Analysis of Cell Size Homeostasis at the Single-Cell and Population Level

Philipp Thomas*

Department of Mathematics, Imperial College London, London, United Kingdom

Growth pervades all areas of life from single cells to cell populations to tissues. Cell size often fluctuates significantly from cell to cell and from generation to generation. Here we present a unified framework to predict the statistics of cell size variations within a lineage tree of a proliferating population. We analytically characterize (i) the distributions of cell size snapshots, (ii) the distribution within a population tree, and (iii) the distribution of lineages across the tree. Surprisingly, these size distributions differ significantly from observing single cells in isolation. In populations, cells seemingly grow to different sizes, typically exhibit less cell-to-cell variability and often display qualitatively different sensitivities to cell cycle noise and division errors. We demonstrate the key findings using recent single-cell data and elaborate on the implications for the ability of cells to maintain a narrow size distribution and the emergence of different power laws in these distributions.

Keywords: stochastic individual-based modeling, cell size control, size distributions, homeostasis, lineages, power laws

1. INTRODUCTION

Cells decide when to divide based on their size. A key question is therefore why cells grow to a certain size, how they maintain their sizes within a narrow distribution and what are the dominant sources of size variations. Recent experiments investigate this issue by taking snapshots of populations [1], trapping single cells [2] or recording movies of growing populations [3, 4]. Currently, we lack the tools with which to compare these different data.

It is known that timing divisions independently of cell size, a so-called timer control, does not lead to a stable size distribution when cells grow exponentially [5, 6]. Since exponential growth is ubiquitously observed in microbes [2, 3, 7–9], cells must control their division sizes to achieve homeostasis [10–12]. A possible strategy for such a control, called a sizer, is to set a stochastic threshold. Many microbes, however, rather grow by a constant size from birth to division, called the adder control [3, 13–15]. Other mixed strategies may be described by sizer- or timer-like controls [9, 15, 16].

Identifying which strategy cells employ is essential to understand how cells compensate for errors in size control and cell division. Early theory focussed on the distribution of cell size snapshots across a growing population, mainly due to the experimental limitations of tracking many cells over time [17–20]. These initial studies formulated sizer models based on the population balance equation, an integro-differential equation that is notoriously difficult to solve in practice [21–24]. With the advent of single-cell traps such as the mother machine [2], the theoretical focus moved toward describing single cells over many divisions [13, 25–27]. This path is more amenable to analysis because it considers only a single individual described by a discrete-time stochastic process or stochastic map [28]. It does not, however, apply to time-lapse movies monitoring cell growth in populations. A quantitative investigation of this issue is essential for our understanding of cellular physiology.

Here, we develop a comprehensive quantitative framework to predict single-cell and population statistics from models of cell size control. We focus on a stochastic branching process that produces a lineage tree (**Figure 1A**). We analyse (i) the size distributions of single dividing cells followed over time, (ii) the distribution in lineages across a population tree, (iii) the distribution across all cells in the tree, and (iv) the distributions of snapshots. Although these measures stem from the same lineage tree and the same underlying stochastic process, we find that they can give quantitatively and qualitatively different results when size varies from cell to cell. Our findings highlight the significance of population dynamics for the analysis of cell size control and size homeostasis.

2. ESTIMATION OF CELL SIZE DISTRIBUTIONS FROM TREE-STRUCTURED DATA

Estimation of cell size from snapshots samples all cells at a given instance in time. Time-lapse microscopy, however, offers various statistical analyses based on the underlying tree-structure of the data. A tree corresponds to an experiment starting from a single cell that resolves all ancestral relationships between cells in the population (see **Figure 1A**). A lineage is a subset of the tree representing a line of descent between the ancestral cell (root of the tree) and a cell of the final population (leave).

To estimate the tree and lineage statistics, we use the lineage counting method proposed by Nozoe et al. [29]. We denote by s_0^{ij} the size measurements at cell birth of the ith of D_j cells in the jth of N_L lineages of N_T trees and by D_j the number of cell divisions in the jth lineage. There are several ways in which these data can be pooled to represent either tree or lineage statistics. The corresponding distributions and their moments therefore generally represent weighted averages of the form

$$\bar{g} = \sum_{j=1}^{N_L} \sum_{i=1}^{D_j} w_{ij} g\left(s_0^{ij}\right),$$

where w_{ij} are the weights and the function g denotes the desired summary statistic. Choosing, for instance, g to be the Dirac delta function (or an appropriate binning function) gives a weighted cell size distribution. Similarly, choosing $g(x) = x^n$ we estimate its nth moment. In this article, we consider the distributions that correspond to the following weights:

$$\text{Forward lineage:} \quad w_{ij} = \frac{2^{-D_j}}{N_T D_j}, \quad (1a)$$

$$\text{Backward lineage:} \quad w_{ij} = \frac{1}{N_L D_j}, \quad (1b)$$

$$\text{Tree distribution:} \quad w_{ij} = \frac{1}{\Lambda_{ij}}. \quad (1c)$$

The *forward lineage* is a random path in a tree traversed from root to leaves in which each daughter cell is chosen with equal probability. The weights (1a) follow from the fact that the probability of choosing such a lineage from N_T binary trees is $2^{-D_j}/N_T$ and there are D_j measurements in the jth lineage (see [9, 29] for details). A similar lineage would be observed in the mother machine [2] assuming mother and daughter cells are statistically indistinguishable.

The *backward lineage* gives equal weight to each lineage and thus characterizes a typical lineage in a tree. The weights (1b) follow because there are D_j measurements in the jth lineage and N_L lineages. Such a lineage can be thought of choosing a random cell of the final population and traversing the tree from leave to root as described by Nozoe et al. [29], Wakamoto et al. [30], and Thomas [31].

Finally, the *tree distribution* gives equal weight to each birth event in the tree. Consequently, the factor Λ_{ij} denotes the number of occurrences of the values s_0^{ij} in the N_L lineages. Note that in practice this measure is computed easily by averaging all birth sizes in the trees. In the following, we develop an analytical framework to obtain these distributions from models of cell size control.

3. SIZE DISTRIBUTION OF A SINGLE CELL OVER TIME

As explained in the previous section, the forward lineage samples either daughter cell with equal probability. We can model such paths as stochastic maps with respect to the number of cell divisions D [28]. Denoting the birth size by $s_0^{(D)}$, division size by $s_d^{(D)}$ and inherited size fraction by $p^{(D)}$, we have

$$s_d^{(D)} = f(s_0^{(D)}) + \eta^{(D)}, \quad (2)$$

$$s_0^{(D+1)} = p^{(D)} s_d^{(D)}. \quad (3)$$

Here the deterministic function f determines how division size depends on birth size and the stochastic part $\eta^{(D)}$ is independent of birth size. An equivalent formulation is the one in terms of probabilities. In the limit of long times, the birth size distribution $\psi_{\text{fw}}(s) = \lim_{N\to\infty} \frac{1}{N} \sum_{D=1}^{N} \delta(s - s_0^{(D)})$ is stationary and satisfies the integral equation

$$\psi_{\text{fw}}(s) = \int_0^\infty dp \int_0^\infty ds_0 \frac{\pi(p)}{p} \varphi(s/p|s_0) \psi_{\text{fw}}(s_0). \quad (4)$$

The distribution $\varphi(s|s_0) = \varphi_\eta(s_d - f(s_0))$ can be written in terms of the distribution of the birth-size independent part of division size η and $\pi(p)$ is the distribution of inherited size fractions p. Because we follow each daughter cell with equal probability, the latter can be written in terms of the marginal distributions of the two daughter cells

$$\pi(p) = \frac{1}{2}\pi_1(p) + \frac{1}{2}\pi_2(p). \quad (5)$$

Because the mother's size is split between the daughters, we have $\pi_1(p) = \pi_2(1 - p)$ which implies that $E[p] = 1/2$ and also models the situation of asymmetric cell divisions. Hence on

FIGURE 1 | Characterization of cell size statistics from tree-structured data. **(A)** Lineage tree resulting from cell growth and division. The snapshot distribution characterizes the final state of the tree (gray). The distribution of birth sizes resulting from all cell divisions in the tree (black nodes) is the tree distribution. Forward lineages (red lineage) traverse the tree from root to leaf and follow randomly each daughter cell after division. Backward lineages start a randomly chosen cell in the final population and traverse the tree from leaf to root. **(B)** The snapshot distribution (gray) for the case of small division errors shows a maximum for newborn cells and decreases progressively toward larger cell sizes. Forward (red), backward lineage (blue) and tree distributions (green) of birth size are qualitatively similar. **(C)** For large fluctuations in cell size controls ($CV_\eta = 0.75$) the snapshot distribution broadens significantly (gray). Distributions of forward and backward lineages are almost indistinguishable, while the tree distribution (green) shows that cells are typically smaller compared to the lineage statistics. **(D)** For large division errors ($CV_p = 0.5$), forward, backward lineage and tree distributions appear all significantly different showing cells in forward lineages are typically smaller than in backward and tree distribution. The analytical results (solid lines) are verified using stochastic simulations in all cases (dots). Theory and simulations use adder rule ($a = 1$), gamma-distributed cell cycle noise with $E[\eta] = 1$ and $CV_\eta = 0.1$, and division errors following a symmetric beta distribution with $CV_p = 0.07$ unless stated otherwise.

average a daughter cell inherits precisely half of the mother's size, as expected.

4. SIZE DISTRIBUTIONS IN GROWING CELL POPULATIONS

While the theory of a single dividing cell is well-established, individual-based frameworks of growing cell populations are less commonly considered but date back to Koch and Schaechter [17]. We go on to develop this theory by deriving the governing equations for the population statistics. Assuming that cell division occurs at a rate $\gamma(\tau, s)$ that depends on the present cell age τ and size s, the rate of change in the number of cells with that age and size $n(\tau, s)$ is

$$\frac{d}{dt} n(\tau(t), s(t), t) = -\gamma(\tau(t), s(t)) n(\tau(t), s(t), t).$$

We further assume that cells grow exponentially with growth rate α

$$s(\tau) = s_0 e^{\alpha \tau},$$

as has been reported for many microbes such as *Escherichia coli* [2], *Caulobacter crescentus* [3], *B. subtilis* [7], *M. smegmatis* [9]

and *S. cerevisiae* [8]. Using this dependence we can expand the total derivative to obtain

$$\left[\frac{\partial}{\partial t} + \frac{\partial}{\partial \tau} + \frac{\partial}{\partial s} \alpha s \right] n(\tau, s, t) = -\gamma(\tau, s) n(\tau, s, t). \quad (6)$$

To account for cell births the cell density must obey the boundary condition

$$n(0, s, t) = 2 \int_0^\infty ds_d \int_0^\infty d\tau \, \mathcal{K}(s|s_d) \gamma(\tau, s_d) n(\tau, s_d, t). \quad (7)$$

The condition implements the fact that the number of newborn cells in the population must equal twice the number of dividing cells. The division kernel $\mathcal{K}(s|s_d)$ denotes the transition probability for a cell of size s_d to inherit size s after cell division. Most conveniently, we can write it in terms of the inherited size fraction via

$$\mathcal{K}(s|s_d) = \int_0^1 dp \, \pi(p) \delta(s - p s_d), \quad (8)$$

where $\pi(p)$ is defined as in Equation (5) and also takes into account a possible asymmetry between the daughter cells. Similar formulations have been used to characterize distributions of size-structured populations [21, 23, 32]. In the following, we derive

a strategy of how to obtain the various population statistics analytically. We will do this iteratively by writing the snapshot distribution in terms of the tree distribution, and writing the tree distribution in terms of the backward lineage distribution that can be efficiently solved.

4.1. Snapshot Distribution
In balanced growth conditions the total number of cells increases exponentially, a condition which dramatically simplifies the analysis. In this limit, the cell density follows

$$n(\tau, s, t) \sim e^{\lambda t} \Pi(\tau, s). \tag{9}$$

The distribution $\Pi(\tau, s)$ is the probability density of observing a cell with age τ and size s in a snapshot, which is independent of time. It can further be shown that the asymptotic population growth rate λ equals the rate of cell-growth α [31]. Using this ansatz in Equation (6) and employing the method of characteristics, we find that the snapshot density satisfies

$$\Pi(\tau, s) = \Pi(0)\Pi(se^{-\alpha\tau}|0)e^{-2\alpha\tau} \int_{\tau}^{\infty} d\tau_d \, \varphi(\tau_d|se^{-\alpha\tau}), \tag{10}$$

where $\Pi(s_0|0)$ is the distribution of birth sizes and

$$\varphi(\tau_d|s_0) = \gamma(\tau_d, s_0 e^{\alpha\tau_d})e^{\int_0^{\tau_d} d\tau \gamma(\tau, s_0 e^{\alpha\tau})}$$

is the distribution of division times for a cell of given birth size. Integrating Equation (6) and using the boundary condition (7) shows that $\Pi(0) = 2\alpha$.

To make the dependence on birth size explicit, we change variables from τ to s_0. The result is the snapshot distribution of cell size and birth size, which reads

$$\Pi(s, s_0) = 2\Pi(s_0|0)\phi(s|s_0)\frac{s_0}{s^2}, \tag{11}$$

where $\phi(s|s_0) = 1 - \int_{s_0}^{s} ds' \varphi(s'|s_0)$ is the probability that a cell born with size s_0 has not divided before reaching size s. The snapshot distribution (11) depends explicitly on the distribution of birth sizes, which we characterize in the following. In the absence of cell division errors and cell cycle noise, the birth size s_0 is deterministic and the snapshot distribution reduces to $\Pi(s) = 2s_0/s^2$ for $s_0 \leq s \leq 2s_0$ as shown by [17].

4.2. Tree Distribution
In balanced growth conditions, the fraction of cells with a certain size is constant. Hence, the birth size distribution in a snapshot is also the distribution of birth sizes in the population tree

$$\psi_{\text{tree}}(s_0) = \Pi(s_0|0). \tag{12}$$

To characterize this measure, we substitute Equations (9) and (10) into the boundary condition (7) to find

$$\psi_{\text{tree}}(s_0) = 2\int_0^{\infty} ds_d \int_0^{\infty} d\tau \, \mathcal{K}(s|s_d)e^{-2\alpha\tau}\varphi(\tau|s_d e^{-\alpha\tau})\psi_{\text{tree}}(s_d e^{-\alpha\tau}). \tag{13}$$

Changing the integration from τ to $s_0 = e^{-\alpha\tau}s_d$ and using $\varphi(\tau|s_0)d\tau = \frac{s_d}{s_0}\varphi(s_d|s_0)ds_0$, we finally arrive at

$$\psi_{\text{tree}}(s) = 2\int_0^{\infty} ds_0 \int_{s_0}^{\infty} ds_d \, \mathcal{K}(s|s_d)\left(\frac{s_0}{s_d}\right)\varphi(s_d|s_0)\psi_{\text{tree}}(s_0), \tag{14}$$

which characterizes the tree distribution. The distribution $\varphi(s_d|s_0)$ is the distribution of division sizes that describes the cell size control as in Equation (4).

4.3. Backward Lineage Distribution
The equation for the tree distribution has no intuitive stochastic process interpretation. However, we notice that substituting

$$\psi_{\text{bw}}(s_0) = s_0 \psi_{\text{tree}}(s_0)E_{\text{bw}}[s_0^{-1}], \tag{15}$$

where $E_{\text{bw}}[s_0^{-1}]$ is a normalizing factor, transforms Equation (14) into

$$\psi_{\text{bw}}(s) = \int_0^{\infty} dp \int_0^{\infty} ds_0 \frac{\pi_{\text{bw}}(p)}{p}\varphi(s/p|s_0)\psi_{\text{bw}}(s_0), \tag{16}$$

with

$$\pi_{\text{bw}}(p) = 2p\pi(p). \tag{17}$$

This equation provides such a simple stochastic interpretation because it is of the same form as the forward lineage equation (4), except that π_{bw} replaces π. In fact, it describes the distribution of birth sizes in a backward lineage. Interestingly, the inherited size fraction in backward lineages π_{bw} is skewed toward cells inheriting larger size fractions. It thus appears as if cells are dividing asymmetrically in backward lineages. The impact of this phenomenon on the cell size distributions is explored in the following.

4.4. Comparison of Cell Size Distributions
A phenomenological linear relation between birth and division size has been observed in experiments [9, 15, 28]

$$s_d = as_0 + \eta. \tag{18}$$

To incorporate these dependencies into our stochastic framework, we set $\varphi(s_d|s_0) = \varphi_\eta(s_d - as_0)$. The parameter a denotes different models of cell size control and φ_η the size control distribution. For $a = 0$ division size is independent of birth size, called the sizer mechanism. For $a = 1$ the size added from birth to division varies independently of birth size, an adder mechanism that is commonly observed in bacteria [3, 10, 13]. For $a = 2$ division size is directly proportional to birth size, which resembles cell-size control based on division timing. Intermediate values would behave either sizer- ($0 < a < 1$) or timer-like ($1 < a < 2$).

Equation (16) could in principle be solved by sampling the stochastic maps given in section 3. More effective, however, is to discretise it and solve the resulting set of linear equations

numerically. This procedure can in principle be used for arbitrary distributions of division errors and size control. In **Figures 1B–D** we show the result of this procedure for beta-distributed division errors and gamma-distributed cell size control. The snapshot distribution is obtained using the numerical result for ψ_{bw} together with Equation (16) in (11).

When errors in cell division and size control are small (**Figure 1B**), the snapshot distribution displays a maximum at newborn sizes and a long decaying shoulder toward the division size consistent with the fact that in populations the number of newborn cells must equal twice the number of dividing cells. The numerical results for the backward distribution ψ_{bw} shown in **Figure 1B** are qualitatively very similar to the tree distribution obtained via Equation (15) and also to the lineage distribution computed by solving Equation (4).

As the fluctuations in size control increase (**Figure 1C**), the snapshot distribution broadens and the birth size distributions become skewed. The distributions of forward and backward lineage are nearly indistinguishable, but the tree distribution leans toward significantly smaller sizes. For large division errors but small fluctuations in size control (**Figure 1D**), the snapshot distribution also broadens, but the size distributions in forward (red), backward lineages (blue) and population trees (green) all differ significantly in this condition. Cells in backward lineages are typically bigger than in the population tree and these are bigger than in forward lineages.

We verify the accuracy of our predictions by agent-based simulations. To this end, we stochastically sample division sizes using Equations (18), calculate the interdivision times $\tau_d = \ln(s_d/s_0)/\alpha$, determine the birth size after division from Equation (3) and assign the remaining size to the other daughter. The distributions from the simulated population trees (dots in **Figure 1**) are in excellent agreement with the theory (solid lines) in all cases.

The predicted differences in the birth sizes can be understood intuitively from the associated interdivision times of cells. In the presence of division errors, cells born larger have shorter division times, and they are hence over-represented in the population tree and the backward lineages when compared to forward lineages. On the contrary, when only cell size control fluctuates, cells dividing at smaller sizes are over-represented in the population tree. Curiously, our theory predicts that the latter phenomenon affects only the tree statistics but not the backward lineage. This is because the lineage statistics have the same division size controls φ (cf. Equations 4, 16). In the following, we provide a quantitative analysis of these effects based on the cell size moments.

5. CONDITIONS FOR CELL SIZE HOMEOSTASIS

5.1. Mean Cell Size and Cell-To-Cell Variability

To obtain expressions for the cell-size moments in a forward lineage, we multiply Equation (4) by s^n, perform

the integration and employ the binomial theorem. The result is

$$E_{\text{fw}}[s_0^n] = E[p^n] \sum_{k=0}^{n} \binom{n}{k} E[\eta^k] a^{n-k} E_{\text{fw}}[s_0^{n-k}], \quad (19)$$

which depends on the moments $E[\eta^k]$ of the birth-size independent part of the division size (cf. Equation 18). Solving this relation for the nth moment $E_{\text{fw}}[s_0^n]$ allows to express the cell size moments recursively in terms of lower order moments

$$E_{\text{fw}}[s_0^n] = \frac{E[p^n]}{1 - a^n E[p^n]} \sum_{k=1}^{n} \binom{n}{k} a^{n-k} E[\eta^k] E_{\text{fw}}[s_0^{n-k}]. \quad (20)$$

Using the above equation, we have explicit expressions for the mean cell size and its fluctuations. Similarly, the moments in backward lineages $E_{\text{bw}}[s_0^n]$ are obtained by replacing $E[p^n]$ with $E_{\text{bw}}[p] = 2E[p^{n+1}]$ (cf. Equation 17), and the moments of the tree distribution follow from $E_{\text{tree}}[s_0^n] = E_{\text{bw}}[s_0^{n-1}]/E_{\text{bw}}[s_0^{-1}]$ (cf. Equation 15). In summary, the mean size and its coefficient of variation ($\text{CV}_x^2 = \text{Var}[s_0]/E^2[s_0]$) are given by

Forward lineage:

$$E_{\text{fw}}[s_0] = \frac{E[\eta]}{2 - a},$$

$$\text{CV}_{\text{fw}}^2[s_0] = \frac{(a-2)^2 \text{CV}_\eta^2 \left(\text{CV}_p^2 + 1\right) + 4\text{CV}_p^2}{4 - a^2(1 + \text{CV}_p^2)}, \quad (21a)$$

Backward lineage:

$$E_{\text{bw}}[s_0] = \frac{E[\eta]\left(\text{CV}_p^2 + 1\right)}{2 - a(1 + \text{CV}_p^2)},$$

$$\text{CV}_{\text{bw}}^2[s_0] = \frac{4\text{CV}_p^2\left(\text{CV}_p^2 - 1\right) - \text{CV}_\eta^2\left(3\text{CV}_p^2 + 1\right)\left(a\text{CV}_p^2 + a - 2\right)^2}{\left(\text{CV}_p^2 + 1\right)^2\left(3a^2\text{CV}_p^2 + a^2 - 4\right)},$$

$$(21b)$$

Tree distribution:

$$E_{\text{tree}}[s_0] = E_{\text{bw}}^{-1}[s_0^{-1}],$$

$$\text{CV}_{\text{tree}}^2[s_0] = \frac{\left(1 + \text{CV}_p^2\right)\left(a + E[\eta]E_{\text{bw}}[s_0^{-1}]\right) - 2}{2 - a(1 + \text{CV}_p^2)}. \quad (21c)$$

These expression are exact but provide exact analytical insights only to the forward and backward lineage statistics, because the tree mean cannot be obtained in closed form and has to be approximated. In what follows, we provide a discussion of quantitative differences between these different statistics, analyse their sensitivities to different noise sources, and compare our findings to experimental single-cell data.

5.1.1. Quantitative Differences of the Cell-Size Statistics

We first compare the lineage statistics. Interestingly, the mean cell size in a forward lineage is independent of the division error CV_p while it increases with CV_p in backward lineages confirming our previous findings (**Figure 1D**). We illustrate this dependence for the sizer ($a = 0$), adder ($a = 1$) and a timer-like size control ($a = 1.5$) in **Figure 2A**. It is seen that the mean cell size in backward lineages is more sensitive to the cell size control parameter a. In **Figure 2B** we compare the dependence of the coefficient of variation of size fluctuations for different size controls and division errors. Importantly, despite the sensitivity of the mean cell size to division errors, for sizer and adder the coefficient of variation is always smaller in the backward than in the forward lineage. It can be shown using Equations (21) that this holds whenever $a \leq 1$.

In the tree statistics, cells are expected to be smaller in the tree statistics than in backward lineages

$$E_{\text{tree}}[s_0] \leq E_{\text{bw}}[s_0], \tag{22}$$

which follows using Jensen's inequality in Equation (21c). To compare the tree statistics with the forward statistics, we approximate the tree mean assuming small fluctuations by expanding Equation (21c) about $E_{\text{bw}}[s_0]$ and substituting the backward moments. A straight-forward calculation including also the skewness skew$[\eta]$ of the size control distribution gives

$$E_{\text{tree}}[s_0] \approx E_{\text{fw}}[s_0]\left(1 + \frac{2a}{4-a^2}CV_p^2 - \frac{2-a}{a+2}CV_\eta^2 \right.$$
$$\left. + \frac{(2-a)^2\text{skew}[\eta]}{a^2+2a+4}CV_\eta^3\right), \tag{23a}$$

$$CV_{\text{tree}}^2[s_0] \approx \frac{4}{4-a^2}CV_p^2 + \frac{2-a}{a+2}CV_\eta^2 - \frac{(2-a)^2\text{skew}[\eta]}{a^2+2a+4}CV_\eta^3. \tag{23b}$$

Note that by Equation (8) division errors do not posses skewness and hence there is no corresponding term.

By inspecting the second and third term in the approximation of the tree mean, we observe that division errors increase the mean while errors in the size control decrease it relative to the forward lineage average. Interestingly, the tree mean of the sizer control ($a = 0$) decreases with size control errors but is insensitive to division errors, while a timer control ($a = 2$) amplifies division errors (**Figure 2A**). For the adder control ($a = 1$), the tree mean can either be smaller or larger than the forward lineage average depending on the relative the relative size of the noise sources (**Figure 2A**).

The differences in the coefficients of variation of the tree statistic and the forward lineage are determined by the skewness skew$[\eta]$ of the size control distribution. This follows because the first two terms in Equation (23b) equal the linearized coefficient of variation in forward lineages. Negative skewness hence increases the noise levels in the tree statistics, while

positive skewness decreases it. The latter case is shown in **Figure 2B** where we find that noise levels are smaller than forward lineage ones. Considering the fact that backward lineages also have smaller noise levels than forward lineages for $a \leq 1$, this suggests a possible mechanism for noise reduction in populations.

5.1.2. Sensitivity Analysis

To investigate the different lineage statistics further, we compute the sensitivities of the coefficient of variation given in Equations (21) to the different noise sources

$$S_p(CV_\eta^2) = \left.\frac{\partial CV^2[s_0]}{\partial CV_p^2}\right|_{CV_p^2=0}, \quad S_\eta(CV_p^2) = \left.\frac{\partial CV^2[s_0]}{\partial CV_\eta^2}\right|_{CV_\eta^2=0}. \tag{24}$$
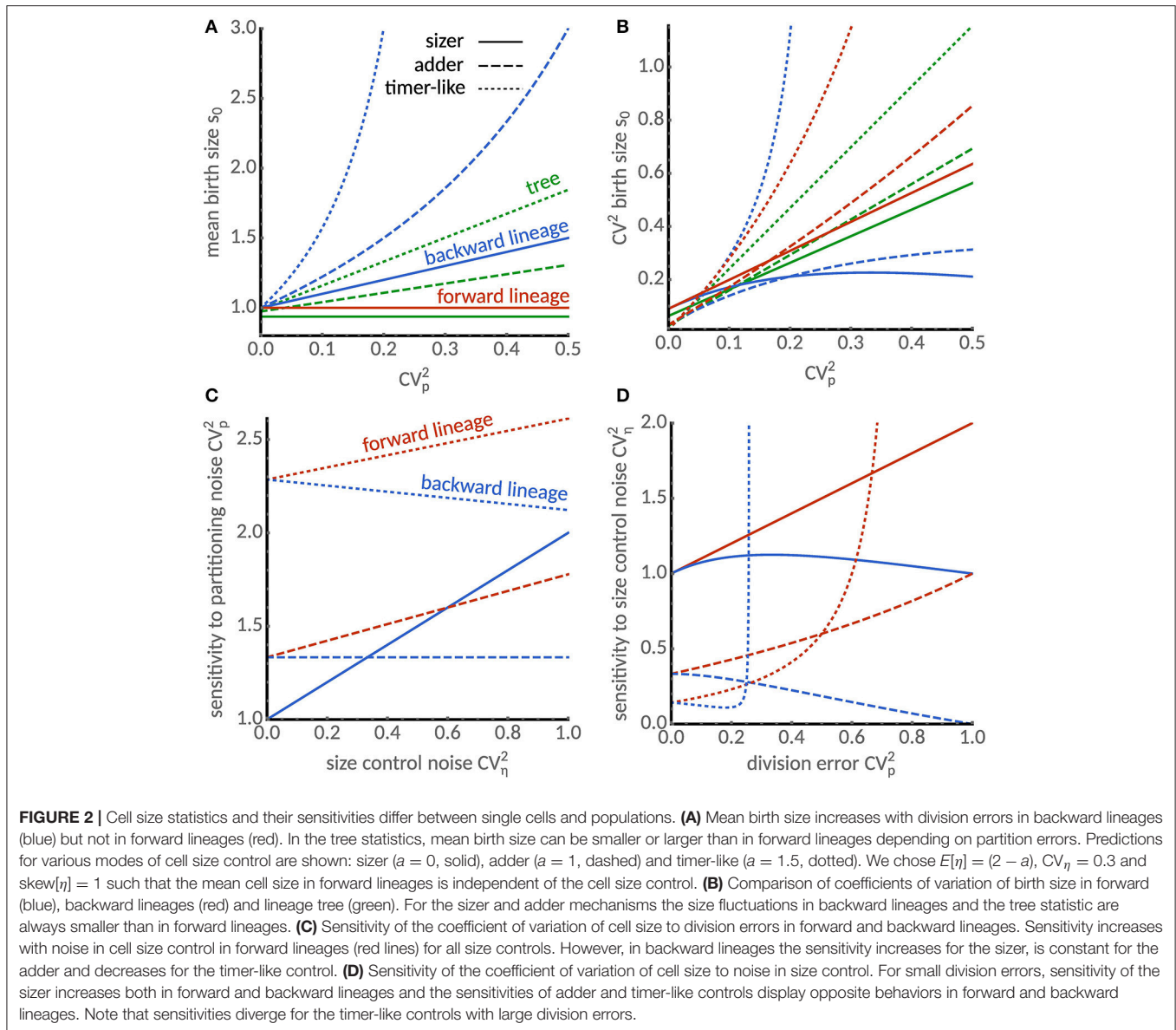
We find that the sensitivity to division errors $S_p(CV_\eta^2)$ shown in **Figure 2C** increases with noise in cell size control in forward lineages (red lines) for all size controls. However, in backward lineages, the sensitivity increases only for the sizer, it is constant for the adder and decreases for the timer control.

The sensitivity of size variation to noise in cell size control $S_\eta(CV_p^2)$ shown in **Figure 2D** increases for the sizer mechanism with small division errors both in forward and backward lineages, but exhibits opposite dependencies for the adder and timer-like controls in the respective lineage statistics. We also observe that an instability occurs for timer-like mechanisms as described previously [25, 27]. Interestingly, this instability occurs for smaller values of a in the backward lineages than those previously reported for forward lineages. This suggests that different conditions for cell size homeostasis apply to forward and backward lineages but also to the tree distribution. Before analyzing these conditions in section 5.2, we compare the predictions of our theory to single-cell data.

5.1.3. Application to Single-Cell Experiments

To validate the predictions of our theory, we analyse a recent dataset acquired using time-lapse microscopy of the mycobacterium *S. smegmatis* [9]. Its rod-shape allows us to use of cell length as a proxy for cell size. **Figure 3** summarizes the lineage-weighted statistics of mean lengths and the corresponding coefficients of variation as introduced in section 2. We find that mean length is higher in the tree statistic and backward lineages than in forward lineages (**Figure 3A**). Similarly, we observe smaller cell-to-variability in the tree and backward lineages (**Figure 3B**). The noise reduction in the backward lineage noise is roughly 10%. These findings qualitatively confirm the predicted differences in the cell size statistics of populations.

To quantitatively compare these measurements with the proposed theory, we fit the expressions in Equations (21a), (21b) and (23) using nonlinear least squares. Because cells grown in glycerol follow the adder rule, we fix $a = 1$ [9]. The best-fit (see caption of **Figure 3** for parameters) provides reasonable agreement with the measured quantities

FIGURE 2 | Cell size statistics and their sensitivities differ between single cells and populations. **(A)** Mean birth size increases with division errors in backward lineages (blue) but not in forward lineages (red). In the tree statistics, mean birth size can be smaller or larger than in forward lineages depending on partition errors. Predictions for various modes of cell size control are shown: sizer ($a = 0$, solid), adder ($a = 1$, dashed) and timer-like ($a = 1.5$, dotted). We chose $E[\eta] = (2 - a)$, $CV_\eta = 0.3$ and skew$[\eta] = 1$ such that the mean cell size in forward lineages is independent of the cell size control. **(B)** Comparison of coefficients of variation of birth size in forward (blue), backward lineages (red) and lineage tree (green). For the sizer and adder mechanisms the size fluctuations in backward lineages and the tree statistic are always smaller than in forward lineages. **(C)** Sensitivity of the coefficient of variation of cell size to division errors in forward and backward lineages. Sensitivity increases with noise in cell size control in forward lineages (red lines) for all size controls. However, in backward lineages the sensitivity increases for the sizer, is constant for the adder and decreases for the timer-like control. **(D)** Sensitivity of the coefficient of variation of cell size to noise in size control. For small division errors, sensitivity of the sizer increases both in forward and backward lineages and the sensitivities of adder and timer-like controls display opposite behaviors in forward and backward lineages. Note that sensitivities diverge for the timer-like controls with large division errors.

considering the experimental error bars. We expect that higher sample sizes and developed models of the size control accounting for the known asymmetric growth of old and new pole cells [9] could improve the quantitative agreement.

5.2. Moment Conditions for Cell Size Homeostasis

Next, we investigate the conditions for cell size homeostasis on the basis of the existence of certain cell size moments. From the recursive relation in Equation (20) we observe that if a certain moment is bounded, all lower order moments must be finite as well. Specifically, we can deduce the condition for the *n*th moment in forward lineages to be bounded by inspecting the denominator in Equation (20). Conditions for the backward lineages follow from substituting $E[p^n] \rightarrow 2E[p^{n+1}]$ and
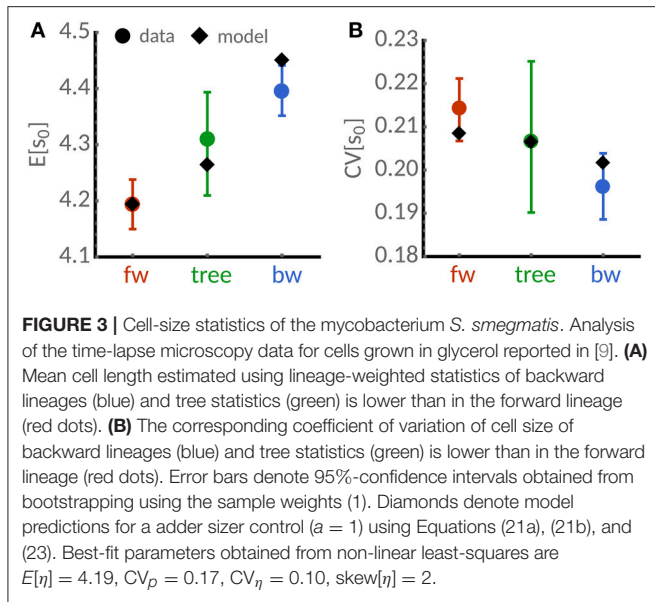
subsequently $n \rightarrow n - 1$ for the tree distribution. In summary, the conditions for the existence of the *n*th moment are:

$$\text{Forward lineage:} \quad 1 > a^n E[p^n], \quad (25a)$$

$$\text{Backward lineage:} \quad 1 > 2a^n E[p^{n+1}], \quad (25b)$$

$$\text{Tree distribution:} \quad 1 > 2a^{n-1} E[p^n]. \quad (25c)$$

We make three important observations: (i) The above conditions are independent of the noise in size control but are identical only when the divisions are perfect $p = 1/2$; (ii) $a < 1$ guarantees all moments to be finite; and (iii) there exists a hierarchy between the different conditions. A finite moment of tree distribution implies the same moment to be finite in the forward lineage. A finite moment in the backward lineage implies the same moment to be

FIGURE 3 | Cell-size statistics of the mycobacterium *S. smegmatis*. Analysis of the time-lapse microscopy data for cells grown in glycerol reported in [9]. **(A)** Mean cell length estimated using lineage-weighted statistics of backward lineages (blue) and tree statistics (green) is lower than in the forward lineage (red dots). **(B)** The corresponding coefficient of variation of cell size of backward lineages (blue) and tree statistics (green) is lower than in the forward lineage (red dots). Error bars denote 95%-confidence intervals obtained from bootstrapping using the sample weights (1). Diamonds denote model predictions for a adder sizer control ($a = 1$) using Equations (21a), (21b), and (23). Best-fit parameters obtained from non-linear least-squares are $E[\eta] = 4.19$, $CV_p = 0.17$, $CV_\eta = 0.10$, skew$[\eta] = 2$.

finite in both the forward lineage and the tree distribution. This hierarchy follows because $2a^n E[p^{n+1}] \leq 2a^{n-1} E[p^n] \leq a^n E[p^n]$.

We will briefly discuss the marginal case of the adder rule ($a = 1$) because it models the size control of many microbes. Its moments are always finite in the forward lineage but not in the backward lineage and tree distribution. This extreme situation ($CV_p^2 = 1$) appears, however, only when one daughter inherits all of the mother's size resulting in a non-growing cell and one that grows without bound. Such a situation has been observed in minicell-producing mutants of *E. coli*, but these cells also allow for divisions at mid-cell reinforcing their overall viability [33]. In the following, we investigate in more detail the conditions for finite mean and variance.

5.2.1. Conditions for Finite Mean Size and Cell-To-Cell Variability

The conditions for the cell size distributions having finite mean values are found by letting $n = 1$ in Equations (25a), which leads to

$$\text{Forward lineage:} \quad a < 2, \tag{26a}$$

$$\text{Backward lineage:} \quad a < \frac{2}{1 + CV_p^2}, \tag{26b}$$

$$\text{Tree distribution:} \quad a < e^{-2E_\pi[p \ln p]}. \tag{26c}$$

The first and second conditions are obtained from noting that $E[p] = 1/2$ and $E[p^2] = \frac{1}{4}(1 + CV_p^2)$. The third condition is not immediately obvious since condition (25c) does not provide direct information about the mean. The result can be obtained by perturbing (25c) by $n \to 1 + \epsilon$ and requiring its leading order term to be positive.

Note that the condition for the forward lineage depends only on the cell size control. Interestingly, this is not the case for the backward lineage and tree distributions, which depend intricately

on the distribution of division errors. We summarize these findings in **Figure 4A**, which shows the regions in parameter space a and CV_p^2 for which the first two moments are finite. We observe that a finite mean in the backward lineage implies a finite mean in the other measures. Similarly, a finite mean of the tree distribution implies the same for the forward lineage but not for the backward lineage, as discussed in the previous section.

Similarly, the conditions for existence of the second moments can be investigated, which hold identically for the coefficients of variation of cell size. The result is given by

$$\text{Forward lineage:} \quad a^2 < \frac{4}{1 + CV_p^2}, \tag{27a}$$

$$\text{Backward lineage:} \quad a^2 < \frac{4}{1 + 3CV_p^2}, \tag{27b}$$

$$\text{Tree distribution:} \quad a < \frac{2}{1 + CV_p^2}. \tag{27c}$$

These conditions are obtained using $E[p^2] = \frac{1}{4}(1 + CV_p^2)$ and noting that $\pi(p)$ is symmetric about the mean and hence has no skewness leading to $E[p^3] = \frac{1}{8}(1 + 3CV_p^2)$. In **Figure 4A** we illustrate the resulting regions where forward, backward lineages and tree variance are bounded.

5.3. The Distributions of Timer-Like Controls Display Different Power Laws in Single Cells and Populations

Divergent moments can be indicative of rare cell sizes and distributions with power law tails. To investigate the emergence of power laws in the different statistical measures, we assume that for large cell sizes the distributions follow $\sim s_0^{-\alpha}$ with power-law exponent α. The case $\alpha = \infty$ corresponds to exponential tails while finite values of α indicate a power law. These can be deduced from the moments conditions because for a given power-law exponent α all moments of order $\alpha - 1$ and above are unbounded. Thus from the conditions (25), we find that the power-law exponents satisfy the following equations

$$\text{Forward lineage:} \quad 1 = a^{\alpha-1} E[p^{\alpha-1}], \tag{28a}$$

$$\text{Backward lineage:} \quad 1 = 2a^{\alpha-1} E[p^\alpha], \tag{28b}$$

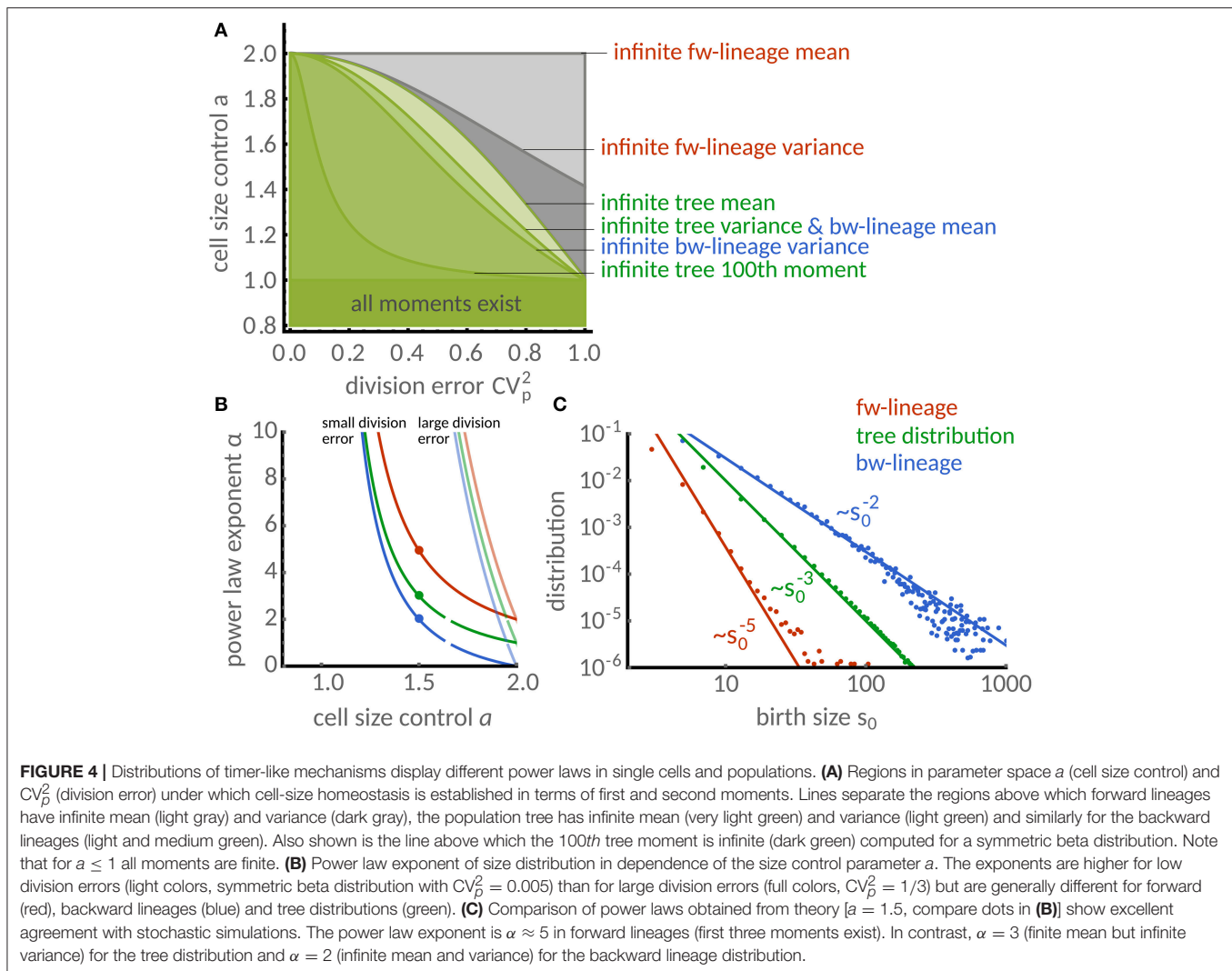$$\text{Tree distribution:} \quad 1 = 2a^{\alpha-2} E[p^{\alpha-1}], \tag{28c}$$

for $\alpha > 1$. Because of the inequality given after Equation (25), it follows that

$$\alpha_{\text{fw}} \geq \alpha_{\text{tree}} \geq \alpha_{\text{bw}}, \tag{29}$$

independently of the distribution of division errors. Note that the exponents are equal when divisions are perfect ($p = 1/2$).

However, it is difficult to give an explicit expression for the exponents without assuming a particular distribution $\pi(p)$ of the division errors. In **Figure 4B**, we study the dependence of the power-law exponent α on the cell size control parameter a for small and large division errors following a symmetric Beta-distribution. We find that the tails are exponential ($\alpha = \infty$)

FIGURE 4 | Distributions of timer-like mechanisms display different power laws in single cells and populations. **(A)** Regions in parameter space a (cell size control) and CV_p^2 (division error) under which cell-size homeostasis is established in terms of first and second moments. Lines separate the regions above which forward lineages have infinite mean (light gray) and variance (dark gray), the population tree has infinite mean (very light green) and variance (light green) and similarly for the backward lineages (light and medium green). Also shown is the line above which the 100th tree moment is infinite (dark green) computed for a symmetric beta distribution. Note that for $a \leq 1$ all moments are finite. **(B)** Power law exponent of size distribution in dependence of the size control parameter a. The exponents are higher for low division errors (light colors, symmetric beta distribution with $CV_p^2 = 0.005$) than for large division errors (full colors, $CV_p^2 = 1/3$) but are generally different for forward (red), backward lineages (blue) and tree distributions (green). **(C)** Comparison of power laws obtained from theory [$a = 1.5$, compare dots in **(B)**] show excellent agreement with stochastic simulations. The power law exponent is $\alpha \approx 5$ in forward lineages (first three moments exist). In contrast, $\alpha = 3$ (finite mean but infinite variance) for the tree distribution and $\alpha = 2$ (infinite mean and variance) for the backward lineage distribution.

whenever $a < 1$. For $a > 1$, however, the exponents of forward, backward lineages and tree distributions are similar only for small division errors and they deviate substantially for large division errors.

To study this case in more detail, we focus on the timer-like case $a = 1.5$ with large division errors. In **Figure 4C** we compare the predicted power laws with stochastic simulations. The power law exponent of the forward lineage distribution is $\alpha \approx 5$ meaning that only the first 3 moments exist. In contrast, the tree distribution shows the exponent $\alpha = 3$, which translates to finite mean cell size but unbounded variance. Interestingly, the backward lineage distribution has exponent $\alpha = 2$, which does not admit a finite mean cell size. In all cases, the simulations (dots) are in excellent agreement with the theoretical exponents (lines).

6. DISCUSSION

Cell size has been traditionally quantified by snapshots, but recent advances in time-lapse microscopy allow to track the growth of hundreds to thousands of cells in a microcolony. These techniques allow to construct entire lineage trees and resolve genealogies and thus offer various statistics with which to quantify individual cell sizes. We presented a unified modeling framework to quantify and compare size distributions in these populations.

We demonstrated that cells in populations exhibit different cell size, often exhibit lower cell-to-cell variability and display different sensitivities to division errors and cell cycle noise compared to cells in isolation. Specifically, we found that positive skewness in the size control distribution decreases cell-to-cell variability in the population tree. In fact, many bacteria implementing the adder control such as *E. coli*, *B. subtilis* [7], and *S. smegmatis* [9], show positively skewed length increments. We also observed that whenever $a \leq 1$, the coefficient of variation of cell size in backward lineages is smaller than in forward lineages. Since backward lineages represent typical lineages in a population tree, we speculate that populations implementing adder or sizer controls, which are ubiquitously found in microbes, could exploit these differences to narrow their size distributions.

A general requirement for size homeostasis, regardless of the model, is that cells born small grow proportionally more than cells born with larger sizes [11]. These larger cells proliferate quicker and are therefore over-represented in populations, which explains the discrepancies we found between different size distributions, their associated moments and power laws in the presence of division errors. We hence expect our conclusions to hold also for developed models of cell size control. We found that these differences are particularly sensitive to division errors, which can be extremely heterogeneous when cells respond to stress [4]. Conversely, we found that cells dividing at smaller sizes are over-represented in the population tree when division size is highly variable. We anticipate that this effect will be important when cell size control is modulated by time-dependent environments [34].

Finally, we elucidated that isolated cells tracked over many generations and cells in populations exhibit different sensitivities to division errors and noise in cell cycle control. Experimental devices such as mother machines and agar pads may thus give qualitatively different conclusions about a cell's ability to control its size and the resulting population heterogeneity. Similar effects have been described for stochastic gene expression [35]. However, our findings raise a more fundamental question: what type of size distribution cells may have evolved to control?

AUTHOR CONTRIBUTIONS

PT designed the study, carried out the analysis, and wrote the paper.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

1. Si F, Li D, Cox SE, Sauls JT, Azizi O, Sou C, et al. Invariance of initiation mass and predictability of cell size in *Escherichia coli*. *Curr Biol.* (2017) **27**:1278–87. doi: 10.1016/j.cub.2017.03.022

2. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, et al. Robust growth of *Escherichia* coli. *Curr Biol.* (2010) **20**:1099–103. doi: 10.1016/j.cub.2010.04.045

3. Campos M, Surovtsev IV, Kato S, Paintdakhi A, Beltran B, Ebmeier SE, et al. A constant size extension drives bacterial cell size homeostasis. *Cell* (2014) **159**:1433–46. doi: 10.1016/j.cell.2014.11.022

4. Wehrens M, Ershov D, Rozendaal R, Walker N, Schultz D, Kishony R, et al. Size laws and division ring dynamics in filamentous *Escherichia coli* cells. *Curr Biol.* (2018) **28**:972–9.e5. doi: 10.1016/j.cub.2018.02.006

5. Bell GI. Cell growth and division: III. Conditions for balanced exponential growth in a mathematical model. *Biophys J.* (1968) **8**:431–44.

6. Hannsgen KB, Tyson JJ, Watson LT. Steady-state size distributions in probabilistic models of the cell division cycle. *SIAM J Appl Math.* (1985) **45**:523–40.

7. Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, et al. Cell-size control and homeostasis in bacteria. *Curr Biol.* (2015) **25**:385–91. doi: 10.1016/j.cub.2014.12.009

8. Soifer I, Robert L, Amir A. Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy. *Curr Biol.* (2016) **26**:356–61. doi: 10.1016/j.cub.2015.11.067

9. Priestman M, Thomas P, Robertson BD, Shahrezaei V. Mycobacteria modify their cell size control under sub-optimal carbon sources. *Front Cell Dev Biol.* (2017) **5**:64. doi: 10.3389/fcell.2017.00064

10. Taheri-Araghi S, Brown SD, Sauls JT, McIntosh DB, Jun S. Single-cell physiology. *Annu Rev Biophys.* (2015) **44**:123–42. doi: 10.1146/annurev-biophys-060414-034236

11. Willis L, Huang KC. Sizing up the bacterial cell cycle. *Nat Rev Microbiol.* (2017) **15**:606–20. doi: 10.1038/nrmicro.2017.79

12. Schmoller KM. The phenomenology of cell size control. *Curr Opin Cell Biol.* (2017) **49**:53–8.

13. Amir A. Cell size regulation in bacteria. *Phys Rev Lett.* (2014) **112**:208102. doi: 10.1103/PhysRevLett.112.208102

14. Deforet M, van Ditmarsch D, Xavier JB. Cell-size homeostasis and the incremental rule in a bacterial pathogen. *Biophys J.* (2015) **109**:521–8. doi: 10.1016/j.bpj.2015.07.002

15. Sauls JT, Li D, Jun S. Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Curr Opin Cell Biol.* (2016) **38**:38–44. doi: 10.1016/j.ceb.2016.02.004

16. Banerjee S, Lo K, Daddysman MK, Selewa A, Kuntz T, Dinner AR, et al. Biphasic growth dynamics control cell division in *Caulobacter crescentus*. *Nat Microbiol.* (2017) **2**:17116. doi: 10.1038/nmicrobiol.2017.116

17. Koch A, Schaechter M. A model for statistics of the cell division process. *Microbiology* (1962) **29**:435–54.

18. Collins J, Richmond M. Rate of growth of *Bacillus cereus* between divisions. *Microbiology* (1962) **28**:15–33.

19. Powell E. A note on Koch & Schaechter's hypothesis about growth and fission of bacteria. *Microbiology* (1964) **37**:231–49.

20. Tyson JJ, Diekmann O. Sloppy size control of the cell division cycle. *J Theor Biol.* (1986) **118**:405–26.

21. Metz JA, Diekmann O. *The Dynamics of Physiologically Structured Populations, Vol. 68*. Berlin: Springer (1986).

22. Rading M, Engel T, Lipowsky R, Valleriani A. Stationary size distributions of growing cells with binary and multiple cell division. *J Stat Phys.* (2011) **145**:1–22. doi: 10.1007/s10955-011-0305-9

23. Robert L, Hoffmann M, Krell N, Aymerich S, Robert J, Doumic M. Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism. *BMC Biol.* (2014) **12**:17. doi: 10.1186/1741-7007-12-17

24. García MR, Vázquez JA, Teixeira IG, Alonso AA. Stochastic Individual-Based Modeling of Bacterial Growth and Division Using Flow Cytometry. *Front Microbiol.* (2018) **8**:2626. doi: 10.3389/fmicb.2017.02626

25. Marantan A, Amir A. Stochastic modeling of cell growth with symmetric or asymmetric division. *Phys Rev E.* (2016) **94**:012405. doi: 10.1103/PhysRevE.94.012405

26. Vargas-Garcia CA, Soltani M, Singh A. Conditions for cell size homeostasis: a stochastic hybrid system approach. *IEEE Life Sci Lett.* (2016) **2**:47–50. doi: 10.1109/LLS.2016.2646383

27. Modi S, Vargas-Garcia CA, Ghusinga KR, Singh A. Analysis of noise mechanisms in cell-size control. *Biophys J.* (2017) **112**:2408–18. doi: 10.1016/j.bpj.2017.04.050

28. Tanouchi Y, Pai A, Park H, Huang S, Stamatov R, Buchler NE, et al. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* (2015) **523**:357–60. doi: 10.1038/nature14562

29. Nozoe T, Kussell E, Wakamoto Y. Inferring fitness landscapes and selection on phenotypic states from single-cell genealogical data. *PLoS Genet.* (2017) **13**:e1006653. doi: 10.1371/journal.pgen.1006653

30. Wakamoto Y, Grosberg AY, Kussell E. Optimal lineage principle for age-structured populations. *Evolution* (2012) **66**:115–34. doi: 10.1111/j.1558-5646.2011.01418.x

31. Thomas P. Single-cell histories in growing populations: relating physiological variability to population growth. *bioRxiv* [Preprint] (2017). doi: 10.1101/100495

32. Bell GI, Anderson EC. Cell growth and division: I. A mathematical model with applications to cell volume distributions in mammalian suspension cultures. *Biophys J.* (1967) 7:329–51.

33. Farley MM, Hu B, Margolin W, Liu J. Minicells, back in fashion. *J Bacteriol.* (2016) **198**:1186–95. doi: 10.1128/JB.00901-15

34. Martins BMC, Tooke AK, Thomas P, Locke JCW. Cell size control driven by the circadian clock and environment in cyanobacteria. *bioRxiv* [Preprint] (2017). doi: 10.1101/183558

35. Thomas P. Making sense of snapshot data: ergodic principle for clonal cell populations. *J R Soc Interface* (2017) **14**:20170467. doi: 10.1098/rsif.2017.0467