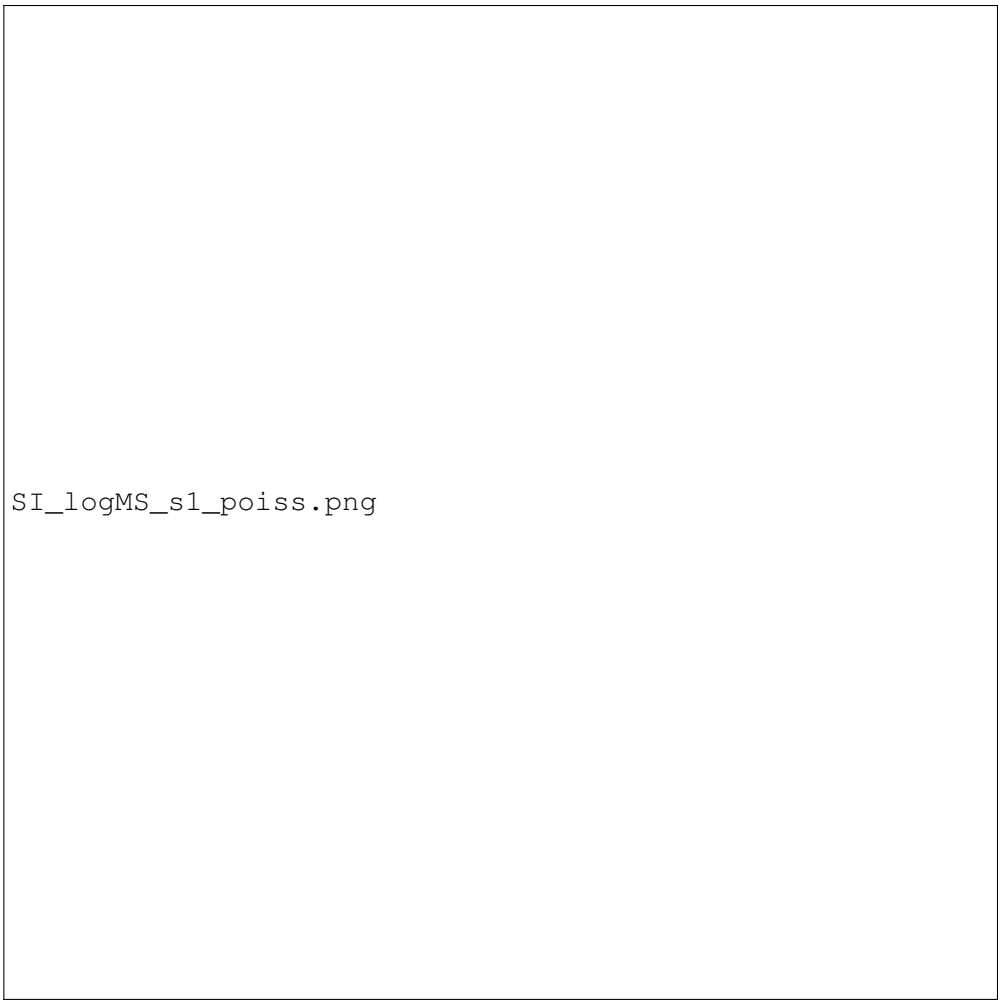

FIG. 3: Mean (red) and standard deviation (blue), for Monte-Carlo simulations (stars) and theory (straight lines). (A) Mean, standard deviation and growth curves for the simulations summarized in Supp. Fig. **??**(C). The law for the division time of individual bacteria is a Gaussian with mean $\tau_0 = 23$ min and standard deviation $\sigma = 0.25\tau_0$, the parameter of the Poisson distribution for the initial number of cells is varied. (B) Mean, standard deviation and growth curves for the simulations summarized in Supp. Fig. **??**(D). The first bacterial generation divides with a mean time $\tau_1 = 2\tau_0$, and a standard deviation σ_1. Subsequent division times follow a normal law of mean τ_0 and standard deviation σ_0. 2000 independent simulations were used for each value of λ (A) or σ_1 (B).

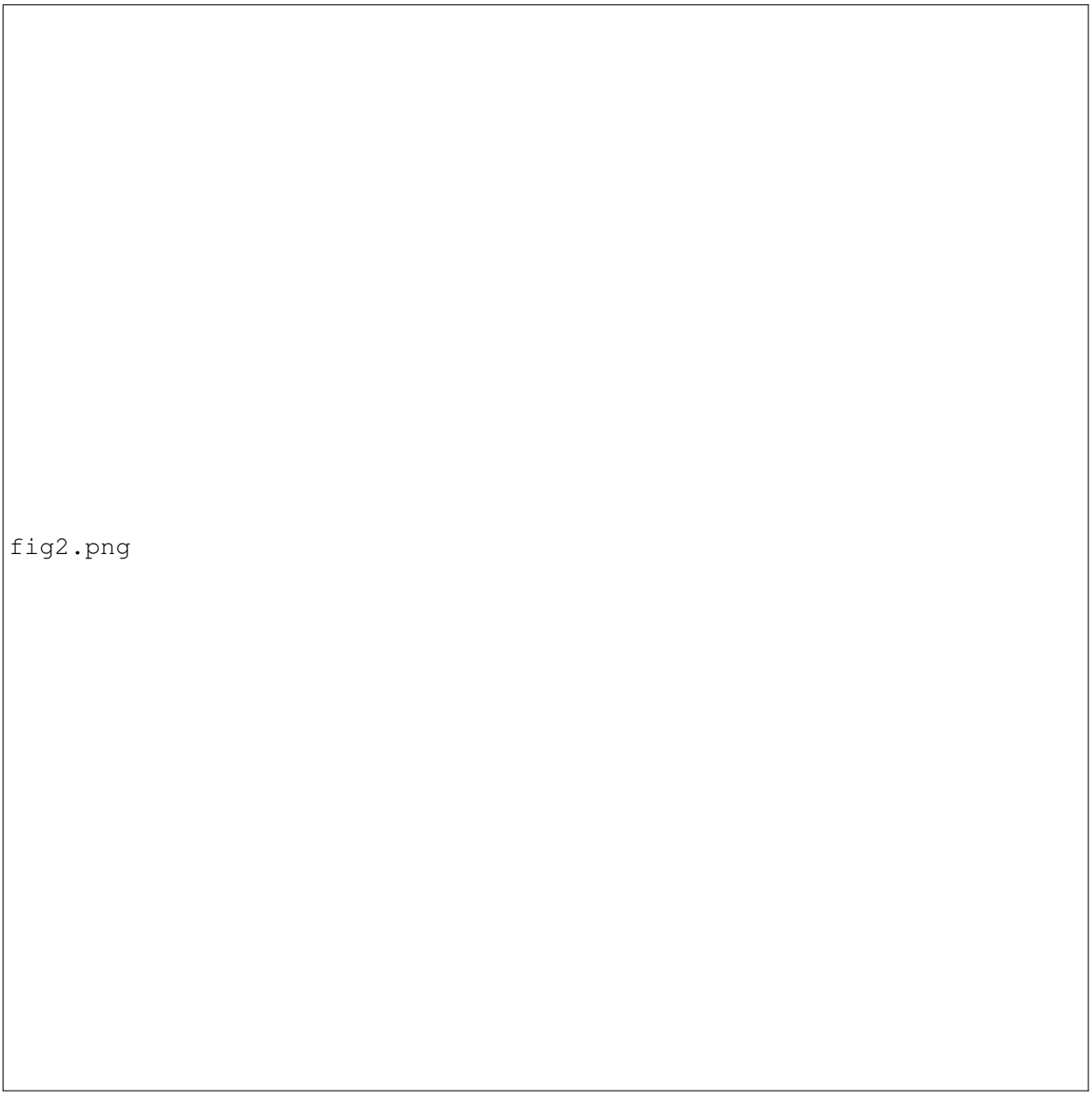

FIG. 4: (A) Time-lapse image of *E. coli* cells dividing under a 90X objective. The images correspond to t=0, 21, 42, 63, 84, and 105 min respectively. (B) Density of the division times obtained from the time-lapse images, for *E. coli*, and for the first four generations. (C) Density of the division times obtained from the time-lapse images, for *B. subtilis*, and for the first four generations. (D) Fitted values of the mean division times and their standard deviations, for both strains.

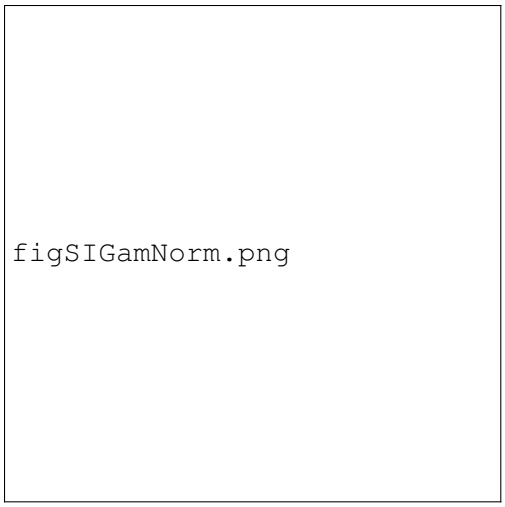

FIG. 5: Comparison of fits for the experimental distribution of division times shown in Supp. Fig. **??**, for the first 4 generations. Blue: fit to a Gaussian distribution. Red: fit to a Gamma distribution.

FIG. 6: Comparison between experiments and simulations for the evolution of the 3rd (A), 4th (B) and 5th (C) central moments of the distribution of the number of cells. Data shown are for *E. coli*.

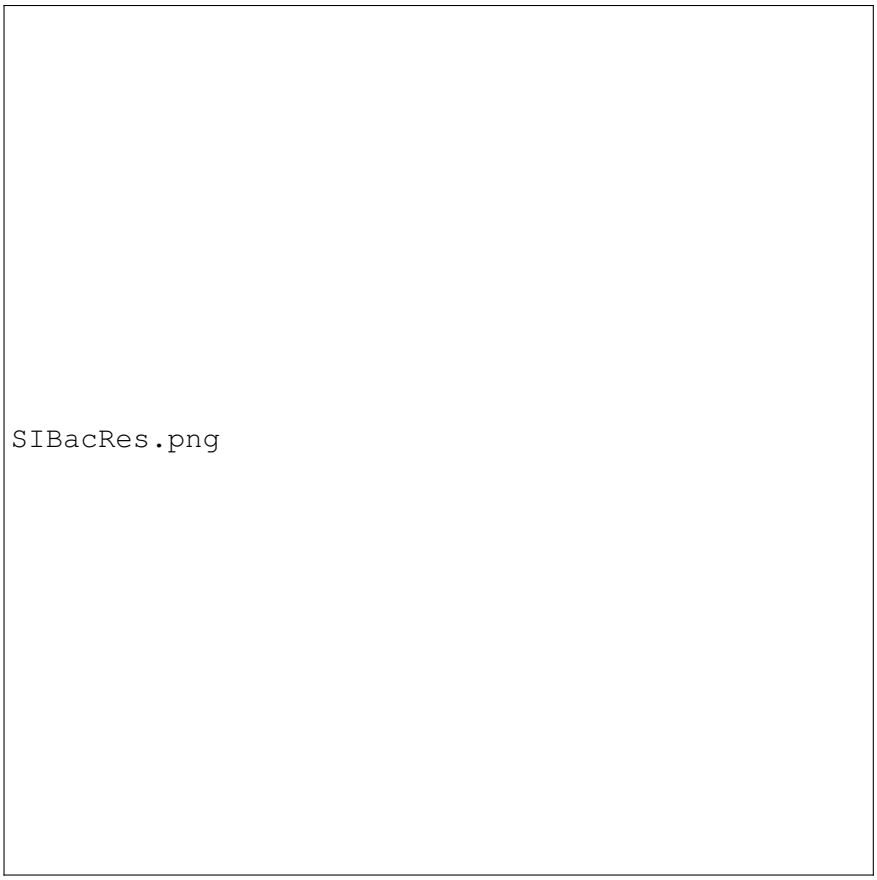

FIG. 7: Comparison between the model (yellow), numerical (orange) and experimental (blue) results, for *B. subtilis*. (A) Mean number of cells as a function of time $M_N(t)$, (B) Standard deviation of the number of cells as a function of time $SD_N(t)$, (C) Coefficient of variation of the number of cells as a function of time CV_N (D) Shape of the distribution of $N(t)/\exp(\alpha t)$, numerical and experimental, kernel fit.

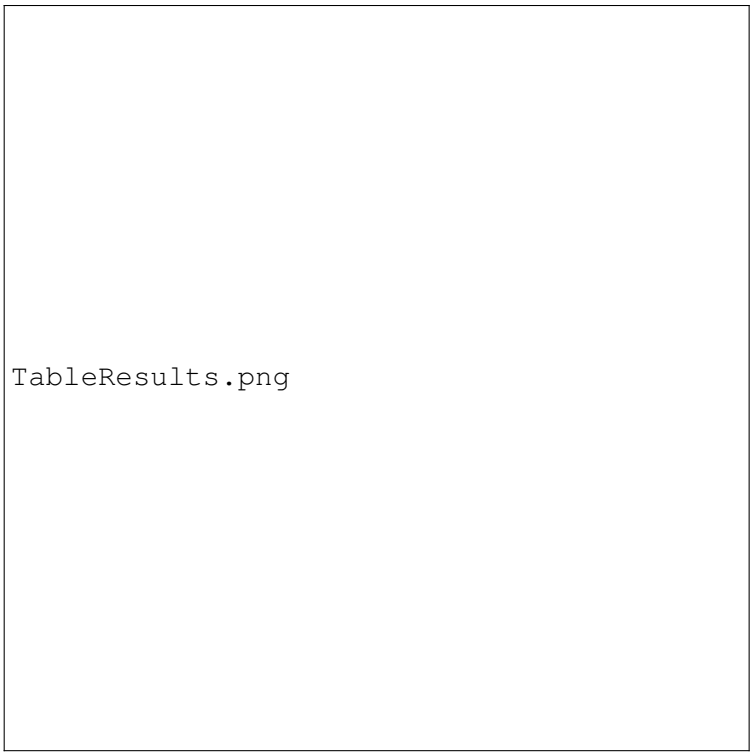

FIG. 8: Experimental and theoretical (using the single-cell data as inputs) results for both bacterial strains.