To bin or not to bin: analyzing single-cell growth data

Prathitha Kar[1,2], Sriram Tiruvadi-Krishnan[3], Jaana Männik[3], Jaan Männik*[3], and Ariel Amir†[1]

[1]*School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02134, USA*

[2]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA*

[3]*Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996, USA*

*Corresponding author- email: jmannik@utk.edu; phone: +1 (865) 974 6018

†Corresponding author- email: arielamir@seas.harvard.edu; phone: +1 (617) 495 5818

Abstract

Collection of high-throughput data has become prevalent in biology. Large datasets allow the use of statistical constructs such as binning and linear regression to quantify relationships between variables and hypothesize underlying biological mechanisms based on it. We discuss several such examples in relation to single-cell data and cellular growth. In particular, we show instances where what appears to be ordinary use of these statistical methods leads to incorrect conclusions such as growth being non-exponential as opposed to exponential and vice versa. We propose that the data analysis and its interpretation should be done in the context of a generative model, if possible. In this way, the statistical methods can be validated either analytically or against synthetic data generated via the use of the model, leading to a consistent method for inferring biological mechanisms from data. On applying the validated methods of data analysis to infer cellular growth on our experimental data, we find the growth of length in *E. coli* to be non-exponential. Our analysis shows that in the later stages of the cell cycle the growth rate is faster than exponential.

1 Introduction

The last decade has seen a tremendous increase in the availability of high-quality large datasets in biology, in particular in the context of single-cell level measurements. Such data are complementary to "bulk" measurements made over a population of cells. They have led to new biological paradigms and motivated the development of quantitative models [1–7]. Nevertheless, they have also led to new challenges in data analysis, and here we will point out some of the pitfalls that exist in handling such data. In particular, we will show that the commonly used procedure of binning data in order to eliminate noise (averaging conditioned on the value of one of the variables) may lead to smooth curves that hint at specific functional relations between the two variables plotted that are inconsistent with the true functional relations. As we shall show, this may come about due to the "hidden" noise sources that affect the binning procedure and the phenomenon of "inspection bias" where certain bins have biased contributions. One of our main take home messages is the significance of having an underlying model (or models) to guide/test/validate data analysis methods. The underlying model is referred to as a generative model in the sense that it leads to similar data to that observed in the experiments. The importance of a so-called generative model has been beautifully advocated in the context of astrophysical data analysis [8], yet biology brings in a plethora of exciting differences: while in physics noise from measurement instruments often dominates, in the biological examples we will dwell on here it is the *intrinsic* biological noise that can obscure the mathematical relation between variables when not handled properly. In the following, we will illustrate this rather philosophical introduction on a concrete and fundamental example, albeit e pluribus unum. We will focus on the analysis of the *Escherichia coli* growth curves obtained via high throughput optical microscopy. Nevertheless we anticipate the conceptual points made here – and demonstrated on a particular example of interest – will translate to other types of measurements, which

make use of microscopy but also beyond.

Binning corresponds to grouping data based on the value of the x-axis variable, and finding the mean of the fluctuating y-axis variable for this group. By removing the fluctuations of the y-variable, the binning process often aims to expose the "true" functional relation between the two variables which can be used to infer the underlying biological mechanism. It is important to discuss the sources of fluctuations in the y-axis variable before we proceed. In biology, fluctuations in the variables arise inevitably from the intrinsic variability within a cell population. Cells growing in the same medium and environment have different characteristics (e.g., growth rate) due to the stochastic nature of biochemical reactions in the cell [9]. For example, the division event is controlled by stochastic reactions, whose variability leads to cell dividing at a size smaller or larger than the mean. In this paper, when modeling the data, we will consider the intrinsic noise as the only source of variability and assume that the measurement error is much smaller than the intrinsic variation in the population.

One example of the use of binning is shown in Figure 1A where size at division (L_d) vs size at birth (L_b) is plotted using experimental data obtained by Tanouchi *et al.* for *E. coli* growing at 25°C [10]. In Figure 1A, the functional relation between length at division and length at birth for *E. coli* is observed to be linear and close to $L_d = L_b + \Delta L$ (see Section 5.11.1 for details). The relation obtained allows us to hypothesize a coarse-grained biological model known as the adder model as shown in Figure 1B in which the length at division is set by addition of length ΔL from birth [4, 11–16]. This example demonstrates the use of statistical analysis on single-cell data to understand the underlying cell regulation mechanisms. Using statistical methods such as binning and linear regression, other phenomenological models apart from adder have also been proposed in *E. coli* where the division length (L_d) is not directly "set" by that at birth [17–19]. The phenomenological models, in turn, can be related to mechanistic (molecular-level) models of cell size and cell cycle regulation [20]. Recent work has shed light on the subtleties involved in interpreting the linear regression results for

the L_d vs L_b plot where seemingly adder behavior in length can be obtained from a sizer model (division occurring on reaching a critical size) due to the interplay of multiple sources of variability [21]. This issue is similar in spirit to those we highlight here.

The volume growth of single bacterial cells has been typically assumed to be exponential [4, 14, 22–25]. Assuming ribosomes to be the limiting component in translation, growth is predicted to be exponential and growth rate depends on the active ribosome content in the cell [26–28]. Under the assumption of exponential growth, the size at birth (L_b), the size at division (L_d), and the generation time (T_d) are related to each other by,

$$\ln(\frac{L_d}{L_b}) = \lambda T_d, \tag{1}$$

where λ is the growth rate. Understanding the mode of growth is important e.g., due to its potential effects on cell size homeostasis. Exponentially growing cells cannot employ a mechanism where they control division by timing a constant duration from birth but such a mechanism is possible in case of linear growth [3, 13, 29]. Linear regression performed on $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot, where $\langle\lambda\rangle$ is the mean growth rate, was used to infer the mode of growth in the archaeon *H. salinarum* [16], and in the bacteria *M. smegmatis* [30] and *C. glutamicum* [31], for example. If the best linear fit follows the y=x trend, the resulting functional relation might point to growth being exponential. A corollary to this is the rejection of exponential growth when the slope and intercept of the best linear fit deviate from one and zero respectively [31]. Thus, binning and linear regression applied on single-cell data appear to provide information about the underlying biology, in this case, the mode of cellular growth. We will test the validity of such inference by analyzing synthetic data generated using generative models. We find that linear regression performed on the plot $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$, surprisingly, does not provide information about the mode of growth. Nonetheless, we show that other methods of statistical analysis such as binning growth rate vs age plots

are adequate in addressing the problem. Using these validated methods on experimental data, we find that *E. coli* grows non-exponentially. In later stages of the cell cycle, the growth rate is higher than that in early stages.

2 Statistical methods like binning and linear regression should be interpreted based on a model.

To illustrate the pitfalls associated with binning, we use data from recent experiments on *E. coli* where the length at birth, the length at division and the generation time were obtained for multiple cells (see Section 5.1 and [32]). Phase-contrast microscopy was used to obtain cell length at equal intervals of time. Note that we consider length as a proxy for cell size as the fluctuations in the width of *E. coli* cells are negligible in a given condition [15, 23, 33, 34]. To investigate if the cell growth was exponential, we plotted $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ for cells growing in M9 alanine minimal medium at 28°C ($\langle T_d\rangle = 214$ min). The linear regression of these data yields a slope of 0.3 and an intercept of 0.4 as shown in Figure 2A. The binned data and the best linear fit deviate significantly from the y=x line (see Table S2). Additionally, the binned data follows a non-linear trend and flattens out at longer generation times. We also found similar deviations in the binned data and best linear fit in glycerol medium ($\langle T_d\rangle =$ 164 min) shown in Figure 2- figure supplement 1A, and glucose-cas medium ($\langle T_d\rangle = 65$ min) shown in Figure 2- figure supplement 1B. Qualitatively similar results have been recently obtained for another bacterium, *C. glutamicum*, in Ref. [31]. These results might point to growth being non-exponential.

Next we will approach the same problem but with a generative model. We will first show that the aforementioned non-linear dependencies are perfectly consistent with purely exponential growth. For the model, we consider exponential growth where the growth rate is distributed normally and independently between cell cycles with mean growth rate $\langle\lambda\rangle$

and standard deviation $CV_\lambda\langle\lambda\rangle$. CV_λ is thus the coefficient of variation (CV) of the growth rate and is assumed to be small. To maintain a narrow distribution of cell size, cells must employ regulatory mechanisms. In our model, we assume that, barring the noise due to stochastic biochemical reactions, cells attempt to divide at a particular size L_d given size at birth L_b. Keeping the model as generic as possible, we can write L_d as a function of L_b, $\mathrm{f}(L_b)$ which can be thought of as a coarse-grained model for the regulatory mechanism. Ref. [13] provides a framework to capture the regulatory mechanisms by choosing $\mathrm{f}(L_b) = 2L_b^{1-\alpha}L_0^{\alpha}$. L_0 is the typical size at birth and α, which can take values between 0 and 2, reflects the strength of regulation strategy. $\alpha = 0$ corresponds to the timer model where division occurs on average after a constant time from birth, and $\alpha = 1$ is the sizer model where a cell divides upon reaching a critical size. $\alpha = 1/2$ can be shown to be equivalent to the adder model where division is controlled by addition of constant size from birth [13]. In addition to the deterministic function (f) specifying division, the size at division is affected by noise ($\frac{\zeta}{\langle\lambda\rangle}$) in division timing. We assume it has a Gaussian distribution with mean zero and standard deviation $\frac{\sigma_n}{\langle\lambda\rangle}$ and that it is independent of the growth rate. Thus, the generation time (T_d) can be mathematically written as $T_d = \frac{1}{\lambda}\ln(\frac{\mathrm{f}(L_b)}{L_b}) + \frac{\zeta}{\langle\lambda\rangle}$ and is influenced by growth rate noise and division timing noise. Note that replacing the time additive division timing noise with a size additive division timing noise will not affect the results qualitatively (see Sections 5.2 and 5.3 for details and Table S1 for variable definitions).

For perfectly symmetrically dividing cells whose sizes are narrowly distributed, the trend in the binned data for $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot is found to be (see Section 5.4),

$$y = x\left(1 + \frac{1 - \frac{x}{\ln(2)}}{1 + \frac{2}{2-\alpha}\frac{\sigma_n^2}{CV_\lambda^2\ln^2(2)}}\right). \tag{2}$$

Fixing $CV_\lambda = \sigma_n = 0.15$, we show using simulations in Figure 2C the non-linear trend in the binned data even though we assumed exponential growth. Similarly, on performing linear

regression on $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot, we find that the slope is not equal to one and the intercept is non-zero (see Eqs. 27 and 28). Eq. 2 shows that the trend in the binned data depends on the ratio of growth rate noise and division timing noise. The slope is equal to one and intercept is zero only if the noise in growth rate is negligible. In experiments that is rarely the case, hence, the binned data trend and the best linear fit deviate from the y=x line even though growth might be exponential. Thus, we cannot rule out exponential growth in the *E. coli* experiments despite the binned data trend being non-linear and the best-fit line deviating from the y=x line.

Why does a non-linear relationship in the binned data for the plot $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ arise even for exponential growth? According to the model, L_d is determined by a deterministic strategy, f(L_b) and a time/size additive division timing noise. The noise component which affects L_d and subsequently the quantity $\ln(\frac{L_d}{L_b})$ is thus the noise in division timing and not the growth rate. The generation time (T_d) plotted on the x-axis is influenced by the noise in division timing as well as the noise in growth rate. Binning assumes that for a fixed value of the x-axis variable, the noise from other sources affects only the y-axis variable (the binned variable). Similarly for linear regression, the underlying assumption is that the independent variable on x-axis is precisely known while the dependent variable on the y-axis is influenced by the independent variable and from external factors other than the independent variable. In this case, only $\langle\lambda\rangle T_d$ plotted on x-axis is influenced by growth rate noise while both $\langle\lambda\rangle T_d$ and $\ln(\frac{L_d}{L_b})$ are influenced by noise in division time. This does not fit the assumption for binning and linear regression and hence, the best linear fit for $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot might deviate from the y=x line even in the case of exponential growth.

Another way of explaining the deviation from the linear y=x trend is by inspection bias, which arises when certain data is over-represented [35]. Cells which have a longer generation time than the mean will most likely have a slower growth rate. Thus, in Figure 2A and Figure 2C, at larger values of $\langle\lambda\rangle T_d$ or T_d, the bin averages are biased by slower growing

cells, thus making $\ln(\frac{L_d}{L_b})$ or λT_d to be lower than expected. This provides an explanation for the flattening of the trend.

It follows from the previous discussion that if one bins data by $\ln(\frac{L_d}{L_b})$ then the assumption for binning is met. Both of the variables $\langle\lambda\rangle T_d$ and $\ln(\frac{L_d}{L_b})$ are influenced by the noise in division time but $\langle\lambda\rangle T_d$ plotted on the y-axis is also influenced by the growth rate noise. Thus, the y-axis variable, $\langle\lambda\rangle T_d$ is determined by the x-axis variable, $\ln(\frac{L_d}{L_b})$, and an external source of noise, in this case, the growth rate noise. Thus, based on our model, we expect the trend in binned data and linear regression performed on the interchanged axes to follow the y=x trend for exponentially growing cells (see Section 5.4). Indeed, on interchanging the axis and plotting $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ for synthetic data, we find that the trend in the binned data follows the y=x line (Figure 2D). We also find that the best linear fit follows the y=x line in the case of alanine (Figure 2B), glycerol (Figure 2- figure supplement 1A) and glucose-cas (Figure 2- figure supplement 1B). A change from non-linear behavior to that of linear on interchanging the axes is also observed in a related problem where growth rate (λ) and inverse generation time ($\frac{1}{T_d}$) are considered (Figure 2- figure supplement 2 and Section 5.10).

Thus far, we showed for a range of models where birth controls division that the binned data trend for $\ln(\frac{L_d}{L_b})$ as function of $\langle\lambda\rangle T_d$ is non-linear and dependent on the noise ratio $\frac{\sigma_n}{CV_\lambda}$ in the case of exponential growth. On interchanging the axes the binned data trend agrees with the y=x line independent of the growth rate and division time noise. However, we will show next that this agreement with the y=x trend cannot be used as a "smoking gun" for inferring exponential growth from the data. To investigate this further, let us consider linear growth, which has also been suggested to be followed by *E. coli* cells [36, 37]. The underlying equation for linear growth is,

$$L_d - L_b = \lambda' T_d, \tag{3}$$

where λ' is the the elongation speed i.e., $\frac{dL}{dt}$. For cells growing linearly, the best linear

fit for the plot $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ is expected to deviate from the y=x line. Surprisingly, we found that for the class of models where birth controls division by a strategy f(L_b) and cells grow linearly, the best linear fit for $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ agrees closely with the y=x trend. On carrying out analytical calculations based on this model, we obtain the slope and the intercept of the $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot to be $\frac{3}{2}\ln(2) \approx 1.04$ and -0.03 respectively, which is very close to that for exponential growth (see Section 5.6). This is shown for simulations of linear growth with cells following an adder model in Figure 3A. Given no information about the underlying model, Figure 3A could be interpreted as cells undergoing exponential growth contrary to the assumption of linear growth in simulations. Thus, when handling experimental data, cells undergoing either exponential or linear growth might seem to agree closely with the y=x trend. Deforet *et al.* [38] used the linear binned data trend in case of $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot to infer exponential growth but as we showed in this section, the linear trend does not rule out linear growth. This again reiterates our message of having a generative model to guide the data analysis methods such as binning and linear regression. For completeness, we also discuss the natural plot for linear growth, $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ and the plot obtained on interchanging the axes in Section 5.5 and Figure 3- figure supplements 1A, 1B. For cells growing exponentially, the best linear fit for the $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot is expected to deviate from the y=x line. This is indeed what is observed in Figure 3- figure supplement 1C where simulations of exponentially growing cells following the adder model are presented (see Section 5.6 for extended discussion).

In all of the cases above, the problem at hand deals with distilling the biologically relevant functional relation between two variables. However, the data is assumed to be subjected to fluctuations of various sources, and it is important to ensure that the statistical construct we are using (e.g. binning) is robust to these. How can we know a priori whether the statistical method is appropriate and a "smoking gun" for the functional relation we are conjecturing? The examples shown above suggest that performing statistical tests on synthetic data ob-

tained using a generative model is a convenient and powerful approach. Note that in cases such as the ones studied here where analytical calculations may be performed, one may not even need to perform any numerical simulations to test the validity of the methods.

3 Growth rate vs age plots are consistent with the underlying growth mode.

In the last section, we showed that the plots $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ and $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ are not decisive in identifying the mode of growth. Recent works on *B. subtilis* [39] and fission yeast [40] have used differential methods of quantifying growth namely growth rate ($= \frac{1}{L}\frac{dL}{dt}$) vs age plots and elongation speed ($=\frac{dL}{dt}$) vs age plots to probe the mode of growth within a cell cycle. Here, L denotes the size of the cell after time t from birth in the cell cycle and age denotes the ratio of time t to T_d within a cell cycle (hence it ranges from 0 to 1 by construction within a cell cycle). In this section, using various models of cell growth and cell cycle, we test the growth rate vs age method. For cells assumed to be growing exponentially, growth rate is constant throughout the cell cycle. On averaging over multiple cell cycles, the trend of binned data is expected to be a horizontal line with value equal to mean growth rate which is indeed what we find in the numerical simulations of the adder and the adder per origin model [17], as shown in Figure 3B. In contrast, for linearly growing cells, the elongation speed is expected to remain constant. We show this constancy using numerical simulations of linearly growing cells following the adder model (Figure 3- figure supplement 3A). In accordance with this result, the growth rate is expected to decrease with cell age for linear growth. This is verified in Figure 3B by again using the numerical simulations of linear growth with cells following the adder model. Thus, the two growth modes (exponential and linear) could be differentiated using the growth rate vs age plot and it appears to be a consistent method to obtain the mode of growth. For further details about the binning

method used in growth rate vs age and elongation speed vs age plots, see Section 5.7.

Using the validated growth rate vs age plots, we obtained the growth rate trend for experimental data on *E. coli* for the three growth conditions studied in this paper (Figures 4A-4C). We found an increase in growth rate in all growth conditions during the course of the cell cycle. One may wonder whether such an increase may be explained by the *E. coli* morphology alone, due to the presence of hemispherical poles. For exponentially growing cell volume and considering a geometry of *E. coli* with spherical caps at the poles, the percentage increase in the growth rate of length over a cell cycle is around 3% which is significantly smaller than that observed in our experimental data. Considering cell size trajectories (cell size, L at time, t data) where cell lengths were tracked beyond the cell division event (by considering cell size in both daughter cells), we also found that the growth rate decreases close to division (age ≈ 1) and returns to a value nearly equal to that observed at the beginning of cell cycle (age ≈ 0) as shown in Figure 4- figure supplements 1A-1C (see Section 5.7 for extended discussion).

The above question of mode of growth within a cell cycle can also be analyzed in relation to a specific event. Several studies have pointed to a change in growth rate at the onset of constriction [41, 42]. This change in growth rate can be probed using growth rate vs time plots where time is taken relative to the onset of constriction as shown in Figure 4- figure supplement 2. These plots show a decrease in growth rates at the two extremes of the plot. These decreases are due to inspection bias, where the growth rate trend is affected by the biased contribution of cells with a higher than average generation time or equivalently slower growth rate (see Section 5.8 for extended discussion). Inspection bias is also observed when timing is considered relative to other cell events such as cell birth (see Section 5.8 and Figure 3- figure supplements 2C, 2D).

It might not always be possible to obtain growth rate trajectories as a function of time/cell age. Godin *et al.* instead obtained the instantaneous biomass growth speed ($\frac{dM}{dt}$) as a func-

tion of its buoyant mass (M) [22]. On applying linear regression for instantaneous mass growth speed vs mass, we expect the slope of the best linear fit obtained to provide the average growth rate ($\langle \frac{1}{M}\frac{dM}{dt} \rangle$) under the assumption of exponential growth while for linear growth the intercept provides the average growth speed. Using this method, biomass was suggested to be growing exponentially. This method can be applied to study the length growth rate within the cell cycle by plotting elongation speed as a function of length [43]. We find that the binned data trend of this plot follows the expected trend for linear and exponential growth as shown in Figure 3- figure supplement 3B and Figure 3- figure supplement 3D, respectively, for a cell cycle model where division is controlled via an adder mechanism from birth. However, the trend obtained appears to be model-dependent as shown in Figure 3- figure supplement 3F where the underlying cell cycle model used in the simulations is the adder per origin model. For this model, the binned data trend is found to be non-linear with the growth rate speeding up at large sizes, despite the synthetic data being generated for perfectly exponential growth. This non-linear trend can lead to growth rate being misinterpreted as non-exponential within the cell cycle (see Section 5.9 for details). Thus, an analysis using the elongation speed vs size plot must be accompanied with an underlying cell cycle model.

In summary, we found that the growth rate vs age plot was a consistent method to determine the changes in growth rate within a cell cycle. Unlike the growth rate vs age plots, the inference from the growth rate vs size plots was found to be model-dependent. Using the growth rate vs age plots, we show that the length growth of *E. coli* can be faster than exponential (super-exponential).

4 Discussion

Statistical methods such as binning and linear regression are useful for interpreting data and generating hypotheses for biological models. However, we show in this paper that predicting the relationships between experimentally measured quantities based on these methods might lead to misinterpretations. Constructing a generic model and verifying the statistical analysis on the synthetic data generated by this model provides a more rigorous way to mitigate these risks.

In the paper, we provide examples in which $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ and $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plots fail as a method to infer the mode of growth. The binned data trend for $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot was found to be dependent upon the noise parameters in the class of models where birth controlled division (Equation 2). We also show that $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot could not differentiate between exponential and linear modes of growth (Figures 2D, 3A). Thus, we conclude that the best linear fit for the above plots might not be a suitable method to infer the mode of growth but they are just one of the many correlations which the correct cell cycle model should be able to predict.

We found growth rate vs age and elongation speed vs age plots to be consistent methods to probe growth within a cell cycle. The method was validated using simulations of various cell cycle models (such as the adder, and adder per origin model, where in the latter, control over division is coupled to DNA replication) and the binned growth rate trend agreed closely with the underlying mode of growth for the wide range of models considered (Figure 3B). In the case of growth rate vs time plots, it was important to take into consideration the effects of inspection bias. We used cell cycle models to show the time regimes where inspection bias could be observed (Figure 3- figure supplement 2). In the regime with negligible inspection bias, we could reconcile the growth rate trend obtained using growth rate vs age (Figures 4A-4C) and growth rate vs time plots (Figure 4- figure supplement 2). The authors in Ref. [31]

circumvent inspection bias in the elongation speed vs time from birth plots by focusing their analysis on the time period from cell birth to the generation time of the fastest dividing cell. The authors of Ref. [44], while investigating the division behavior in the cells undergoing nutrient shift within their cell cycle, use both models and experimental data from steady-state conditions to identify inspection bias. These serve as good examples of using models to aid data analysis.

Statistics obtained from linear regression such as in Figure 1A help narrow down the landscape of cell cycle models, but many have potential pitfalls lurking which might lead to misinterpretations (Figure 2C, Figure 3A). There are additional issues beyond those concerning linear regression and binning discussed here. For example, Ref. [45] discusses Simpson's paradox [46] where distinct cellular sub-populations might lead to erroneous interpretation of cell cycle mechanisms. Examples of such distinct sub-populations are found in asymmetrically dividing bacteria such as *M. smegmatis* [30, 47].

Single cell size in *E. coli* has been reported to grow exponentially [4, 14, 22–25], linearly [36], bilinearly [48] or trilinearly [41]. These are inconsistent with our observations in Figures 4A-4C where we find that growth can be super-exponential. The non-monotonic behavior in the fastest-growth condition is reminiscent of the results reported in Ref. [39] for *B. subtilis*. The authors of Ref. [39] attribute the increase in growth rate to a multitude of cell cycle processes such as initiation of DNA replication, divisome assembly, septum formation. In the two slower growth conditions (Figures 4A-4B), we find that the growth rate increase starts before the time when the septal cell wall synthesis starts i.e., the constriction event. However, in the fastest growth condition (Figure 4C), the timing of growth rate increase seems to coincide with the onset of constriction which is in agreement with previous findings [41, 42].

It is important to distinguish between length growth and biomass growth. Ref. [49] measures biomass and cell volume and finds the mass-density variations within the cell-cycle

to be small. In this paper, since we observe the length growth to be non-exponential (Figure 4), it remains to be seen whether biomass growth also follows a similar non-exponential behavior or if it is exponential as previously suggested [22, 49].

In conclusion, the paper draws the attention of the readers to the careful use of statistical methods such as linear regression and binning. Although shown in relation to cell growth, this approach to data analysis seems ubiquitous. The general framework of carrying out data analysis is presented in Figure 5. It proposes the construction of a generative model based on the experimental data collected. Of course, we do not always know whether the model used is an adequate description of the system. What is the fate of the methodology described here in such cases? First, we should be reminded of Box's famous quote "all models are wrong, some are useful". The goal of a model is not to provide as accurate a description of a system as possible, but rather to capture the essence of the phenomena we are interested in and stimulate further ideas and understanding. In our context, the goal of the model is to provide a rigorous framework in which data analysis tools can be critically tested. If verified within the model, it is by no means proof of the success of the model and the method itself, and further comparisons with the data may falsify it leading to the usual (and productive) cycle of model rejection and improvement via comparison with experiments. However, if the best model we have at hand shows that the data analysis method is non-informative, as we have shown here on several methods used to identify the mode of growth, then clearly we should revise the analysis as it provides us with a non-consistent framework, where our modeling is at odds with our data analysis. Furthermore, testing the methods on a simplified model is still advantageous compared with the option of using the methods without any validation. To mitigate the risk of using irrelevant models, in some cases it may be desirable to test the analysis methods on as broad a class of models as possible as we have done in the paper, for example by our use of a general value of α to describe the size-control strategy within our models. Thus, guided by the model, the data analysis methods can be ultimately applied to

experimental data and underlying functional relationships can be inferred. Reiterating the message of the authors in Ref. [8], the data analysis using this framework aims to justify the methods being used, thus, reducing arbitrariness and promoting consensus among the scientists working in the field.

5 Methods

5.1 Experimental methods

Strain engineering: STK13 strain (ΔftsN::frt-Ypet-FtsN, ΔdnaN::frt-mCherry-dnaN) is derivative of *E. coli* K12 BW27783 (CGSC#: 12119) constructed by λ-Red engineering [50] and by P1 transduction [51]. For chromosomal replacement of ftsN with fluorescence derivative, we used primers carrying 40nt tails with identical sequence to the *ftsN* chromosomal locus and a plasmid carrying a copy of *ypet* preceded by a kanamycin resistance cassette flanked by *frt* sites (frt-kan^R-*frt-Ypet-linker*) as PCR template (a kind gift from R. Reyes-Lamothe McGill University, Canada; [52]). The resulting PCR product was transformed by electroporation into a strain carrying the λ-Red-expressing plasmid pKD46. Colonies were selected by kanamycin resistance, verified by fluorescence microscopy and by PCR using primers annealing to regions flanking ftsN gene. After removal of kanamycin resistance by expressing the Flp recombinase from plasmid pCP20 [53], we transferred the mCherry-dnaN gene fusion (BN1682 strain; a kind gift from Nynke Dekker from TUDelft, The Netherlands, [54]) into the strain by P1 transduction. To minimize the effect of the insertion on the expression levels of the gene we removed the kanamycin cassette using Flp recombinase expressing plasmid pCP20.

Cells growth, preparation, and culturing *E. coli* in mother machine microfluidic devices: All cells were grown and imaged in M9 minimal medium (Teknova) supplemented with 2 mM magnesium sulfate (Sigma) and corresponding carbon sources at 28°C.

Three different carbon sources were used: 0.5% glucose supplemented by 0.2% casamino acids (Cas) (Sigma), 0.3% glycerol (Fisher) and 0.3% alanine (Fisher) supplemented with 1x trace elements (Teknova).

For microscopy, we used mother machine microfluidic devices made of PDMS (polydimethylsiloxane). These were fabricated following to previously described procedure [55]. To grow and image cells in microfluidic device, we pipetted 2-3 μl of resuspended concentrated overnight culture of $OD_{600} \sim 0.1$ into main flow channel of the device and let cells to populate the dead-end channels. Once these channels were sufficiently populated (about 1 hr), tubing was connected to the device, and the flow of fresh M9 medium with BSA (0.75 μg/ml) was started. The flow was maintained at 5 μl/min during the entire experiment by an NE-1000 Syringe Pump (New Era Pump Systems, NY). To ensure steady-state growth, the cells were left to grow in channels for at least 14 hr before imaging started.

Microscopy: A Nikon Ti-E inverted epifluorescence microscope (Nikon Instruments, Japan) with a 100X (NA = 1.45) oil immersion phase contrast objective (Nikon Instruments, Japan), was used for imaging the bacteria. Images were captured on an iXon DU897 EMCCD camera (Andor Technology, Ireland) and recorded using NIS-Elements software (Nikon Instruments, Japan). Fluorophores were excited by a 200W Hg lamp through an ND8 neutral density filter. A Chroma 41004 filtercube was used for capturing mCherry images, and a Chroma 41001 (Chroma Technology Corp., VT) for Ypet images. A motorized stage and a perfect focus system were utilized throughout time-lapse imaging. Images in all growth conditions were obtained at 4 min frame rate.

Image analysis: Image analysis was carried out using Matlab (MathWorks, MA) scripts based on Matlab Image Analysis Toolbox, Optimization Toolbox, and DipImage Toolbox (https://www.diplib.org/). Cell lengths were determined based on segmented phase contrast images. Dissociation of Ypet-FtsN label from cell middle was used to determine the exact timing of cell divisions.

Further experimental details can also be found in Ref. [32].

5.2 Model

Consider a model of cell cycle characterized by two events: cell birth and division. In our model, we assume that, barring the noise, cells tend to divide at a particular size v_d given size at birth v_b, via some regulatory mechanism. Hence, we can write v_d as a function of v_b, f(v_b). Ref. [13] provides a framework to capture the regulatory mechanisms by choosing f(v_b) = $2v_b^{1-\alpha}v_0^{\alpha}$. v_0 is the typical size at birth and α captures the strength of regulation strategy. $\alpha = 0$ corresponds to the timer model where division occurs after a constant time from birth, and $\alpha = 1$ is the sizer where a cell divides on reaching a critical size. $\alpha = 1/2$ can be shown to be equivalent to an adder where division is controlled by addition of constant size from birth [13]. From here on, we would be using the length of the cell (L_b, L_d, etc.) as a proxy for size (v_b, v_d, etc.). All of the variable definitions are summarized in Table S1. We also define $l_b = \frac{L_b}{\langle L_b \rangle}$ and $l_d = \frac{L_d}{\langle L_b \rangle}$. Using this, we can write the division strategy f(l_b) to be l_d = f(l_b) = 2 $l_b^{1-\alpha}$. The total division size obtained will be a combination of f(l_b) and noise in the division timing, the source of which could be the stochasticity in biochemical reactions controlling division.

We will assume that division is perfectly symmetric i.e., size at birth in the $(n+1)^{th}$ generation (l_b^{n+1}) is half of size at division in the n^{th} generation (l_d^n). Using the size additive division timing noise ($\zeta_s(0, \sigma_{bd})$) and f(l_b) specified above, we obtain,

$$x_{n+1} = (1-\alpha)x_n + \ln\left(1 + \frac{\zeta_s(0,\sigma_{bd})}{2(1+x_n)^{1-\alpha}}\right), \tag{4}$$

where $x_n = \ln(l_b^n)$. Size at birth (L_b) is narrowly distributed, hence $l_b \approx 1$ and we can write

$x = \ln(l_b) = \ln(1 + \delta)$ where δ is a small number. We obtain $x \ll 1$ and,

$$x \approx \delta = l_b - 1. \tag{5}$$

The size additive noise, $\zeta_s(0, \sigma_{bd})$ is assumed to be small and has a normal distribution with mean 0 and standard deviation σ_{bd}. Note that σ_{bd} is a dimensionless quantity. Since $\zeta_s(0, \sigma_{bd})$ is assumed to be small and $x_n \ll 1$, we can Taylor expand the last term of Equation 4 to first order,

$$x_{n+1} \approx (1 - \alpha)x_n + \frac{\zeta_s(0, \sigma_{bd})}{2}. \tag{6}$$

Equation 6 shows a recursive relation for cell size and it is agnostic of the mode of growth. We will show later for exponential growth that replacing the size additive noise with time additive noise does not change the structure of Equation 6.

5.3 Exponential growth

Next, we will try to obtain the generation time (T_d) in the case of exponentially growing cells. For exponential growth, the time at division T_d is given by,

$$T_d = \frac{1}{\lambda} \ln(\frac{L_d}{L_b}). \tag{7}$$

For simplicity, we will assume a constant growth rate (λ) within the cell-cycle. Growth rate is fixed at the start of the cell-cycle and is given by $\lambda = \langle\lambda\rangle + \langle\lambda\rangle\xi(0, CV_\lambda)$, where $\langle\lambda\rangle$ is the mean growth rate and $\xi(0, CV_\lambda)$ is assumed to be small with a normal distribution that has mean 0 and standard deviation CV_λ. CV_λ denotes the coefficient of variation (CV) of the growth rate. This captures the variability in growth rate within cells arising from the stochastic nature of biochemical reactions occurring within the cell.

5.3.1 Size additive noise

Here we will calculate the generation time using the division strategy f(l_b) and a size additive division timing noise ($\zeta_s(0, \sigma_{bd})$) as described previously. On substituting $L_d = (f(l_b) + \zeta_s)\langle L_b \rangle$ into Equation 7 we obtain,

$$T_d = \frac{1}{\langle\lambda\rangle + \langle\lambda\rangle\xi(0, CV_\lambda)} \ln(\frac{2l_b^{1-\alpha} + \zeta_s(0, \sigma_{bd})}{l_b}), \tag{8}$$

where the size additive noise ($\zeta_s(0, \sigma_{bd})$) is Gaussian with mean 0 and standard deviation σ_{bd}.

The noise $\zeta_s(0, \sigma_{bd})$ is assumed to be small, and we obtain to first order,

$$T_d \approx \frac{1}{\lambda}\left(\ln(2) - \alpha x_n + \frac{\zeta_s(0, \sigma_{bd})}{2(1 + x_n)^{1-\alpha}}\right). \tag{9}$$

Since $x_n \ll 0$, on Taylor expanding $\frac{1}{(1+x_n)^{1-\alpha}}$ to first order,

$$T_d \approx \frac{1}{\lambda}\left(\ln(2) - \alpha x_n + \frac{\zeta_s(0, \sigma_{bd})}{2}(1 + (1 - \alpha)x_n)\right). \tag{10}$$

Assuming noise in growth rate to be small and expanding to first order, we obtain,

$$T_d \approx \frac{1}{\langle\lambda\rangle}\left(\ln(2) - \alpha x_n - \ln(2)\xi(0, CV_\lambda) + \frac{\zeta_s(0, \sigma_{bd})}{2}\right). \tag{11}$$

Equation 11 gives the generation time for the class of models where birth controls division under the assumption that growth is exponential.

5.3.2 Time additive noise

Next, we ensure that the recursive relation for size at birth and the expression for the generation time given by Equations 6 and 11, respectively, are robust to the nature of noise

assumed. In this section, the generation time is obtained using the division strategy f(l_b) as described previously along with a time additive division timing noise ($\frac{\zeta}{\langle\lambda\rangle}$). In such a case, T_d is obtained to be,

$$T_d = \frac{1}{\lambda}(\ln(2) - \alpha x_n) + \frac{\zeta(0,\sigma_n)}{\langle\lambda\rangle}. \tag{12}$$

The time additive noise, $\frac{\zeta(0,\sigma_n)}{\langle\lambda\rangle}$, is assumed to be small and has a normal distribution with mean 0 and standard deviation $\frac{\sigma_n}{\langle\lambda\rangle}$. Note that σ_n is a dimensionless quantity.

Assuming noise in growth rate to be small, we find T_d to first order to be,

$$T_d \approx \frac{1}{\langle\lambda\rangle}\left(\ln(2) - \alpha x_n - \ln(2)\xi(0, CV_\lambda) + \zeta(0,\sigma_n)\right). \tag{13}$$

Equation 13 is same as Equation 11, if the time additive noise term, $\zeta(0,\sigma_n)$, in Equation 12 is replaced by $\zeta_s(0,\sigma_{bd})/2$. Using Equation 13, the variance in T_d (σ_t^2) is,

$$\sigma_t^2 = \frac{1}{\langle\lambda\rangle^2}\left(\ln^2(2)CV_\lambda^2 + \frac{2\sigma_n^2}{2-\alpha}\right). \tag{14}$$

For exponential growth, we also find,

$$\ln(\frac{L_d}{L_b}) = x_{n+1} - x_n + \ln(2) = \lambda T_d. \tag{15}$$

On substituting Equation 12 into Equation 15 we obtain to first order,

$$x_{n+1} \approx (1-\alpha)x_n + \zeta(0,\sigma_n). \tag{16}$$

On replacing the time additive noise term, $\zeta(0,\sigma_n)$, in Equation 16 with $\zeta_s(0,\sigma_{bd})/2$, we recover the recursive relation for size at birth obtained in the case of size additive noise shown in Equation 6. Hence, the model is insensitive to noise being size additive or time additive with a simple mapping for going from one noise type to another in the small noise

limit.

At steady state, x has a normal distribution with mean 0 and variance σ_x^2 whose value is given by,

$$\sigma_x^2 = \frac{\sigma_n^2}{\alpha(2-\alpha)}. \tag{17}$$

We note that some of the derivations above have also been presented in Ref. [16], but are provided here for completeness.

5.4 Predicting the results of statistical constructs applied on $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ and $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$

5.4.1 Obtaining the best linear fit

Next, we calculate the equation for the best linear fit for the choice of $\ln(\frac{L_d}{L_b})$ as y-axis and $\langle\lambda\rangle T_d$ as x-axis and vice versa. For simplicity, in this section, we will consider time additive division timing noise. However, the results obtained here will hold for size additive noise as well because the model is robust to the type of noise added as shown in the previous section.

First, we calculate the correlation coefficient (ρ_{exp}) for $\ln(\frac{L_d}{L_b})$ and time of division T_d,

$$\rho_{exp} = \frac{\langle(\ln(\frac{L_d}{L_b}) - \langle\ln(\frac{L_d}{L_b})\rangle)(T_d - \langle T_d\rangle)\rangle}{\sigma_l \sigma_t}, \tag{18}$$

where σ_l is the standard deviation in $\ln(\frac{L_d}{L_b})$. Using Equations 15 and 16 we obtain,

$$\ln(\frac{L_d}{L_b}) \approx \ln(2) - \alpha x_n + \zeta(0, \sigma_n). \tag{19}$$

Substituting Equations 13 and 19 into the numerator of Equation 18,

$$\langle(\ln(\frac{L_d}{L_b}) - \langle\ln(\frac{L_d}{L_b})\rangle)(T_d - \langle T_d\rangle)\rangle$$
$$= \langle(-\alpha x_n + \zeta(0,\sigma_n))\frac{(-\alpha x_n - \ln(2)\xi(0,CV_\lambda) + \zeta(0,\sigma_n))}{\langle\lambda\rangle}\rangle. \quad (20)$$

As the terms $\zeta(0,\sigma_n)$, $\xi(0,CV_\lambda)$ and x_n are independent of each other, $\langle\xi(0,CV_\lambda)\zeta(0,\sigma_n)\rangle = 0$, $\langle\xi(0,CV_\lambda)x_n\rangle = 0$ and $\langle x_n\zeta(0,\sigma_n)\rangle = 0$. Equation 20 simplifies to,

$$\langle(\ln(\frac{L_d}{L_b}) - \langle\ln(\frac{L_d}{L_b})\rangle)(T_d - \langle T_d\rangle)\rangle = (\alpha^2\sigma_x^2 + \sigma_n^2)\frac{1}{\langle\lambda\rangle}. \quad (21)$$

The variance of $\ln(\frac{L_d}{L_b})$ obtained using Equation 19 is,

$$\sigma_l^2 = \alpha^2\sigma_x^2 + \sigma_n^2 = \frac{2\sigma_n^2}{2-\alpha}. \quad (22)$$

Inserting Equations 14, 21 and 22 into Equation 18, we get,

$$\rho_{exp} = \sqrt{\frac{1}{1 + \frac{(1-\frac{\alpha}{2})\ln^2(2)CV_\lambda^2}{\sigma_n^2}}}. \quad (23)$$

The slope of a linear regression line is given by,

$$m = \rho\frac{\sigma_y}{\sigma_x}, \quad (24)$$

where σ_x, σ_y and ρ are the standard deviation of the x-variable, the standard deviation of the y-variable and the correlation coefficient of the (x,y) pair, respectively. The intercept is,

$$c = \langle y\rangle - m\langle x\rangle. \quad (25)$$

On the x-axis, we plot $\langle\lambda\rangle T_d$ and the y-axis is chosen as $\ln(\frac{L_d}{L_b})$. The slope for this choice (m_{tl}) can be calculated by,

$$m_{tl} = \rho_{exp}\frac{\sigma_l}{\sigma_t\langle\lambda\rangle}. \tag{26}$$

On substituting the values we get,

$$m_{tl} = \frac{1}{1+\frac{(1-\frac{\alpha}{2})\ln^2(2)CV_\lambda^2}{\sigma_n^2}}. \tag{27}$$

Only for $CV_\lambda \ll \sigma_n$ we would expect a slope close to 1.

The intercept (c_{tl}) for the $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot is given by,

$$c_{tl} = \langle\ln(\frac{L_d}{L_b})\rangle - m_{tl}\langle\langle\lambda\rangle T_d\rangle = \ln(2)\left(1-\frac{1}{1+\frac{(1-\frac{\alpha}{2})\ln^2(2)CV_\lambda^2}{\sigma_n^2}}\right). \tag{28}$$

However, if we choose the x-axis as $\ln(\frac{L_d}{L_b})$ and the y-axis is chosen as $\langle\lambda\rangle T_d$, we obtain the slope m_{lt},

$$m_{lt} = \rho_{exp}\frac{\sigma_t\langle\lambda\rangle}{\sigma_l}. \tag{29}$$

On substituting the values we obtain $m_{lt} = 1$ independent of the noise parameters and find that the intercept is zero.

5.4.2 Non-linearity in binned data

In the Main text, for the plot $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$, we find the binned data to be non-linear (see Figure 2C of the Main text). In this section, we explain the non-linearity observed using the model developed in the previous sections.

Binning data based on the x-axis means taking an average of the y-variable conditioned on the value of the x-variable. Mathematically, this amounts to calculating $\mathbb{E}[y \mid x]$ i.e., the conditional expectation of the y-variable given that x is fixed. In our case, we need to

523 calculate $\mathbb{E}[\ln(\frac{L_d}{L_b}) \mid \langle\lambda\rangle T_d]$. $\ln(\frac{L_d}{L_b}) = \lambda T_d$ by definition of exponential growth, hence,

$$\mathbb{E}[\ln(\frac{L_d}{L_b}) \mid \langle\lambda\rangle T_d] = \mathbb{E}[\lambda T_d \mid \langle\lambda\rangle T_d]. \tag{30}$$

524 Since T_d is fixed, this is equivalent to calculating $\mathbb{E}[\lambda \mid T_d]$. Using Equation 13,

$$\mathbb{E}[\lambda \mid T_d] = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda p(x,\xi,\zeta)\ \delta(T_d - (\frac{\ln(2)}{\langle\lambda\rangle} - \alpha\frac{x}{\langle\lambda\rangle} - \frac{\ln(2)\xi}{\langle\lambda\rangle} + \frac{\zeta}{\langle\lambda\rangle})\ dx\ d\xi\ d\zeta}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,\xi,\zeta)\ \delta(T_d - (\frac{\ln(2)}{\langle\lambda\rangle} - \alpha\frac{x}{\langle\lambda\rangle} - \frac{\ln(2)\xi}{\langle\lambda\rangle} + \frac{\zeta}{\langle\lambda\rangle})\ dx\ d\xi\ d\zeta}. \tag{31}$$

$p(x,\xi,\zeta)$ is the joint probability distribution of x and noise parameters ξ and ζ. Since, they are independent of each other, the joint distribution is product of the individual distributions $f_1(x)$, $f_2(\xi)$ and $f_3(\zeta)$, the distributions being Gaussian with mean 0 and standard deviation σ_x, CV_λ and σ_n, respectively. σ_x, σ_n are related by Equation 17. Since x, ξ, and ζ are narrowly distributed around zero, the contribution from large positive or negative values is extremely small. This ensures that T_d is also close to its mean and non-negative despite the limits of the integral being $-\infty$ to ∞. Using $\lambda = \langle\lambda\rangle + \langle\lambda\rangle\xi(0, CV_\lambda)$ in Equation 31,

$$\begin{aligned}&\mathbb{E}[\lambda \mid T_d]\\ &= \langle\lambda\rangle\left(1 + \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \xi f_1(x)f_2(\xi)f_3(\zeta)\ \delta(T_d - (\frac{\ln(2)}{\langle\lambda\rangle} - \alpha\frac{x}{\langle\lambda\rangle} - \frac{\ln(2)\xi}{\langle\lambda\rangle} + \frac{\zeta}{\langle\lambda\rangle}))\ dx\ d\xi\ d\zeta}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x)f_2(\xi)f_3(\zeta)\ \delta(T_d - (\frac{\ln(2)}{\langle\lambda\rangle} - \alpha\frac{x}{\langle\lambda\rangle} - \frac{\ln(2)\xi}{\langle\lambda\rangle} + \frac{\zeta}{\langle\lambda\rangle}))\ dx\ d\xi\ d\zeta}\right).\end{aligned} \tag{32}$$

525 On evaluating the integrals, we obtain,

$$\mathbb{E}[\lambda \mid T_d] = \langle\lambda\rangle\left(1 + \frac{1}{1 + \frac{2}{2-\alpha}\frac{\sigma_n^2}{CV_\lambda^2 \ln^2(2)}} - \frac{\frac{\langle\lambda\rangle T_d}{\ln(2)}}{1 + \frac{2}{2-\alpha}\frac{\sigma_n^2}{CV_\lambda^2 \ln^2(2)}}\right). \tag{33}$$

Thus, the trend of binned data is found to be,

$$\mathbb{E}[\ln(\frac{L_d}{L_b}) \mid \langle\lambda\rangle T_d] = \langle\lambda\rangle T_d \left(1 + \frac{1}{1 + \frac{2}{2-\alpha}\frac{\sigma_n^2}{CV_\lambda^2 \ln^2(2)}} - \frac{\frac{\langle\lambda\rangle T_d}{\ln(2)}}{1 + \frac{2}{2-\alpha}\frac{\sigma_n^2}{CV_\lambda^2 \ln^2(2)}}\right). \tag{34}$$

In the regime $CV_\lambda \ll \sigma_n$, the last two terms on the RHS of Equation 34 vanish and the binned data follows the trend y=x.

For the $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot, we need to calculate $\mathbb{E}[\langle\lambda\rangle T_d \mid \ln(\frac{L_d}{L_b})]$. Using Equations 13 and 19, we obtain,

$$\langle\lambda\rangle T_d = \ln(\frac{L_d}{L_b}) - \ln(2)\xi(0, CV_\lambda). \tag{35}$$

$\ln(\frac{L_d}{L_b})$ is independent of $\xi(0, CV_\lambda)$. Using this, we can write $\mathbb{E}[\langle\lambda\rangle T_d \mid \ln(\frac{L_d}{L_b})]$ as,

$$\begin{aligned}&\mathbb{E}[\langle\lambda\rangle T_d \mid \ln(\frac{L_d}{L_b})] \\ &= \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\langle\lambda\rangle T_d)\ f_2(\xi)\ f_4(\ln(\frac{L_d}{L_b}))\ \delta\left(\langle\lambda\rangle T_d\ - (\ln(\frac{L_d}{L_b}) - \ln(2)\xi)\right)\ d(\langle\lambda\rangle T_d)\ d\xi}{f_4(\ln(\frac{L_d}{L_b}))}.\end{aligned} \tag{36}$$

Note that the integral over $\langle\lambda\rangle T_d$ goes from $-\infty$ to ∞ although $\langle\lambda\rangle T_d$ cannot be negative. As before, this is not an issue because we assume $\langle\lambda\rangle T_d$ to be tightly regulated around $\ln(2)$ and the contribution to the integral from $-\infty$ to 0 is negligible. $f_4(\ln(\frac{L_d}{L_b}))$ denotes the probability distribution for $\ln(\frac{L_d}{L_b})$, the distribution being Gaussian with mean $\ln(2)$, and standard deviation σ_l which is calculated in Equation 22. Putting the Gaussian form of $f_2(\xi)$ into the integral and simplifying we get,

$$\mathbb{E}[\langle\lambda\rangle T_d \mid \ln(\frac{L_d}{L_b})] = \ln(\frac{L_d}{L_b}). \tag{37}$$

The trend of binned data to first order in noise and x is $\mathbb{E}[\langle\lambda\rangle T_d \mid \ln(\frac{L_d}{L_b})] = \ln(\frac{L_d}{L_b})$. This is shown in Figure 2D of the Main text where the binned data follows the y=x line.

5.5 Linear growth

In this section, we will focus on finding the equation of the best linear fit for relevant plots in the case of linear growth. The time at division for linear growth is given by,

$$T_d = \frac{L_d - L_b}{\lambda'}. \tag{38}$$

Note that λ' has units of [length/time] and is defined as the elongation speed. This is different from the exponential growth rate which has units [1/time]. Here, we will work with the normalized length at birth (l_b) and division (l_d),

$$T_d = \frac{l_d - l_b}{\lambda_{lin}}. \tag{39}$$

Consider the normalized elongation speed to be $\lambda_{lin} = \langle\lambda_{lin}\rangle + \langle\lambda_{lin}\rangle\xi_{lin}(0, CV_{\lambda,lin})$, where $\langle\lambda_{lin}\rangle$ is the mean normalized elongation speed for a lineage of cells and $\xi_{lin}(0, CV_{\lambda,lin})$ is normally distributed with mean 0 and standard deviation $CV_{\lambda,lin}$. Thus, the CV of elongation speed is $CV_{\lambda,lin}$. The regulation strategy which the cell undertakes is equivalent to that in previous sections and is given by $\mathrm{g}(l_b) = 2 + 2(1-\alpha)(l_b - 1)$. Note that we can obtain $\mathrm{g}(l_b)$ by Taylor expanding $\mathrm{f}(l_b)$ around $l_b = 1$. Using the regulation strategy $\mathrm{g}(l_b)$ and adding a size additive noise $\zeta_s(0, \sigma_{bd})$ which is independent of l_b, we find,

$$T_d = \frac{2 + 2(1-\alpha)(l_b^n - 1) + \zeta_s(0, \sigma_{bd}) - l_b^n}{\langle\lambda_{lin}\rangle(1 + \xi_{lin}(0, CV_{\lambda,lin}))}. \tag{40}$$

Note that we chose size additive division timing noise ($\zeta_s(0, \sigma_{bd})$) for convenience in this section. However, it can be shown as done previously that the model is robust to the noise in division timing being size additive or time additive. Assuming that the noise terms

$\xi_{lin}(0, CV_{\lambda,lin})$ and $\zeta_s(0, \sigma_{bd})$ are small, we obtain to first order,

$$T_d \approx \frac{(1-2\alpha)(l_b - 1) + 1 + \zeta_s(0, \sigma_{bd}) - \xi_{lin}(0, CV_{\lambda,lin})}{\langle \lambda_{lin} \rangle}. \tag{41}$$

The terms l_b, $\zeta_s(0, \sigma_{bd})$ and $\xi_{lin}(0, CV_{\lambda,lin})$ are independent of each other. The standard deviation of T_d (σ_t) can be calculated to be,

$$\sigma_t^2 = \frac{(1-2\alpha)^2 \sigma_b^2 + \sigma_{bd}^2 + CV_{\lambda,lin}^2}{\langle \lambda_{lin} \rangle^2}. \tag{42}$$

Assuming perfectly symmetric division and using $l_d^n = g(l_b^n) + \zeta_s(0, \sigma_{bd})$, we find the recursive relation for l_b^n to be,

$$l_d^n - l_b^n = 2l_b^{n+1} - l_b^n = (1-2\alpha) l_b^n + 2\alpha + \zeta_s(0, \sigma_{bd}). \tag{43}$$

Note that Equation 43 is the same as Equation 6 under the approximation $x_n = l_b^n - 1$. At steady state, the standard deviation of l_b is denoted by σ_b and using Equation 43 its value is obtained to be,

$$\sigma_b^2 = \frac{\sigma_{bd}^2}{4\alpha(2-\alpha)}. \tag{44}$$

Similarly, the standard deviation of l_d-l_b, or equivalently $\lambda_{lin} T_d$, denoted by $\sigma_{l,lin}$, is calculated to be,

$$\sigma_{l,lin}^2 = \frac{4\alpha + 1}{4\alpha(2-\alpha)} \sigma_{bd}^2. \tag{45}$$

For linear growth, a natural plot is l_d-l_b vs $\langle \lambda_{lin} \rangle T_d$ (reminiscent of the $\ln(\frac{L_d}{L_b})$ vs $\langle \lambda \rangle T_d$ plot for exponential growth). To calculate the slope of the best linear fit, we have to calculate the correlation coefficient ρ_{lin} given by,

$$\rho_{lin} = \frac{\langle (l_d - l_b - \langle l_d - l_b \rangle) \left(\langle \lambda_{lin} \rangle T_d - \langle \langle \lambda_{lin} \rangle T_d \rangle \right) \rangle}{\langle \lambda_{lin} \rangle \sigma_{l,lin} \sigma_t}. \tag{46}$$

Again using the independence of terms l_b, $\zeta_s(0, \sigma_{bd})$ and $\xi_{lin}(0, CV_{\lambda,lin})$ from each other, we get,

$$\rho_{lin} = \frac{(1-2\alpha)^2\sigma_b^2 + \sigma_{bd}^2}{\langle\lambda_{lin}\rangle\sigma_{l,lin}\sigma_t} = \frac{\sigma_{l,lin}}{\langle\lambda_{lin}\rangle\sigma_t}. \tag{47}$$

The slope of best linear fit for the plot $l_d - l_b$ vs $\langle\lambda_{lin}\rangle T_d$ is given by,

$$m_{tl,lin} = \rho_{lin}\frac{\sigma_{l,lin}}{\langle\lambda_{lin}\rangle\sigma_t} = \frac{1}{1 + \frac{CV_{\lambda,lin}^2 4\alpha(2-\alpha)}{\sigma_{bd}^2(4\alpha+1)}}. \tag{48}$$

The intercept $c_{tl,lin}$ is found to be,

$$c_{tl,lin} = \langle l_d - l_b\rangle - m_{tl,lin}\langle\langle\lambda_{lin}\rangle T_d\rangle = 1 - \frac{1}{1 + \frac{CV_{\lambda,lin}^2 4\alpha(2-\alpha)}{\sigma_{bd}^2(4\alpha+1)}}. \tag{49}$$

On flipping the axis, the slope ($m_{lt,lin}$) for the plot $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ is obtained to be,

$$m_{lt,lin} = \rho_{lin}\frac{\langle\lambda_{lin}\rangle\sigma_t}{\sigma_{l,lin}} = 1. \tag{50}$$

The intercept $c_{lt,lin}$ is found to be,

$$c_{lt,lin} = \langle\langle\lambda_{lin}\rangle T_d\rangle - m_{lt,lin}\langle l_d - l_b\rangle = 0. \tag{51}$$

The best linear fit for the $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot follows the trend y=x.

Simulations of the adder model for linearly growing cells were carried out. The deviation of the best linear fit for the $l_d - l_b$ vs $\langle\lambda_{lin}\rangle T_d$ plot from the y=x line is shown in Figure 3- figure supplement 1A, while in Figure 3- figure supplement 1B, the best linear fit for the plot $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ is shown to agree with the y=x line.

5.6 Differentiating linear from exponential growth

In this section, we explore the equation for the best linear fit of $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot in the case of exponential growth and $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot for linear growth. Intuitively, we expect the best linear fit in both cases to deviate from the y=x line. In this section, we will calculate the best linear fit explicitly. Surprisingly, we will find that, in the case of linear growth, the best linear fit for the $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot follows the y=x line closely.

Let us begin with exponential growth with growth rate, $\lambda = \langle\lambda\rangle + \langle\lambda\rangle\xi(0, CV_\lambda)$ as defined previously. Again, $\xi(0, CV_\lambda)$ has a normal distribution with mean 0 and standard deviation CV_λ, it being the CV of the growth rate. The time at division is given by Equation 7. The average growth rate $\langle\lambda\rangle = \langle\frac{\ln(2)}{T_d}\rangle \approx \frac{\ln(2)}{\langle T_d\rangle}$. For exponential growth, we will plot $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$. As previously defined, $\langle\lambda_{lin}\rangle$ is the mean normalized elongation speed and $\langle\lambda_{lin}\rangle = \langle\frac{1}{T_d}\rangle \approx \frac{1}{\langle T_d\rangle}$. $\langle\lambda_{lin}\rangle$ is related to $\langle\lambda\rangle$ by,

$$\langle\lambda_{lin}\rangle = \frac{\langle\lambda\rangle}{\ln(2)}. \tag{52}$$

$l_d - l_b$ can be calculated by using the regulation strategy f(l_b) introduced in Section 5.2 and a normally distributed size additive noise $\zeta_s(0, \sigma_{bd})$. Note that we have chosen the noise in division timing to be size additive. However, the model is robust to the choice of type of noise as we showed in Section 5.3. Using Equations 5 and 6 we obtain,

$$l_d^n - l_b^n \approx 1 + (1 - 2\alpha)x_n + \zeta_s(0, \sigma_{bd}). \tag{53}$$

Using Equation 11, $\langle\lambda_{lin}\rangle T_d$ is obtained to be,

$$\langle\lambda_{lin}\rangle T_d = 1 - \frac{\alpha x}{\ln(2)} - \xi(0, CV_\lambda) + \frac{\zeta_s(0, \sigma_{bd})}{2\ln(2)}. \tag{54}$$

To calculate the expression for $m_{lt,lin}$, the slope of the best linear fit for $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot,

we first calculate ρ_{lin} given by Equation 46. The expression for $\sigma_{l,lin}$ (standard deviation of $l_d - l_b$) and σ_t (standard deviation of T_d) are found to be,

$$\sigma_{l,lin}^2 = (1-2\alpha)^2\sigma_x^2 + \sigma_{bd}^2, \tag{55}$$

$$\sigma_t^2 = \frac{1}{\langle\lambda_{lin}\rangle^2}\left((\frac{\alpha\sigma_x}{\ln(2)})^2 + CV_\lambda^2 + (\frac{\sigma_{bd}}{2\ln(2)})^2 \right). \tag{56}$$

σ_x is related to σ_n via Equation 17. In Section 5.3, we also showed that $\sigma_n = \frac{\sigma_{bd}}{2}$. Using these, we can write,

$$\sigma_x^2 = \frac{\sigma_{bd}^2}{4\alpha(2-\alpha)}. \tag{57}$$

Now using the expressions for σ_t, $\sigma_{l,lin}$ and the fact that x, $\xi(0, CV_\lambda)$ and $\zeta_s(0, \sigma_{bd})$ are independent of each other, we get,

$$\rho_{lin} = \frac{\frac{(2\alpha-1)\alpha\sigma_x^2}{\ln(2)} + \frac{\sigma_{bd}^2}{2\ln(2)}}{\langle\lambda_{lin}\rangle\sigma_{l,lin}\sigma_t}. \tag{58}$$

For the plot $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$, the slope $m_{lt,lin}$ is given by,

$$m_{lt,lin} = \rho_{lin}\frac{\sigma_t\langle\lambda_{lin}\rangle}{\sigma_{l,lin}} = \frac{\frac{(2\alpha-1)\alpha\sigma_x^2}{\ln(2)} + \frac{\sigma_{bd}^2}{2\ln(2)}}{\sigma_{l,lin}^2}. \tag{59}$$

Inserting Equation 55 into Equation 59 and substituting σ_x^2 given by Equation 57, we obtain,

$$m_{lt,lin} = \frac{1}{\ln(2)}\frac{3\alpha}{4\alpha+1}. \tag{60}$$

The intercept $c_{lt,lin}$ is found to be,

$$c_{lt,lin} = \langle\langle\lambda_{lin}\rangle T_d\rangle - m_{lt,lin}\langle l_d - l_b\rangle = 1 - \frac{1}{\ln(2)}\frac{3\alpha}{4\alpha+1}. \tag{61}$$

For the adder model ($\alpha = \frac{1}{2}$), we get the value of slope $m_{lin,lt} = \frac{1}{2\ln(2)} \approx 0.7213$ and intercept $c_{lin,lt} = 1 - \frac{1}{2\ln(2)} \approx 0.279$. This is different from the best linear fit obtained for same regulatory mechanism controlling division in linearly growing cells where we found that the best linear fit follows the y=x line. Intuitively, we expect the best linear fit of $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot to deviate from y=x line in the case of exponential growth. We showed analytically that for a class of models where birth controls division, it is indeed the case. This is also shown using simulations of the adder model in Figure 3- figure supplement 1C.

In Section 5.4.1, we found the best linear fit for $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot to follow the y=x line for exponentially growing cells where division is regulated by birth event via regulation strategy f(l_b). Next, we calculate the equation for the best linear fit of $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot given growth is linear. The model for division control will be same as that in Section 5.5 i.e., the regulation strategy for division is given by $g(l_b) = 2 + 2(1-\alpha)(l_b - 1)$ which is also equivalent to f(l_b). The linearly growing cells grow with elongation speed $\lambda_{lin} = \langle\lambda_{lin}\rangle(1 + \xi_{lin}(0, CV_{\lambda,lin}))$. As discussed before, $\xi_{lin}(0, CV_{\lambda,lin})$ has a normal distribution with mean 0 and standard deviation $CV_{\lambda,lin}$, it being the CV of the elongation speed. Using Equations 5 and 6, we get,

$$\ln(\frac{L_d}{L_b}) = \ln(2) - \alpha x^n + \frac{\zeta_s(0, \sigma_{bd})}{2}. \tag{62}$$

Using Equations 5 and 52, we obtain from Equation 41,

$$\langle\lambda\rangle T_d = \ln(2) + (1-2\alpha)\ln(2)x + \ln(2)\zeta_s(0, \sigma_{bd}) - \ln(2)\xi_{lin}(0, CV_{\lambda,lin}). \tag{63}$$

Since x, $\xi_{lin}(0, CV_{\lambda,lin})$ and $\zeta_s(0, \sigma_{bd})$ are uncorrelated, the standard deviation of $\ln(\frac{L_d}{L_b})$ and T_d denoted by σ_l and σ_t respectively are calculated to be,

$$\sigma_l^2 = \alpha^2\sigma_x^2 + \frac{\sigma_{bd}^2}{4}, \tag{64}$$

$$\sigma_t^2 = \frac{\ln^2(2)}{\langle\lambda\rangle^2}((1-2\alpha)^2\sigma_x^2 + \sigma_{bd}^2 + CV_{\lambda,lin}^2). \tag{65}$$

We calculate the correlation coefficient for the pair $(\ln(\frac{L_d}{L_b}),\ \langle\lambda\rangle T_d)$. Since the correlation coefficient is unaffected by multiplying one of the variables with a positive constant, we can calculate the correlation coefficient for the pair $(\ln(\frac{L_d}{L_b}),\ T_d)$ or ρ_{exp} as given by Equation 18. Using the independence of terms x, $\xi_{lin}(0, CV_{\lambda,lin})$ and $\zeta_s(0, \sigma_{bd})$,

$$\rho_{exp} = \frac{\ln(2)(\sigma_x^2(2\alpha-1)\alpha + \frac{\sigma_{bd}^2}{2})}{\langle\lambda\rangle\sigma_l\sigma_t}. \tag{66}$$

For the plot $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$, the slope m_{lt} of the best linear fit is given by,

$$m_{lt} = \rho_{exp}\frac{\sigma_t\langle\lambda\rangle}{\sigma_l} = \frac{\ln(2)(\sigma_x^2(2\alpha-1)\alpha + \frac{\sigma_{bd}^2}{2})}{\sigma_l^2}. \tag{67}$$

Inserting Equation 64 into Equation 67 and using Equation 57, we get,

$$m_{lt} = \frac{3}{2}\ln(2) \approx 1.0397. \tag{68}$$

Similarly the intercept (c_{lt}) for the plot $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ is found to be,

$$c_{lt} = \langle\langle\lambda\rangle T_d\rangle - m_{lt}\langle\ln(\frac{L_d}{L_b})\rangle = \ln(2)(1 - \frac{3}{2}\ln(2)) \approx -0.0275. \tag{69}$$

This is very close to y=x trend obtained for the same regulatory mechanism controlling division in exponentially growing cells (Figure 3A).

5.7 Growth rate vs age and elongation speed vs age plots.

In the previous sections, we found that binning and linear regression on the plot $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$, and the plot obtained by interchanging the axes, were inadequate to identify the mode

of growth. In this section, we try to validate the growth rate vs age plot as a method to elucidate the mode of growth.

In addition to cell size at birth and division and the generation time, cell size trajectories (cell size, L vs time from birth, t) were obtained for multiple cell cycles. In our case, the cell size trajectories were collected either via simulations (in Figure 3B) or from experiments (for Figures 4A-4C) at intervals of 4 min. For each trajectory, growth rate at time t or age $\frac{t}{T_d}$ is calculated as $\frac{1}{L(t)}\frac{L(t+\Delta t)-L(t)}{\Delta t}$ where Δt is the time between consecutive measurements. To obtain elongation speed vs age plots, the formula before needs to be replaced with $\frac{L(t+\Delta t)-L(t)}{\Delta t}$. The growth rate is interpolated to contain 200 points at equal intervals of time for each cell trajectory. The growth rate trends appear to be robust with regards to a different number of interpolated points (from 100 to 500 points). To obtain the growth rate trend as a function of cell age, we use the method previously applied in Ref. [39]. In this method, growth rate is binned based on age for each individual trajectory (50 bins) and the average growth rate is obtained in each of the bins. The binned data trend for growth rate vs age is then found by taking the average of the growth rate in each bin over all trajectories. Binning the growth rate for each trajectory ensures that each trajectory has an equal contribution to the final growth rate trend so as to avoid inspection bias. This step is especially important when data collected at equal intervals of time is analyzed. In such a case, cells with larger generation times have a greater number of measurements than cells with smaller generation times. Obtaining the growth rate trend without binning growth rate for each trajectory would have biased the binned data trend for the growth rate vs age plot to a smaller value because of over-representation by slower-growing cells (or equivalently cells with longer generation time). This bias towards lower growth rate values in the growth rate vs age plots is an instance of inspection bias.

In Figures 4A-4C, we find the growth rate obtained from *E. coli* experiments to change within the cell cycle. In the two slower growth media (Figures 4A, 4B), the growth rate is

found to increase with cell age while for the fastest growth media (Figure 4C) the growth rate follows a non-monotonic behaviour similar to that observed in Ref. [39] for *B. subtilis*. Abrupt changes in growth rate are reported at constriction in Refs. [41, 42]. We find that the growth rate changes start before constriction in the two slower growth conditions considered. One possibility is that this increase is due to preseptal cell wall synthesis [56]. Preseptal cell wall synthesis does not require activity of PBP3 (FtsI) but instead relies on bifunctional glycosyltransferases PBP1A and PBP1B that link to FtsZ via ZipA. One hypothesis that can be tested in future works is that at the onset of constriction, activity from PBP1A and PBP1B starts to gradually shift to the PBP3/FtsW complex and therefore no abrupt change in growth rate is observed. In the fastest growth condition (glucose-cas medium), we find that the increase in growth rate approximately coincides with onset of constriction, in agreement with the previous findings [41, 42].

In Figures 4A-4C, the growth rate trends are not obtained for age close to one. This is because growth rate at age $= 1$ is given by $\frac{1}{L(T_d)}\frac{L(T_d+\Delta t)-L(T_d)}{\Delta t}$ and this requires knowing the cell lengths beyond the division event ($L(T_d + \Delta t)$). To estimate growth rates at age close to one, we approximate $L(T_d + \Delta t)$ to be the sum of cell sizes of the two daughter cells. In order to minimize inspection bias, we considered only those cell size trajectories which had $L(t)$ data for 12 min after division (corresponding to an age of approximately 1.1). However, the growth rate trends in all three growth media were robust with regards to a different time for which $L(t)$ was considered (4 min to 20 min after division). We use the binning procedure discussed before in this section. To validate this method, we applied it on synthetic data obtained from the simulations of exponentially growing cells following the adder and the adder per origin model. Cells were assumed to divide in a perfectly symmetric manner and both of the daughter cells were assumed to grow with the same growth rate, independent of the growth rate in the mother cell. The growth rate trends for the two models considered (adder and adder per origin) are expected to be constant even for cell age

> 1. We found that the growth rate trends were indeed approximately constant as shown in Figure 4- figure supplement 1D. We also considered linear growth with division controlled via an adder model. The daughter cells were assumed to grow with the same elongation speed, independent of the elongation speed in the mother cell. In this case, we expect the elongation speed trend to be constant for cell age > 1. This is indeed what we observed as shown in the inset of Figure 4- figure supplement 1D. We used this method on *E. coli* experimental data and found that the growth rate trends obtained for the three growth conditions (Figure 4- figure supplements 1A-1C) were consistent with that shown in Figures 4A-4C in the relevant age ranges. For cell age close to one, we found that the growth rate decreased to a value close to the growth rate near cell birth (age ≈ 0) for all three growth conditions considered.

In summary, we find that the growth rate vs age plots are a consistent method to probe the mode of cell growth within a cell cycle.

5.8 Growth rate vs time from specific event plots are affected by inspection bias

To probe the growth rate trend in relation to a specific cell cycle event, for example cell birth, growth rate vs time from birth plots are obtained for simulations of exponentially growing cells following the adder model. In the growth rate vs time from birth plot, the rate is found to stay constant and then decrease at longer times (Figure 3- figure supplement 2C) even though cells are exponentially growing. Because of inspection bias (or survivor bias), at later times, only the cells with larger generation times (or slower growth rates) "survive". The average generation time of the cells averaged upon in each bin of Figure 3- figure supplement 2C is shown in Figure 3- figure supplement 2D. The decrease in growth rate in Figure 3- figure supplement 2C occurs around the same time when an increase in generation time is observed in Figure 3- figure supplement 2D. Thus, the trend in growth rate is biased towards

lower values at longer times. The problem might be circumvented by restricting the time on the x-axis to the smallest generation time of all the cell cycles considered [31].

To check for growth rate changes at constriction, we used plots of growth rate vs time from constriction ($t - T_n$). Growth rate trends obtained from *E. coli* experimental data show a decrease at the edges of the plots (Figure 4- figure supplements 2A, 2C, and 2E). These deviate from the trends obtained using the growth rate vs age plots (Figures 4A-4C). To investigate this discrepancy, we use a model which takes into account the constriction and the division event. Currently it is unknown how constriction is related to division. For the purpose of methods validation, we use a model where cells grow exponentially, constriction occurs after a constant size addition from birth, and division occurs after a constant size addition from constriction. Note that other models where constriction occurs after a constant size addition from birth while division occurs after a constant time from constriction, as well as a mixed timer-adder model proposed in Ref. [42], lead to similar results. We expect the growth rate trend to be constant for exponentially growing cells. However, we find using numerical simulations that it decreases at the plot edges both before and after the constriction event (Figure 3- figure supplement 2A). This decrease can be attributed to inspection bias. The average growth rate in time bins at the extremes are biased by cells with smaller growth rates. This is shown in Figure 3- figure supplement 2B where the average generation time for the cells contributing in each of the bins of Figure 3- figure supplement 2A is plotted. The time at which the growth rate decreases on both sides of the constriction event is close to the time at which the average generation time increases. For example, in alanine medium, the generation time for each of the bins is plotted in Figure 4- figure supplement 2B. The average generation time for the cells contributing to each of the bins is almost constant for the timings between -80 min to 20 min. Thus, for this time range the changes in growth rate are not because of inspection bias but are a real biological effect. The behavior of growth rate within this time range in Figure 4- figure supplement 2A is in agreement with the trend

in growth rate vs age plot of Figure 4A. On accounting for inspection bias, the growth rate vs age plots agree with the growth rate vs time from constriction plots in other growth media as well (Figure 4- figure supplement 2C, Figure 4- figure supplement 2E). Thus, growth rate vs time plots are also a consistent method to probe growth rate modulation in the time range when avoiding the regimes prone to inspection bias.

5.9 Results of elongation speed vs size plots are model-dependent.

Cells assumed to undergo exponential growth have elongation speed proportional to their size. In the case of exponential growth, the binned data trend of the plot elongation speed vs size is expected to be linear with the slope of the best linear fit providing the value of growth rate and intercept being zero. In this section, we use the simulations to test if binning and linear regression on the elongation speed vs size plots are suitable methods to differentiate exponential growth from linear growth [43].

To test the method, we generate cell size trajectories using simulations of the adder model with a size additive division timing noise and assuming exponential growth. Elongation speed at size $L(t)$ is calculated for each trajectory as $\frac{L(t+\Delta t)-L(t)}{\Delta t}$ where Δt is the time between consecutive measurements ($= 4$ min in our case). Each trajectory is binned into 10 equally sized bins based on their cell sizes and the average elongation speed is obtained for each bin. The final trend of elongation speed as a function of size is then obtained by binning (based on size) the pooled average elongation speed data of all the cell cycles.

We find that the binned data trend is linear with the slope of the best linear fit close to the average growth rate considered in the simulations (Figure 3- figure supplement 3D). This is in agreement with our expectations for exponential growth. In order to check if this method could differentiate between exponential growth and linear growth, we used simulations of the adder model undergoing linear growth to generate cell size trajectories for multiple cell cycles. For linear growth, elongation speed is expected to be constant, independent of its

cell size. The binned data trend for the elongation speed vs size plot is also obtained to be constant for the simulations of linearly growing cells (Figure 3- figure supplement 3B). The intercept of the best linear fit obtained is close to the average elongation speed considered in the simulations. The binned data trend for linear and exponential growth are clearly different as shown in Figure 3- figure supplement 3B and Figure 3- figure supplement 3D, respectively, and this result holds for a broad class of models where the division event is controlled by birth and the growth rate (for exponential growth)/elongation speed (for linear growth) is distributed normally and independently between cell-cycles.

Next, we consider the adder per origin cell cycle model for exponentially growing cells [17]. In this model space, the cell initiates DNA replication by adding a constant size per origin from the previous initiation size. The division occurs on average after a constant time from initiation. For exponentially growing cells, the binned data trend is still expected to be linear as before. Instead, we find using simulations that the trend is non-linear and it might be misinterpreted as non-exponential growth (Figure 3- figure supplement 3F).

Thus, the results of binning and linear regression for the plot elongation speed vs size is model-dependent.

5.10 Interchanging axes in growth rate vs inverse generation time plot might lead to different interpretations.

So far, our discussion was focused on the question of mode of single-cell growth. A related problem regards the relation between growth rate (λ) and the inverse generation time ($\frac{1}{T_d}$). On a population level, the two are clearly proportional to each other. However, single-cell studies based on binning showed an intriguing non-linear dependence between the two, with the two variables becoming uncorrelated in the faster-growth media. [25, 57]. Within the same medium, the binned data curve for the plot λ vs $\frac{1}{T_d}$ flattened out for faster dividing

cells. The trend in the binned data was different from the trend of y= ln(2)x line as observed for the population means. A priori one might speculate that the flattening in faster dividing cells could be because the faster dividing cells might have less time to adapt their division rate to transient fluctuations in the environment. Kennard *et al.* [57] insightfully also plotted $\frac{1}{T_d}$ vs λ and found a collapse of the binned data for all growth conditions onto the y = ln(2)x line. These results are reminiscent of what we previously showed for the relation of $\ln(\frac{L_d}{L_b})$ and $\langle\lambda\rangle T_d$.

In the following, we will elucidate why this occurs in this case using an underlying model and predicting the trend based on it. We use simulations of the adder model undergoing exponential growth. The parameters for size added in a cell cycle and mean growth rates are extracted from the experimental data. CV of growth rate is assumed lower in faster-growth media as observed by Kennard *et al.* Using this model, we could obtain the same pattern of flattening at faster-growth conditions that is observed in the experiments (Figure 2- figure supplement 2A). The population mean for λ and $\frac{1}{T_d}$ follows the expected y=ln(2)x equation (shown as black dashed line) as was the case in experiments. Intuitively, such a departure from the expected y=ln(2)x line for the single cell data can again be explained by determining the effect of noise on variables plotted on both axes. As previously stated T_d is affected by both growth rate noise and noise in division timing while growth rate fluctuates independently of other sources of noise. This does not agree with the assumption for binning as noise in division timing affects the x-axis variable rather than the y-axis variable. In such a case, the trend in the binned data might not follow the expected y=ln(2)x line. However, on interchanging the axes, we would expect the assumptions of binning to be met and the trend to follow the y=$\frac{1}{\ln(2)}$x line (Figure 2- figure supplement 2B).

5.11 Data and simulations

5.11.1 Experimental data

Experimental data obtained by Tanouchi *et al.* [10] was used to plot L_d vs L_b shown in Figure 1A. *E. coli* cells were grown at 25°C in a mother machine device and the length at birth and division were collected for multiple cell cycles. L_d vs L_b plot was obtained using these cells and linear regression performed on it provided a best linear fit.

Data from recent mother machine experiments on *E. coli* was used to make all other plots. Details are provided in Section 5.1 and Ref. [32]. The experiments were conducted at 28°C in three different growth conditions - alanine, glycerol and glucose-cas (also see Section 5.1). Cell size trajectories were collected for multiple cell cycles and all of the data collected were considered while making the plots in the paper.

5.11.2 Simulations

MATLAB R2021a was used for simulations. Simulations of the adder model for exponentially growing cells were carried out over a single lineage of 2500 generations (Figures 2C, 2D, Figure 3- figure supplement 1C). The mean length added between birth and division was set to 1.73 μm in line with the experimental results for alanine medium. Growth rate was variable and sampled from a normal distribution at the start of each cell cycle. The mean growth rate was set to $\frac{\ln(2)}{\langle T_d \rangle}$, where $\langle T_d \rangle = 212$ min and coefficient of variation (CV) = $CV_\lambda = 0.15$. The noise in division timing was assumed to be time additive with mean 0 and standard deviation $\frac{\sigma_n}{\langle \lambda \rangle}$, where $\sigma_n = 0.15$. The binning data trends and the best linear fits obtained using these simulations could be compared with the analytical results obtained in Sections 5.4.2 and 5.6.

For simulations of linear growth (Figures 3A-3B, Figure 3- figure supplements 1A, 1B, 3A, 3B, Figure 4- figure supplement 1D), the mean growth rate was set to $\frac{\langle L_d - L_b \rangle}{\langle T_d \rangle}$, with the values

of $\langle L_d - L_b \rangle$ and $\langle T_d \rangle$ used as mentioned previously. The noise in division timing was size additive with standard deviation = $0.15\langle L_b \rangle$. Noise was also considered to be size additive with the same standard deviation for the simulations of exponentially growing cells shown in Figure 3B, Figure 3- figure supplements 2C, 3C, 3D, and Figure 4- figure supplement 1D.

For Figure 3B, Figure 3- figure supplements 3E, 3F, Figure 4- figure supplement 1D, simulations were carried out over a lineage of 2500 generations for exponentially growing cells following the adder per origin model. In the simulations, the time increment is 0.01 min. The initial condition for the simulations is that cells are born and initiate DNA replication at time t=0 but the results are independent of initial conditions. The number of origins is also tracked throughout the simulations beginning with an initial value of 2. Cells divide into two daughter cells in a perfectly symmetrical manner (no noise in division ratio), and one of the daughter cells is discarded for the next cell cycle. In simulations, the growth rate was fixed within a cell cycle but varied between different cell cycles. On division, the growth rate for that cell cycle was drawn from a normal distribution with mean $\langle \lambda \rangle$ and coefficient of variation (CV_λ) whose values were fixed using the experimental data from alanine medium. The total length at which the next initiation happens is determined by,

$$L_i^{tot,next} = L_i + O\Delta_{ii}, \tag{70}$$

where Δ_{ii} is the length added per origin and O is the number of origins. To determine $L_i^{tot,next}$, Δ_{ii} was drawn on reaching initiation length from a normal distribution. The mean and CV of Δ_{ii} was obtained from experiments done in alanine medium. In the adder per origin model, division happens after a C+D time from initiation. The division length (L_d) is obtained to be,

$$L_d = L_i e^{\lambda(C+D)}. \tag{71}$$

In the simulations, once the initiation length was reached, the corresponding division oc-

curred a time C+D after initiation. C+D timings for each initiation event were again drawn from a normal distribution with the same mean and CV as that of the experiments in alanine medium.

For Figure 3- figure supplement 2A, cells were assumed to grow exponentially in the simulations. The constriction length (L_n) was set to be,

$$L_n = L_b + \Delta_{bn}. \tag{72}$$

The length added (Δ_{bn}) was assumed to have a normal distribution with the mean length added between birth and constriction set to 1.18 μm and the CV = 0.23, in line with the experimental results for alanine medium. The length at division was set as,

$$L_d = L_n + \Delta_{nd}. \tag{73}$$

The length added (Δ_{nd}) was also assumed to have a normal distribution with the mean length added set to 0.53 μm and the CV = 0.26, again in line with the experimental results for alanine medium.

For Figure 3B, Figure 3- figure supplements 2A-2D, 3A-3F, Figure 4- figure supplement 1D, the cell sizes are recorded within the cell cycle at equal intervals of 4 min, similar to that in the *E. coli* experiments of Ref. [32].

For simulations shown in Figure 4- figure supplement 1D, the cell size trajectories are obtained at intervals of 4 min beyond the current cell-cycle. The size after the division event is said to be the sum of the sizes of the daughter cells. It is also further assumed that the daughter cells are equal in size (perfectly symmetric division) and they both grow with the same growth rate (for exponential growth) or elongation speed (for linear growth). The growth rates/elongation speeds for the daughter cells are sampled from a normal distribution with a mean and CV as discussed before. The cell size trajectories are recorded for 80 min

after the division event in the current cell cycle.

In Figure 2- figure supplement 2, simulations of the adder model for exponentially growing cells were carried out until a population of 5000 cells was reached. The parameters for size added in a cell cycle and mean growth rates were extracted from the experimental data [57]. The value of σ_n used in all growth conditions was 0.17 while CV_λ decreased in faster growth conditions (0.2 in the three slowest growth conditions, 0.12 and 0.07 in the second fastest and fastest growth conditions respectively).

6 Acknowledgements

The authors thank Ethan Levien, Jie Lin for useful discussions, Jane Kondev, Xili Liu, and Marco Cosentino Lagomarsino for their useful feedback on the manuscript, Da Yang and Scott Retterer for help in microfluidic chip making, and Rodrigo Reyes-Lamothe for a kind gift of strain. Authors acknowledge technical assistance and material support from the Center for Environmental Biotechnology at the University of Tennessee. A part of this research was conducted at the Center for Nanophase Materials Sciences, which is sponsored at Oak Ridge National Laboratory by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy. This work has been supported by the US-Israel BSF research grant 2017004 (JM), the National Institutes of Health award under R01GM127413 (JM), NSF CAREER 1752024 (AA), NIH grant 103346 (PK) and NSF award 1806818 (PK).

7 Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

References

1. Osella, M., Tans, S.J., and Lagomarsino, M.C. (2017). Step by Step, Cell by Cell: Quantification of the Bacterial Cell Cycle. Trends in Microbiology *25*, 250–256.

2. Facchetti, G., Chang, F., and Howard, M. (2017). Controlling cell size through sizer mechanisms. Current Opinion in Systems Biology *5*, 86–92.

3. Ho, P.Y., Lin, J., and Amir, A. (2018). Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects. Annual Review of Biophysics *47*, 251–271.

4. Soifer, I., Robert, L., and Amir, A. (2016). Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy. Current Biology *26*, 356–361.

5. Jun, S., Si, F., Pugatch, R., and Scott, M. (2018). Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. Reports on Progress in Physics *81*, 056601.

6. Amir, A. and Balaban, N.Q. (2018). Learning from noise: how observing stochasticity may aid microbiology. Trends in Microbiology *26*, 376–385.

7. Kohram, M., Vashistha, H., Leibler, S., Xue, B., and Salman, H. (2021). Bacterial growth control mechanisms inferred from multivariate statistical analysis of single-cell measurements. Current Biology *31*, 955–964.

8. Hogg, D.W., Bovy, J., and Lang, D. (2010). Data analysis recipes: Fitting a model to data. arXiv preprint arXiv:1008.4686 .

9. Kiviet, D.J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., and Tans, S.J. (2014). Stochasticity of metabolism and growth at the single-cell level. Nature *514*, 376–379.

10. Tanouchi, Y., Pai, A., Park, H., Huang, S., Buchler, N.E., and You, L. (2017). Long-term growth data of *Escherichia coli* at a single-cell level. Scientific Data *4*, 1–5.

11. Harris, L.K. and Theriot, J.A. (2016). Relative rates of surface and volume synthesis set bacterial cell size. Cell *165*, 1479–1492.

12. Si, F., Le Treut, G., Sauls, J.T., Vadia, S., Levin, P.A., and Jun, S. (2019). Mechanistic origin of cell-size control and homeostasis in bacteria. Current Biology *29*, 1760–1770.

13. Amir, A. (2014). Cell size regulation in bacteria. Physical Review Letters *112*, 208102.

14. Campos, M., Surovtsev, I.V., Kato, S., Paintdakhi, A., Beltran, B., Ebmeier, S.E., and Jacobs-Wagner, C. (2014). A constant size extension drives bacterial cell size homeostasis. Cell *159*, 1433–1446.

15. Taheri-Araghi, S., Bradde, S., Sauls, J.T., Hill, N.S., Levin, P.A., Paulsson, J., Vergassola, M., and Jun, S. (2015). Cell-size control and homeostasis in bacteria. Current Biology *25*, 385–391.

16. Eun, Y.J., Ho, P.Y., Kim, M., LaRussa, S., Robert, L., Renner, L.D., Schmid, A., Garner, E., and Amir, A. (2018). Archaeal cells share common size control with bacteria despite noisier growth and division. Nature Microbiology *3*, 148–154.

17. Ho, P.Y. and Amir, A. (2015). Simultaneous regulation of cell size and chromosome replication in bacteria. Frontiers in Microbiology *6*, 662.

18. Micali, G., Grilli, J., Osella, M., and Lagomarsino, M.C. (2018). Concurrent processes set *E. coli* cell division. Science Advances *4*, eaau3324.

19. Witz, G., van Nimwegen, E., and Julou, T. (2019). Initiation of chromosome replication controls both division and replication cycles in *E. coli* through a double-adder mechanism. eLife *8*, e48063.

20. Barber, F., Ho, P.Y., Murray, A.W., and Amir, A. (2017). Details matter: noise and model structure set the relationship between cell size and cell cycle timing. Frontiers in Cell and Developmental Biology *5*, 92.

21. Facchetti, G., Knapp, B., Chang, F., and Howard, M. (2019). Reassessment of the basis of cell size control based on analysis of cell-to-cell variability. Biophysical Journal *117*, 1728–1738.

22. Godin, M., Delgado, F.F., Son, S., Grover, W.H., Bryan, A.K., Tzur, A., Jorgensen, P., Payer, K., Grossman, A.D., Kirschner, M.W., et al. (2010). Using buoyant mass to measure the growth of single cells. Nature Methods *7*, 387–390.

23. Wang, P., Robert, L., Pelletier, J., Dang, W.L., Taddei, F., Wright, A., and Jun, S. (2010). Robust growth of *Escherichia coli*. Current Biology *20*, 1099–1103.

24. Cermak, N., Olcum, S., Delgado, F.F., Wasserman, S.C., Payer, K.R., Murakami, M.A., Knudsen, S.M., Kimmerling, R.J., Stevens, M.M., Kikuchi, Y., et al. (2016). High-throughput measurement of single-cell growth rates using serial microfluidic mass sensor arrays. Nature Biotechnology *34*, 1052–1059.

25. Iyer-Biswas, S., Wright, C.S., Henry, J.T., Lo, K., Burov, S., Lin, Y., Crooks, G.E., Crosson, S., Dinner, A.R., and Scherer, N.F. (2014). Scaling laws governing stochastic growth and division of single bacterial cells. Proceedings of the National Academy of Sciences *111*, 15912–15917.

26. Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099–1102.

27. Lin, J. and Amir, A. (2018). Homeostasis of protein and mRNA concentrations in growing cells. Nature Communications *9*, 1–11.

968 28. Metzl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y., and Barkai, N. (2017). Prin-
969 ciples of cellular resource allocation revealed by condition-dependent proteome profiling.
970 eLife *6*, e28034.

971 29. Kafri, M., Metzl-Raz, E., Jonas, F., and Barkai, N. (2016). Rethinking cell growth
972 models. FEMS Yeast Research *16*.

973 30. Logsdon, M.M., Ho, P.Y., Papavinasasundaram, K., Richardson, K., Cokol, M., Sassetti,
974 C.M., Amir, A., and Aldridge, B.B. (2017). A parallel adder coordinates mycobacterial
975 cell-cycle progression and cell-size homeostasis in the context of asymmetric growth and
976 organization. Current Biology *27*, 3367–3374.

977 31. Messelink, J., Meyer, F., Bramkamp, M., and Broedersz, C. (2020). Single-cell growth
978 inference of *Corynebacterium glutamicum* reveals asymptotically linear growth. bioRxiv,
979 10.1101/2020.05.25.115055 .

980 32. Tiruvadi Krishnan, S., Männik, J., Kar, P., Lin, J., Amir, A., and Männik, J. (2021).
981 Replication-related control over cell division in *Escherichia coli* is growth-rate dependent.
982 bioRxiv, 10.1101/2021.02.18.431686 .

983 33. Furchtgott, L., Wingreen, N.S., and Huang, K.C. (2011). Mechanisms for maintaining
984 cell shape in rod-shaped Gram-negative bacteria. Molecular Microbiology *81*, 340–353.

985 34. Adiciptaningrum, A., Osella, M., Moolman, M.C., Lagomarsino, M.C., and Tans, S.J.
986 (2015). Stochasticity and homeostasis in the *E. coli* replication and division cycle.
987 Scientific Reports *5*, 1–8.

988 35. Stein, W.E. and Dattero, R. (1985). Sampling bias and the inspection paradox. Mathe-
989 matics Magazine *58*, 96–99.

990 36. Mitchison, J. (2005). Single cell studies of the cell cycle and some models. Theoretical
991 Biology and Medical Modelling *2*, 4.

992 37. Abner, K., Aaviksaar, T., Adamberg, K., and Vilu, R. (2014). Single-cell model of
993 prokaryotic cell cycle. Journal of Theoretical Biology *341*, 78–87.

994 38. Deforet, M., Van Ditmarsch, D., and Xavier, J.B. (2015). Cell-size homeostasis and the
995 incremental rule in a bacterial pathogen. Biophysical Journal *109*, 521–528.

996 39. Nordholt, N., van Heerden, J.H., and Bruggeman, F.J. (2020). Biphasic cell-size and
997 growth-rate homeostasis by single *Bacillus subtilis* cells. Current Biology *30*, 2238–2247.

998 40. Knapp, B.D., Odermatt, P., Rojas, E.R., Cheng, W., He, X., Huang, K.C., and Chang, F.
999 (2019). Decoupling of rates of protein synthesis from cell expansion leads to supergrowth.
1000 Cell Systems *9*, 434–445.

1001 41. Reshes, G., Vanounou, S., Fishov, I., and Feingold, M. (2008). Cell shape dynamics in
1002 *Escherichia coli*. Biophysical Journal *94*, 251–264.

1003 42. Banerjee, S., Lo, K., Daddysman, M.K., Selewa, A., Kuntz, T., Dinner, A.R., and
1004 Scherer, N.F. (2017). Biphasic growth dynamics control cell division in *Caulobacter*
1005 *crescentus*. Nature Microbiology *2*, 1–6.

1006 43. Cadart, C., Venkova, L., Recho, P., Lagomarsino, M.C., and Piel, M. (2019). The physics
1007 of cell-size regulation across timescales. Nature Physics *15*, 993–1004.

1008 44. Panlilio, M., Grilli, J., Tallarico, G., Iuliani, I., Sclavi, B., Cicuta, P., and Lagomarsino,
1009 M.C. (2021). Threshold accumulation of a constitutive protein explains *E. coli* cell-
1010 division behavior in nutrient upshifts. Proceedings of the National Academy of Sciences
1011 *118*.

45. Willis, L. and Huang, K.C. (2017). Sizing up the bacterial cell cycle. Nature Reviews Microbiology *15*, 606–620.

46. Simpson, E.H. (1951). The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society: Series B (Methodological) *13*, 238–241.

47. Aldridge, B.B., Fernandez-Suarez, M., Heller, D., Ambravaneswaran, V., Irimia, D., Toner, M., and Fortune, S.M. (2012). Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. Science *335*, 100–104.

48. Kubitschek, H. (1981). Bilinear cell growth of *Escherichia coli*. Journal of Bacteriology *148*, 730–733.

49. Oldewurtel, E.R., Kitahara, Y., Cordier, B., Özbaykal, G., and van Teeffelen, S. (2019). Bacteria control cell volume by coupling cell-surface expansion to dry-mass growth. bioRxiv, https://doi.org/10.1101/769786 .

50. Datsenko, K.A. and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences *97*, 6640–6645.

51. Thomason, L.C., Costantino, N., and Court, D.L. (2007). *E. coli* genome manipulation by P1 transduction. Current Protocols in Molecular Biology *79*, 1–17.

52. Reyes-Lamothe, R., Sherratt, D.J., and Leake, M.C. (2010). Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. Science *328*, 498–501.

53. Cherepanov, P.P. and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9–14.

1034 54. Moolman, M.C., Krishnan, S.T., Kerssemakers, J.W., Van Den Berg, A., Tulinski, P.,
1035 Depken, M., Reyes-Lamothe, R., Sherratt, D.J., and Dekker, N.H. (2014). Slow unload-
1036 ing leads to DNA-bound β 2-sliding clamp accumulation in live *Escherichia coli* cells.
1037 Nature Communications *5*, 5820.

1038 55. Yang, D., Jennings, A.D., Borrego, E., Retterer, S.T., and Männik, J. (2018). Analysis
1039 of factors limiting bacterial growth in PDMS mother machine devices. Frontiers in
1040 Microbiology *9*, 871.

1041 56. Pazos, M., Peters, K., Casanova, M., Palacios, P., VanNieuwenhze, M., Breukink, E.,
1042 Vicente, M., and Vollmer, W. (2018). Z-ring membrane anchors associate with cell wall
1043 synthases to initiate bacterial cell division. Nature Communications *9*, 1–12.

1044 57. Kennard, A.S., Osella, M., Javer, A., Grilli, J., Nghe, P., Tans, S.J., Cicuta, P., and
1045 Lagomarsino, M.C. (2016). Individuality and universality in the growth-division laws of
1046 single *E. coli* cells. Physical Review E *93*, 012408.

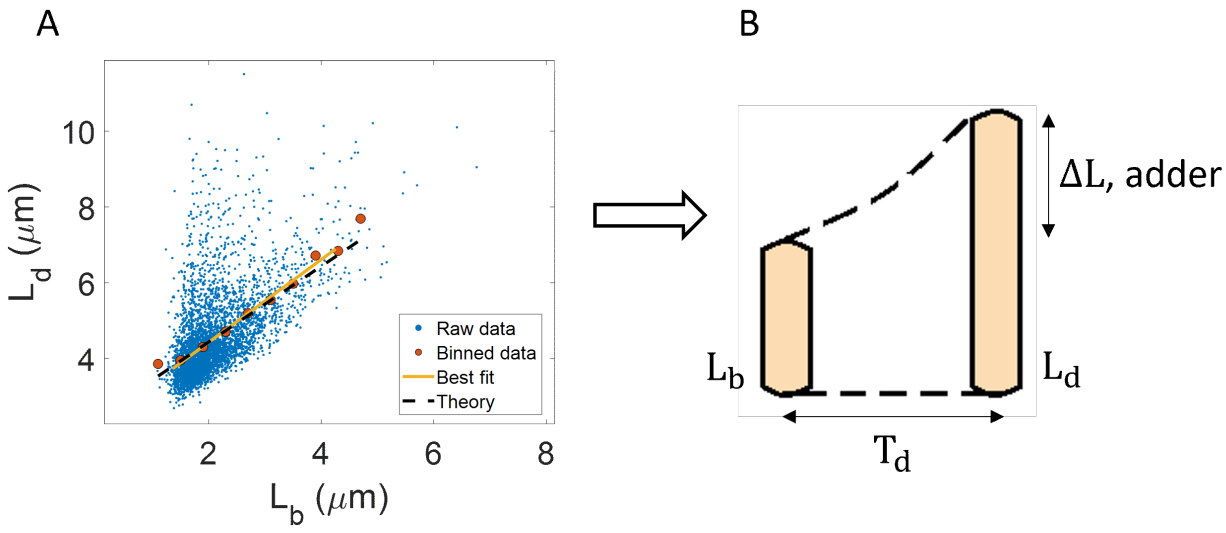


Figure 1: **Utility of binning and linear regression:** **A**. Length at division (L_d) vs length at birth (L_b) is plotted using data obtained by Tanouchi *et al.* [10]. Raw data is shown as blue dots. We find the trend in binned data (red) to be linear with the underlying best linear fit (yellow) following the equation, $L_d = 1.09L_b + 2.24\mu m$. This is close to the adder behavior with an underlying equation given by $L_d = L_b + \Delta L$, where ΔL is the mean size added between birth and division (shown as black dashed line). **B**. A schematic of the adder mechanism is shown where the cell grows over its generation time (T_d) and divides after addition of length ΔL from birth. This ensures cell size homeostasis in single cells.

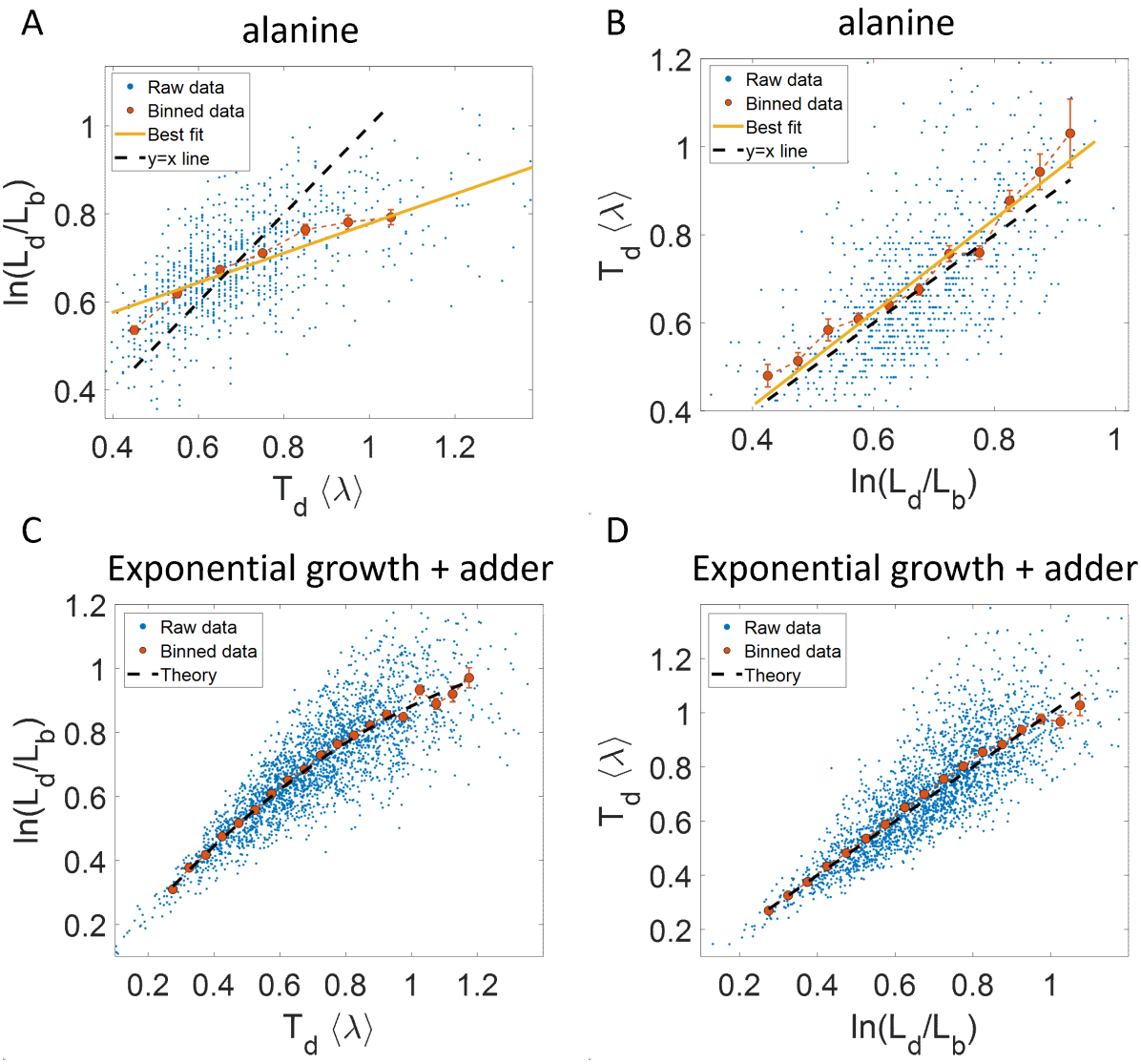


Figure 2: **Plots that could potentially lead to misinterpreting exponential growth: A, B**. Data is obtained from experiments in M9 alanine medium ($\langle T_d \rangle$ = 214 min, N = 816 cells). **A**. $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot is shown. The blue dots are the raw data, the red correspond to the binned data trend, the yellow line is the best linear fit obtained by performing linear regression on the raw data and the black dashed line is the y=x line. *A priori*, non-linear trend in binned data might point to growth being non-exponential. **B**. $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot is shown for the same experiments. **C, D**. Simulations of exponentially growing cells following the adder model are carried out for N = 2500 cells. The parameters used are provided in Section 5.11.2. **C**. $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot is shown. The trend in binned data shown in red is non-linear. The black dashed line is the expected trend obtained from theory (Equation 2). For parameters used in the simulations here, the black dashed line follows $\ln(\frac{L_d}{L_b}) = 1.26\langle\lambda\rangle T_d - 0.38(\langle\lambda\rangle T_d)^2$. **D**. $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot is shown with binned data in red closely following the expected trend of y=x line (black dashed line). In all of these plots, the binned data is shown only for those bins with more than 15 data points in them.

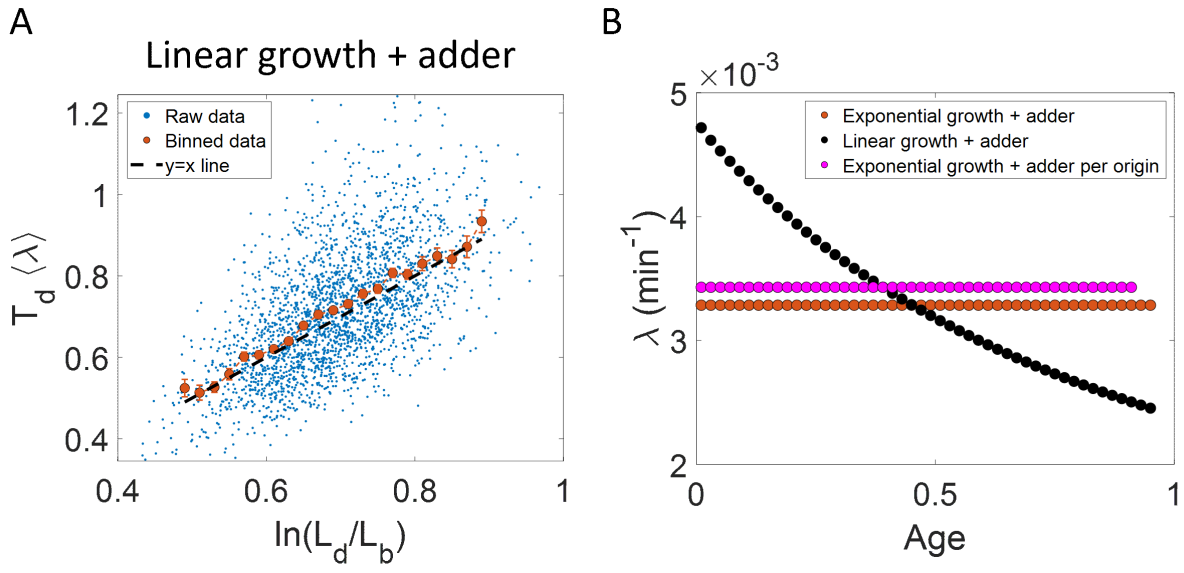


Figure 3: **Differentiating linear growth from exponential growth: A**. $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot is shown for simulations of linearly growing cells following the adder model for N = 2500 cell cycles. The binned data closely follows the y=x trend which could be incorrectly interpreted as cells undergoing exponential growth. **B**. The binned data trend for growth rate vs age plot is shown in red for simulations of N= 2500 cell cycles of exponentially growing cells following the adder model. We observe the trend to be nearly constant as expected for exponential growth. Since the growth rate is fixed at the beginning of each cell cycle in the above simulations, we do not show error bars for each bin within the cell cycle. Also shown in black is the growth rate vs age plot for simulations of N= 2500 cell cycles of linearly growing cells following the adder model. As expected for linear growth, the binned growth rate decreases with age. The binned growth rate trend is also found to be nearly constant for the simulations of exponentially growing cells following the adder per origin model (shown in magenta). Thus, the plot growth rate vs age provides a consistent method to identify the mode of growth. Parameters used in the above simulations of exponential and linear growth are derived from the experimental data in alanine medium. Details are provided in the Section 5.11.2.

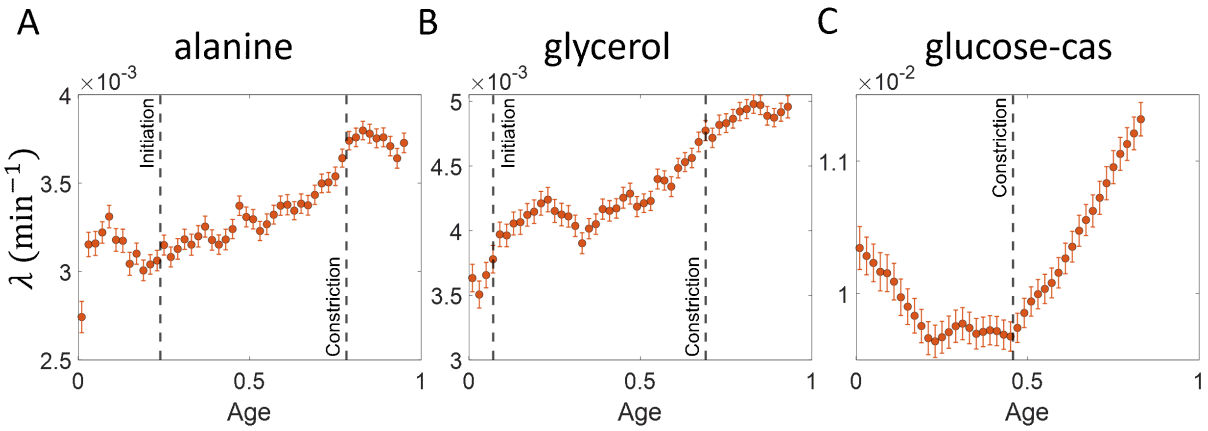


Figure 4: **Growth rate vs age obtained from experiments:** Growth rate vs age plots are shown for *E. coli* experimental data. The red dots correspond to the binned data trends showing the variation in growth rate. The medium in which the experiments were conducted are **A**. Alanine ($\langle T_d \rangle$ = 214 min) **B**. Glycerol ($\langle T_d \rangle$ = 164 min) **C**. Glucose-cas ($\langle T_d \rangle$ = 65 min). The error bars show the standard deviation of the growth rate in each bin scaled by $\frac{1}{\sqrt{N}}$, where N is the number of cells in that bin. The dashed vertical lines mark the age at initiation of DNA replication (left line) and the start of septum formation (right line). In case of glucose-cas, the initiation age is not marked as it occurs in the mother cell.

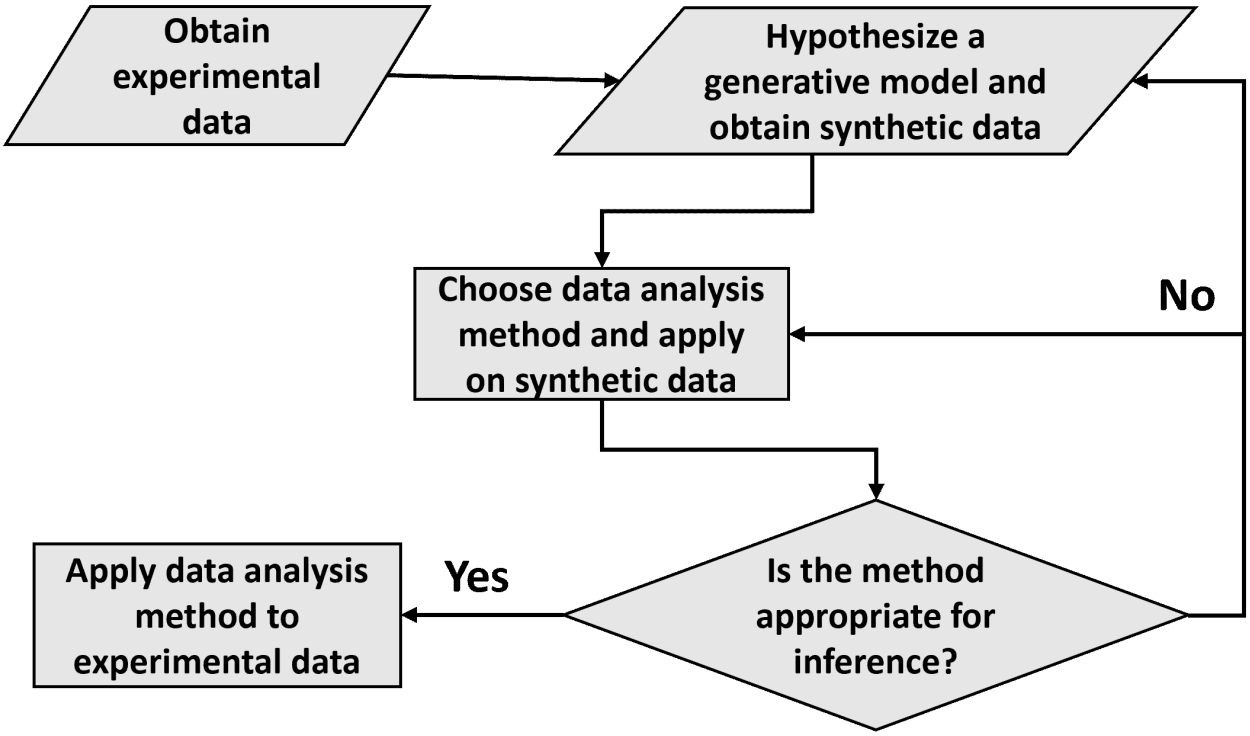


Figure 5: A flowchart of the general framework proposed in the paper to carry out data analysis.

8 Supplementary Figures and Tables

Table S1: Variable definitions.

Variables	**Description**
L_b	Length of the cell at birth and also a proxy for size at birth
L_d	Length of the cell at division and also a proxy for size at division
l_b	$\frac{L_b}{\langle L_b \rangle}$, where $\langle L_b \rangle$ is mean size at birth
l_d	$\frac{L_d}{\langle L_b \rangle}$, where $\langle L_b \rangle$ is mean size at birth
f(l_b)	Mathematical function which captures the regulation strategy determining division given size at birth. f(l_b) $= 2l_b^{1-\alpha}$
T_d	Generation time
σ_t	Standard deviation of generation time
x_n or x	$x_n = \ln(l_b^n)$. Since $l_b \approx 1$, $x_n \approx l_b^n - 1$
σ_x	Standard deviation of x_n
$f_1(x_n)$	Gaussian describing the distribution of x_n. $f_1(x_n) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{x_n^2}{2\sigma_x^2}\right)$
$\langle \lambda \rangle$	Mean growth rate
CV_λ	Coefficient of variation of growth rate
$\xi(0, CV_\lambda)$	Normally distributed growth rate noise. Growth rate is defined as $\lambda = \langle \lambda \rangle + \langle \lambda \rangle \xi(0, CV_\lambda)$
$f_2(\xi)$	Gaussian describing the distribution of random variable $\xi(0, CV_\lambda)$. $f_2(\xi) = \frac{1}{\sqrt{2\pi CV_\lambda^2}} \exp\left(-\frac{\xi^2}{2CV_\lambda^2}\right)$
$\frac{\zeta(0,\sigma_n)}{\langle \lambda \rangle}$	Normally distributed time additive division timing noise with mean 0 and standard deviation $\frac{\sigma_n}{\langle \lambda \rangle}$

$f_3(\zeta)$	Gaussian describing the distribution of random variable $\zeta(0, \sigma_n)$. $f_3(\zeta) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{\zeta^2}{2\sigma_n^2}\right)$
$\zeta_s(0, \sigma_{bd})$	Normally distributed size additive division timing noise with mean 0 and standard deviation σ_{bd}
σ_l	Standard deviation of $\ln(\frac{L_d}{L_b})$
$f_4\left(\ln(\frac{L_d}{L_b})\right)$	Gaussian describing the distribution of $\ln(\frac{L_d}{L_b})$. $f_4\left(\ln(\frac{L_d}{L_b})\right) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{\left(\ln(\frac{L_d}{L_b})-\ln(2)\right)^2}{2\sigma_l^2}\right)$
ρ_{exp}	Correlation coefficient of the pair $(\ln(\frac{L_d}{L_b}), \langle\lambda\rangle T_d)$
m_{tl}	Slope of the best linear fit for $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot
c_{tl}	Intercept of the best linear fit for $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot
m_{lt}	Slope of the best linear fit for $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot
c_{lt}	Intercept of the best linear fit for $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot
$\langle\lambda_{lin}\rangle$	Mean normalized elongation speed
$CV_{\lambda,lin}$	Coefficient of variation of normalized elongation speed
$\xi_{lin}(0, CV_{\lambda,lin})$	Normally distributed normalized elongation speed noise. Normalized elongation speed is defined as $\lambda_{lin} = \langle\lambda_{lin}\rangle + \langle\lambda_{lin}\rangle\xi_{lin}(0, CV_{\lambda,lin})$
$\sigma_{l,lin}$	Standard deviation of $l_d - l_b$
ρ_{lin}	Correlation coefficient of the pair $(l_d - l_b, \langle\lambda_{lin}\rangle T_d)$
$m_{tl,lin}$	Slope of the best linear fit for $l_d - l_b$ vs $\langle\lambda_{lin}\rangle T_d$ plot
$c_{tl,lin}$	Intercept of the best linear fit for $l_d - l_b$ vs $\langle\lambda_{lin}\rangle T_d$ plot
$m_{lt,lin}$	Slope of the best linear fit for $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot
$c_{lt,lin}$	Intercept of the best linear fit for $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot
L_i	Cell size at the start of DNA replication (initiation)

$L_i^{tot,next}$	Total cell size of the daughter cells at the start of DNA replication
Δ_{ii}	Size added per origin between initiations
O	Number of origins just after initiation
C+D	Time between initiation and division
T_n	Timing of start of septum formation/onset of constriction
L_n	Cell size at time T_n

Table S2: The slope and the intercept of the best linear fit along with their 95% confidence intervals (CI) obtained on performing linear regression on experimental data. The data is collected for cells growing in M9 alanine, glycerol and glucose-cas media [32].

Media	**No. of cells**	**T_d (min)**	**$\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plot**		**$\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot**	
			Slope (with 95% CI)	Intercept (with 95% CI)	Slope (with 95% CI)	Intercept (with 95% CI)
Alanine	816	214	0.34 (0.31, 0.36)	0.44 (0.42, 0.46)	1.06 (0.98, 1.14)	-0.01 (-0.07, 0.04)
Glycerol	648	164	0.34 (0.32, 0.37)	0.43 (0.41, 0.44)	1.26 (1.16, 1.35)	-0.13 (-0.20, -0.07)
Glucose-cas	737	65	0.31 (0.28, 0.34)	0.42 (0.40, 0.44)	0.91 (0.83, 1.00)	0.09 (0.03, 0.15)

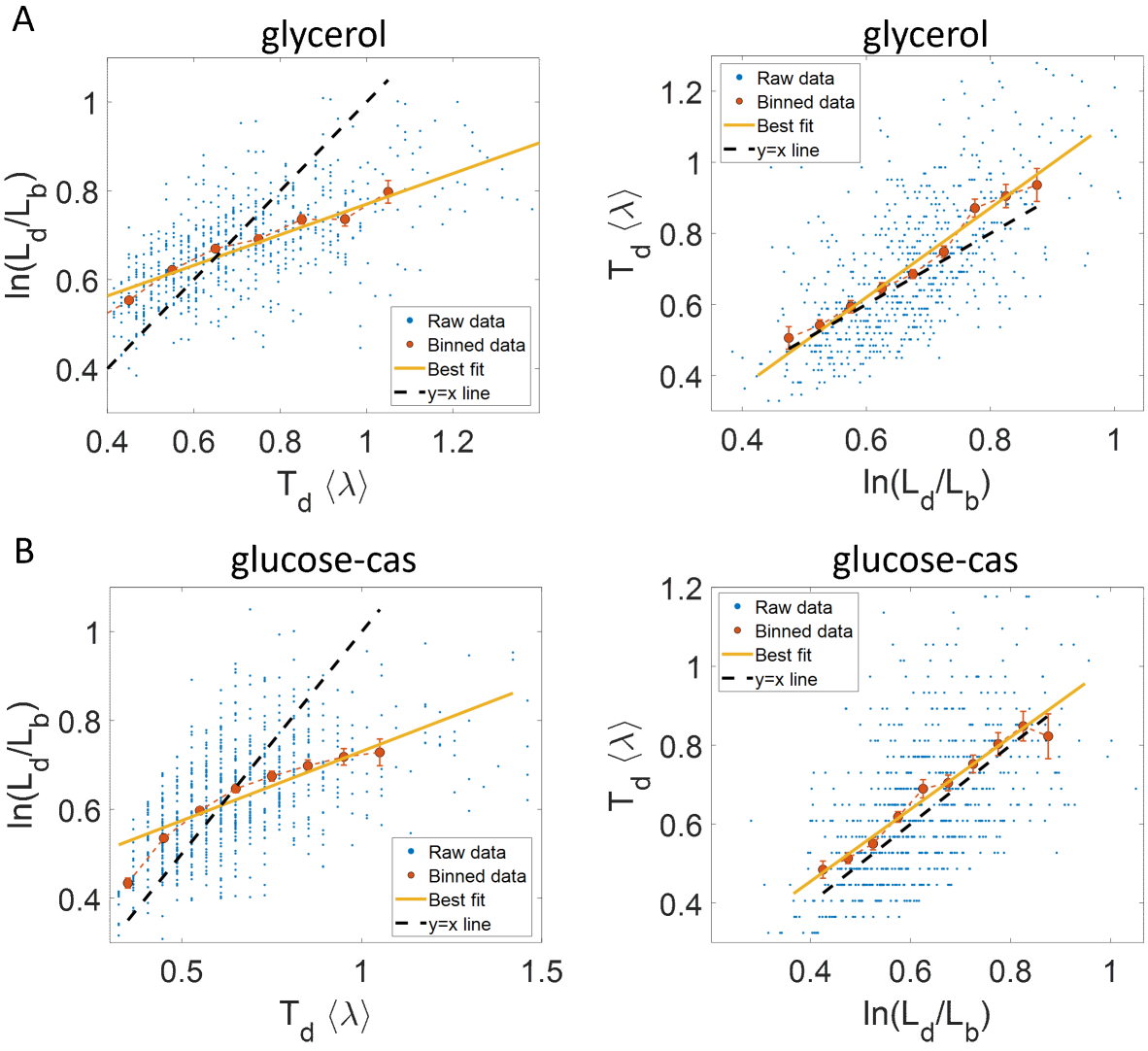


Figure 2- figure supplement 1: **Experimental data:** $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ (left) and $\langle\lambda\rangle T_d$ vs $\ln(\frac{L_d}{L_b})$ plot (right) is shown for, **A**. Cells growing in glycerol medium ($\langle T_d\rangle$ = 164 min, N = 648 cells). **B**. Cells growing in glucose-cas medium ($\langle T_d\rangle$ = 65 min, N = 737 cells). Binned data (red), and the best linear fit (yellow) obtained by performing linear regression on the raw data deviate from the y=x line (black dashed line) in the case of $\ln(\frac{L_d}{L_b})$ vs $\langle\lambda\rangle T_d$ plots in both media. However, both binned data and the best linear fit are in close agreement with the y=x line (black dashed line) on interchanging the axes. In all of these plots, the binned data is shown only for those bins with more than 15 data points in them.

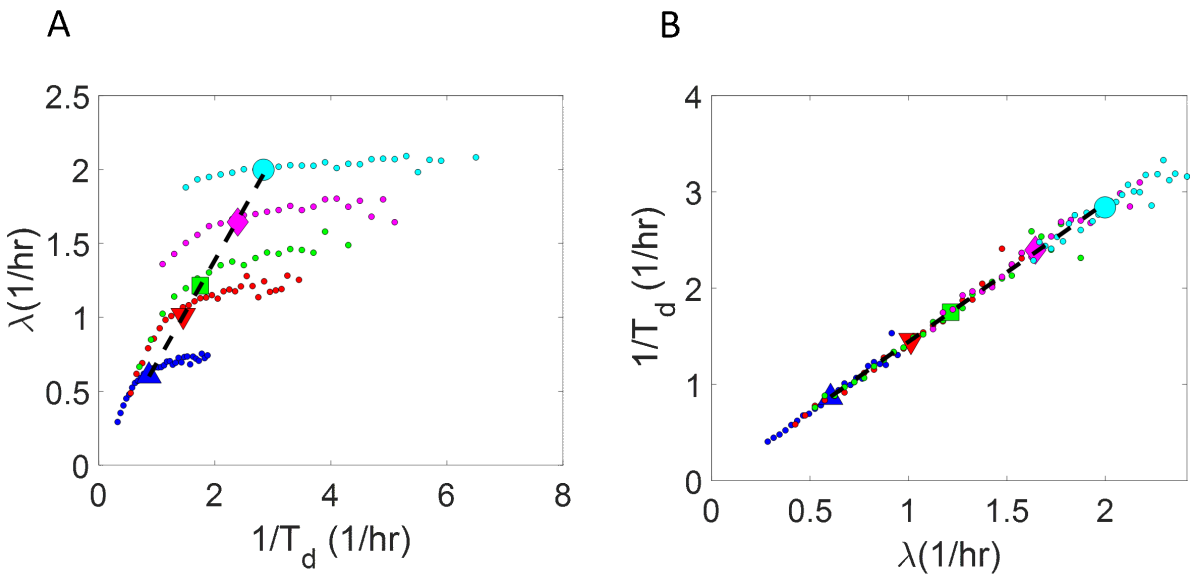


Figure 2- figure supplement 2: **Binned data trend in growth rate (λ) and inverse generation time ($\frac{1}{T_d}$) plots: A-B**. Simulations of the adder model for exponentially growing cells were carried out at multiple growth rates for N = 2500 cells. The size added between birth and division and the mean growth rates were extracted from Kennard *et al.*, [57]. The CV of growth rates was greater for cells growing in slower-growth media. See Section 5.11.2 for the parameter values. For these simulations, we show **A**. λ vs $\frac{1}{T_d}$ plot. **B**. $\frac{1}{T_d}$ vs λ plot. The smaller circles show the trend in binned data within a growth medium. Different colors correspond to different growth media. Population means are shown as larger markers. The population means agree with the expected y=ln(2)x line (black dashed line) in Figure 2- figure supplement 2A but the trend within a single growth medium is non-linear and deviates from the y=ln(2)x line. However, in Figure 2- figure supplement 2B, population means across growth conditions and the trend in binned data within a single growth medium follow the expected y=$\frac{1}{\ln(2)}$x line (black dashed line).

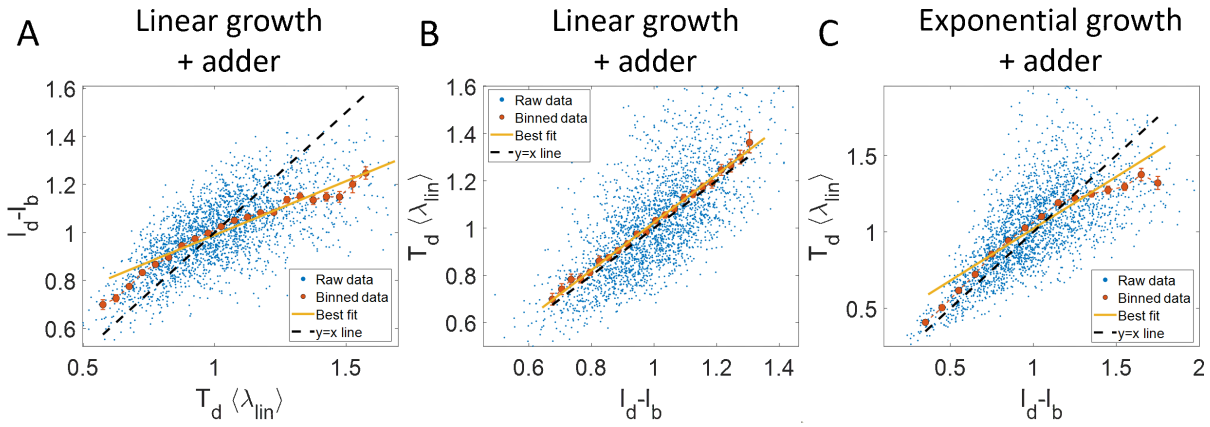


Figure 3- figure supplement 1: **Predicting statistics based on a model of linear growth: A-B**. Simulations of linearly growing cells following the adder model are carried out for N = 2500 cell cycles. **A**. $l_d - l_b$ vs $\langle\lambda_{lin}\rangle T_d$ plot is shown. The raw data is shown as blue dots. The binned data (in red) and the best linear fit (in yellow) deviate from the y=x line (black dashed line). Such a deviation can be predicted based on a model as discussed in detail in Section 5.5. **B**. $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot is shown. The binned data (in red) and the best linear fit (in yellow) agree with the y=x line (in black). **C**. Simulations of exponentially growing cells following the adder model are carried out for N = 2500 cell cycles. $\langle\lambda_{lin}\rangle T_d$ vs $l_d - l_b$ plot is shown. The binned data (in red) and the best linear fit (in yellow) deviate from the y=x line (in black) as expected for exponential growth. Parameters used in the simulations above are provided in Section 5.11.2.

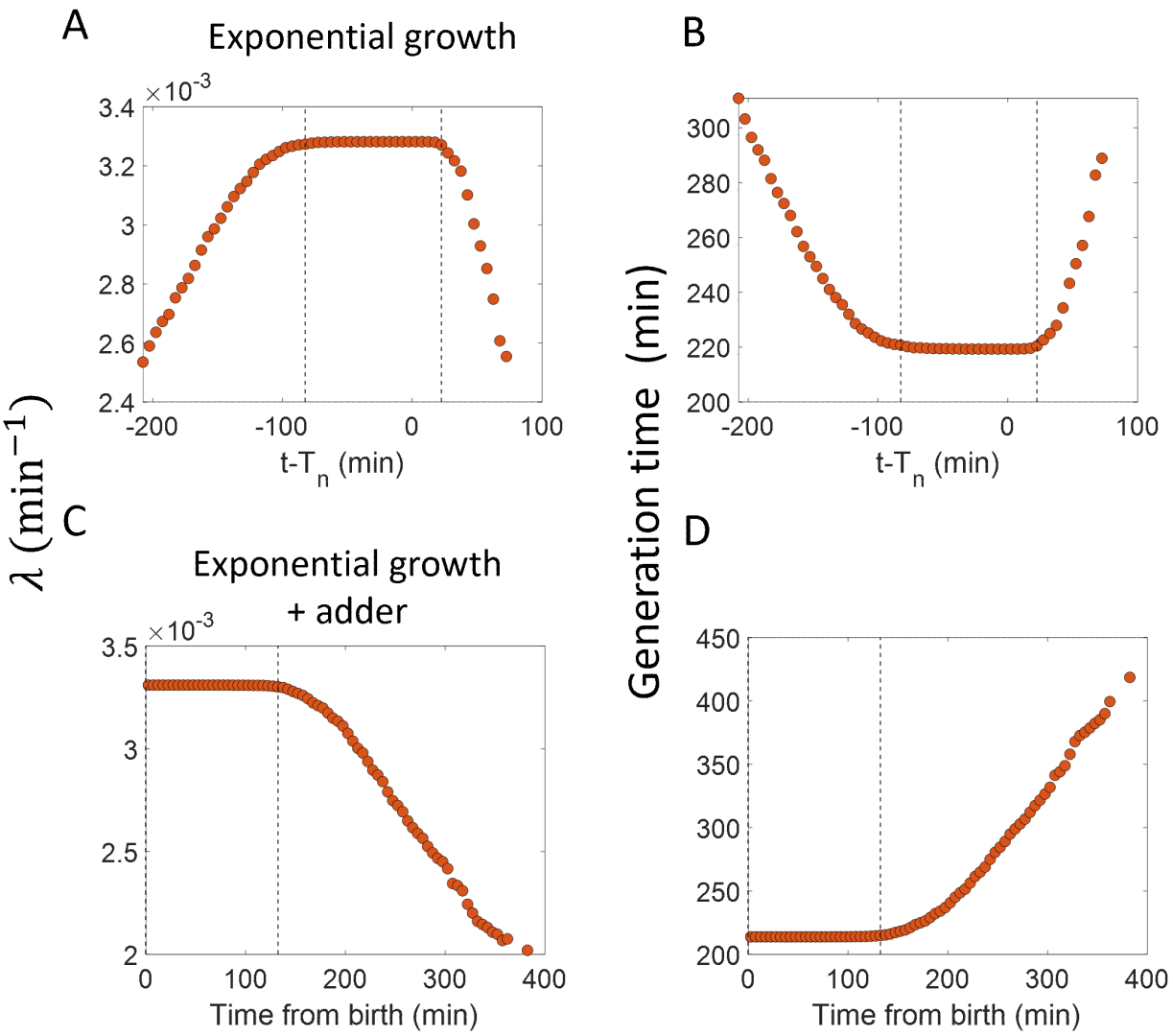


Figure 3- figure supplement 2: **Inspection bias in the growth rate vs time plots obtained from simulations:** **A**. The binned growth rate trend as a function of time from the onset of constriction (t-T_n) is shown in red. Time t-$T_n = 0$ corresponds to onset of constriction. The plot is shown for simulations of exponentially growing cells carried out over N = 2500 cell cycles. Constriction length is determined by a constant length addition from birth and division occurs after a constant length addition from constriction. **B**. The average generation time for the cells present in each bin of Figure 3- figure supplement 2A is shown. **C**. For simulations of exponentially growing cells following the adder model (N=2500), the binned growth rate (in red) vs time from birth plot is shown. **D**. The average generation time for the cells present in each bin of Figure 3- figure supplement 2C is shown. The vertical dashed lines show the time range in which the generation times are approximately constant and hence, the effects of inspection bias are negligible. Within that time range, the growth rate trend is found to be constant, consistent with the assumption of exponential growth.

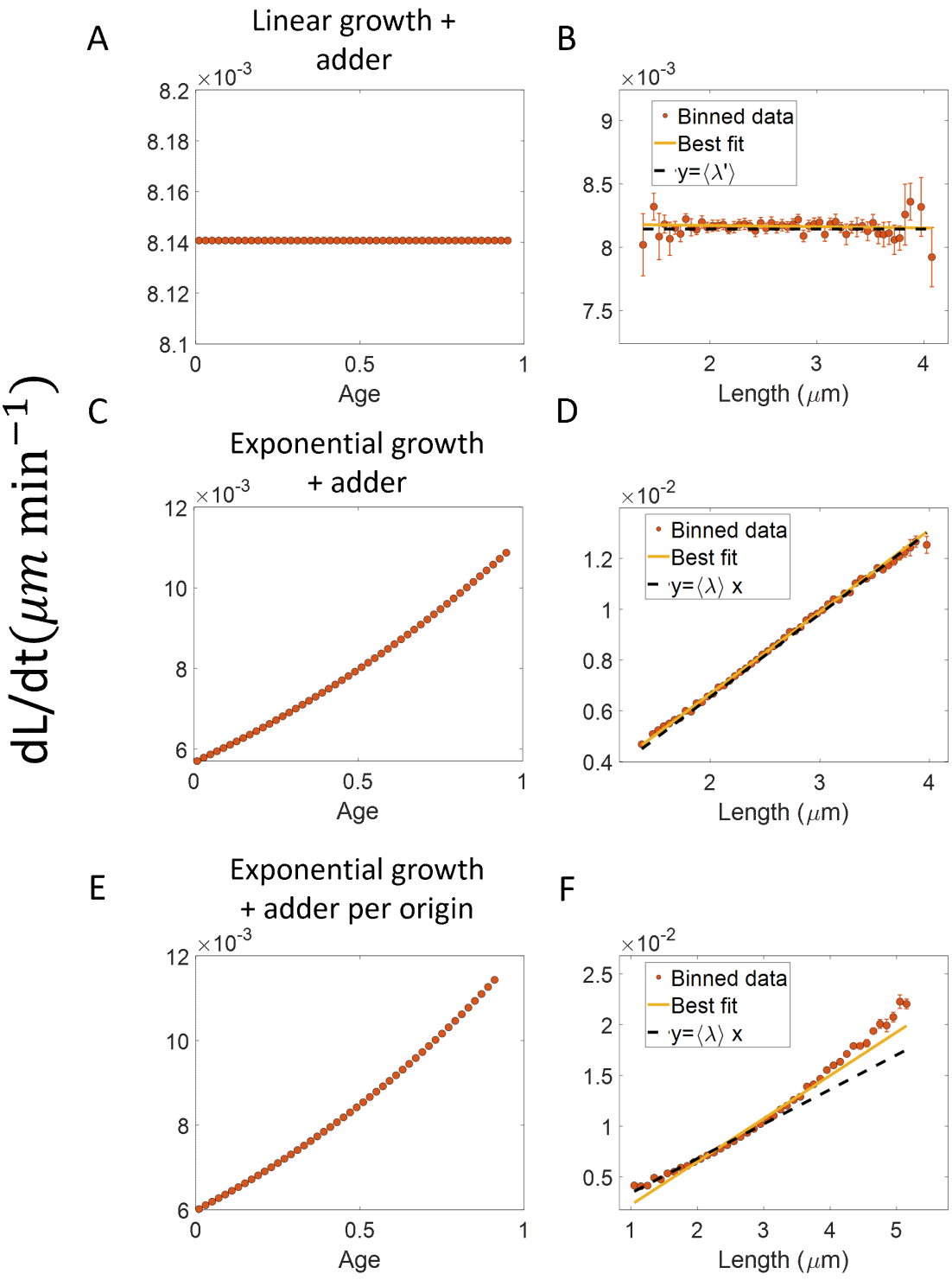


Figure 3- figure supplement 3: **Differential methods of quantifying growth: A-B**. Simulations of linearly growing cells following the adder model are carried out for N = 2500 cell cycles. Cell size (L) data is recorded as a function of time within the cell cycle. **A**. The red dots show the binned data for elongation speed as a function of age. The trend is almost constant in agreement with the linear growth assumption. **B**. Elongation speed is also constant with cell size as expected for linear growth with the intercept value being the average elongation speed. **C-D**. Simulations of exponentially growing cells following the adder model are carried out for N = 2500 cell cycles. **C**. Elongation speed trend (in red) increases with age in agreement with the exponential growth assumption. **D**. Elongation speed trend (in red) increases linearly with size with a slope equal to the average growth rate. **E-F**. Simulations of exponentially growing cells following the adder per origin model are carried out for N = 2500 cell cycles. **E**. Again, the elongation speed trend (in red) increases with age in agreement with the exponential growth assumption. **F**. Elongation speed trend (in red) deviates from the expected linear trend (black dashed line). This could be misinterpreted as non-exponential growth. Thus, we find that the binned data trend for the plot elongation speed vs size is model-dependent.

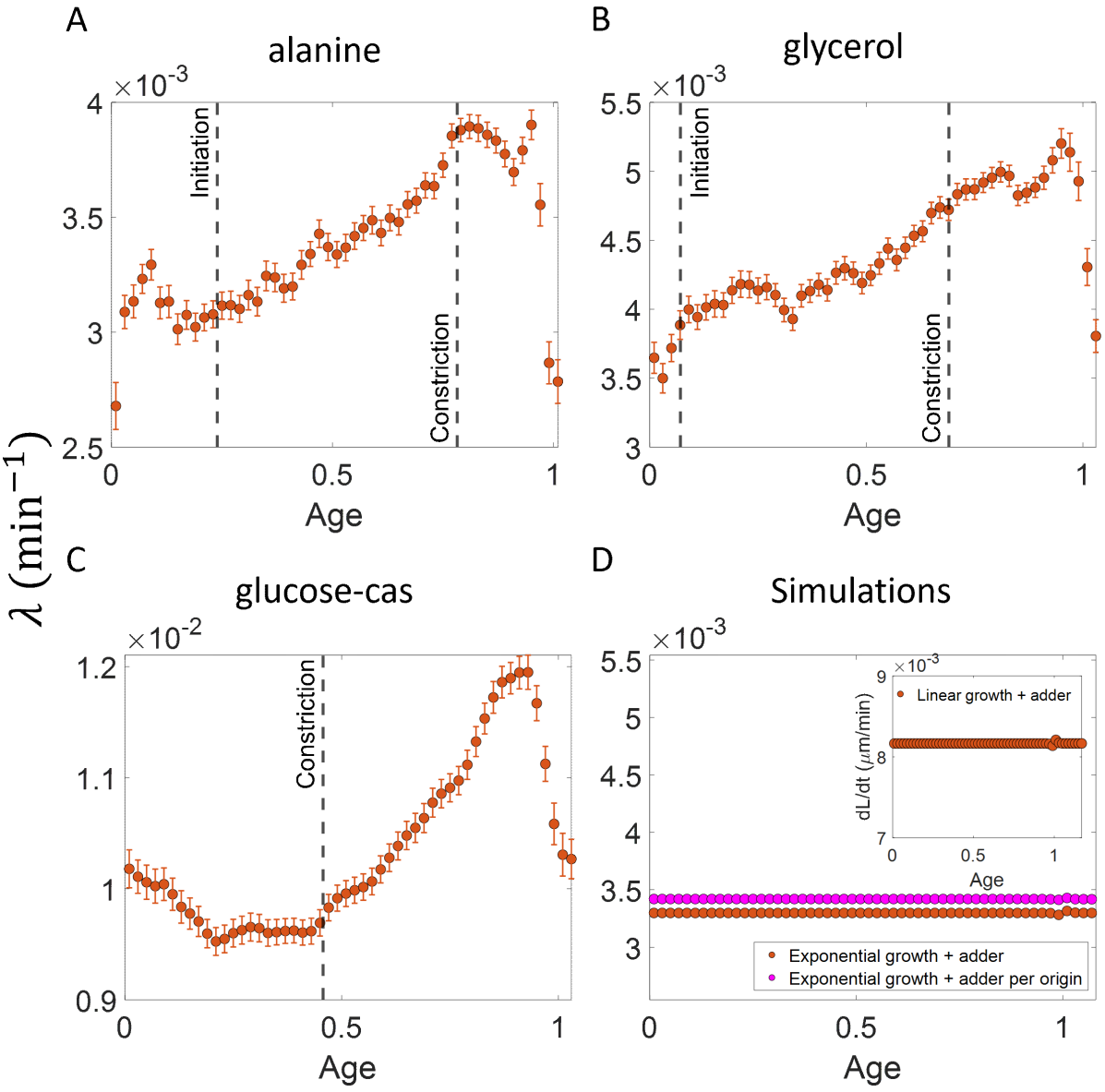
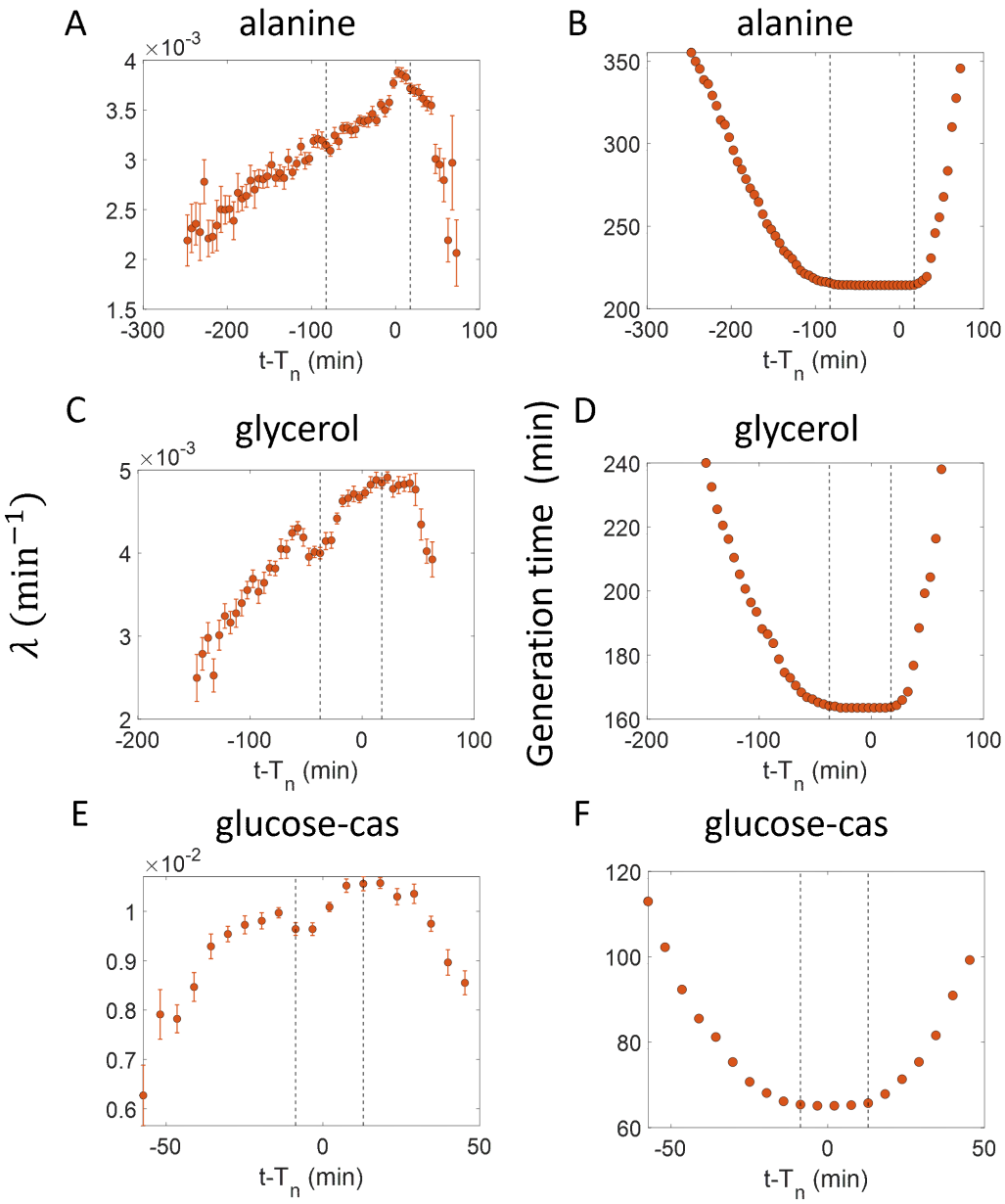


Figure 4- figure supplement 1: **Growth rate vs age curves extended beyond the division event: A,B,C**. The binned growth rate trend is shown in red as a function of age for *E. coli* experimental data. The trends are obtained using the cell size trajectories extending beyond the division event (age>1). The plots are shown for **A**. Alanine medium (N = 720 cells) **B**. Glycerol medium (N = 594 cells). **C**. Glucose-cas medium (N = 664 cells). The error bars in all three plots represent the standard deviation of the growth rate in each bin scaled by $\frac{1}{\sqrt{N}}$, where N is the number of cells in that bin. The growth rate trend appears to be periodic in each of the growth media i.e., λ at age ≈ 1 is close to λ at age ≈ 0. These trends agree with that of Figure 4 in the appropriate age ranges. **D**. Simulations are carried out for N= 2500 cell cycles. The cell size trajectories are collected beyond the division event (age>1). The binned data trend for growth rate vs age plot is shown in red for exponentially growing cells following the adder model. We observe the trend to be nearly constant as expected for exponential growth. The binned growth rate trend is also found to be nearly constant for the simulations of exponential growing cells following the adder per origin model (shown in magenta). (Inset) Shown in red is the elongation speed vs age plot for simulations of N= 2500 cell cycles of linearly growing cells following the adder model. As expected for linear growth, the binned elongation speed trend remains approximately constant with age. The growth rate trends for the models with exponential growth agree with that of Figure 3B. The elongation speed trend (inset) also agrees with the trend in Figure 3- figure supplement 3A.



1162

Figure 4- figure supplement 2: **Inspection bias in the growth rate vs time from constriction plots obtained from experiments: A,C,E**. The binned growth rate trend is shown in red as a function of time from the onset of constriction (t-T_n). Time t-$T_n = 0$ corresponds to the onset of constriction for all cells considered. The plots are shown for **A**. Alanine medium. **C**. Glycerol medium. **E**. Glucose-cas medium. The error bars in all three plots represent the standard deviation of the growth rate in each bin scaled by $\frac{1}{\sqrt{N}}$, where N is the number of cells in that bin. **B,D,F**. The average generation time for the cells present in each bin of **B.** Alanine medium (Figure 4- figure supplement 2A) **D.** Glycerol medium (Figure 4- figure supplement 2C) **F.** Glucose-cas medium (Figure 4- figure supplement 2E) are shown. The vertical dashed lines represent the time range within which the average generation time remains approximately constant. The growth rate trends within this time range are consistent with that in Figure 4 for the respective growth condition as there is negligible inspection bias.