Deciphering the Replication-Division Coordination in *E. coli*: A Unified Mathematical framework for Systematic Model Comparison

Alexandre Perrin[1,*], Marie Doumic[1,†, ¶], Meriem El Karoui[2,‡, ¶], Sylvie Méléard[1,§,¶],

[1] CMAP, Ecole polytechnique, CNRS, Institut polytechnique de Paris, Inria, route de Saclay, 91128 Palaiseau Cedex-France
[2] University of Edinburgh, School of Biological Sciences, Institute of Cell Biology, UK

*alexandre.perrin@polytechnique.edu
† marie.doumic@polytechnique.edu
‡ meriem.elkaroui@ed.ac.uk
§ sylvie.meleard@polytechnique.edu
¶ These authors contributed equally

Abstract

While significant efforts have been made to model bacterial cell division, few models have incorporated DNA replication into the control of this process. To date, models that attempt to capture the coordination between replication and division cycles are based on fundamentally different assumptions, and yet conflicting results have emerged. As a result, key questions regarding how replication affects cell size at division remain unclear. To address this, we develop in a first part, a robust mathematical framework to study models of coordination of replication and division cycles proposed in the literature. Through theoretical analysis, we highlight necessary and sufficient conditions to apply to replication-agnostic and division-agnostic models to ensure physiologically-coherent behaviors. Then, in a second part, we lead a comprehensive statistical analysis to assess the ability of the models to reproduce the division volume distribution conditioned on cell covariates. This in-depth analysis highlighted remarkable performances of a novel model, yielding promising results for future refinements toward a universal model of replication-division coordination in *E. coli*.

Introduction

The regulation of the cell cycle in *E. coli* has been a subject of extensive research for over half a century. A long-standing intuition has been to view this cycle as governed by cellular clocks that reset at the beginning of each cycle, effectively "forgetting" past events [1, 2, 3]. With the advent of advanced microscopy techniques, single-cell observations have revealed significant variability in both division size and inter-generation time between individual isogenic cells, highlighting the stochastic nature of these cellular clocks [4, 5, 6, 7]. This stochasticity has challenged the deterministic approaches historically used to describe cell division in *E. coli* [8], necessitating alternative frameworks that better capture its regulation. In line with this intuition of resetting cellular clocks, the stochastic modeling of cell cycles has relied on Markov chain frameworks. Particularly in *E. coli*, division volume has been assumed to be a function of the previous division volume and independent

random variables. From this approach, the adder model for cell division emerged, identifying the independent random variable not as division size (sizer model) or inter-division time (timer model), but as the volume added between birth and division [9, 10, 11, 12].

In parallel to these developments, fundamental questions have arisen regarding the control of chromosome replication initiation. Initially predicted by [13], it was later demonstrated that replication initiation events are tightly linked to the cell volume per replication origin [14, 5, 15]. Similar to division cycle modeling, replication initiation cycles have been described using a Markov chain approach, where the key variable is the volume per origin of replication at the time of initiation. An initiation adder model has then been proposed, which assumes that the independent random variable driving the Markov chain is the volume added per replication origin between successive initiation events [16, 17, 18].

Then, a fundamental challenge has been to understand how division and replication initiation cycles coordinate, particularly given that cells must ensure an exact distribution of genetic material to their progeny. This problem is especially intricate in bacteria such as *E. coli*, where replication cycles may undergo complex dynamics across growth conditions. Indeed, whereas under slow growth conditions, the cell replicates its chromosome between birth and division analogously to the eukaryotic cell cycle, under faster growth conditions, replication may initiate in the mother cell or in the grandmother cell. The definition of the generic C+D period, which spans from replication initiation to the division that separates the resulting sister chromosomes, has allowed to describe this phenomenon [8] (see Fig 1). When replication initiation takes place in earlier generations, C+D periods overlap, challenging our understanding of the coordination of replication and division cycles. While numerous molecular studies have provided mechanistic insights into this coordination [19, 20, 21, 22, 23], various theoretical models have been developed over the past decade to bridge between division cycles and replication cycles modeling. The objective of these models, systematically formulated as Markov chains, was to identify the key independent random variables governing this coordination. Since most of these models assume that replication initiation cycles are governed solely by the initiation adder, their main question has been to determine the extent to which replication controls division, if at all. Among these models three categories can be distinguished. The first is a replication-agnostic category of model, which states that division is independent of replication [24, 17, 25]. The second is a division agnostic category of model which states that replication is the only factor that controls division [16, 18, 26]. Finally, the third category of models assumes that division is triggered by the slowest of two independent processes, one linked to the C+D period and the other to the inter-division period. [27, 28, 29, 30, 31].

To date, the most extensive approach to test these models has been to verify whether they recover the correlations observed in the data [27, 28, 17, 18, 32, 29, 31]. While all models proposed so far have been corroborated by some correlations, some have been falsified through this approach. For example, the correlation between initiation volume and division volume has challenged the validity of replication-agnostic models [18, 31]. Another approach for testing these models has been to perturb the cell cycles. In [17], perturbations targeting replication initiation showed minimal impact on division timing, leading the authors to conclude a potential independence between these processes and to support a replication-agnostic model. Conversely, in [30] the duration of chromosome replication was increased, resulting in stronger replication-division coupling. This observation has been interpreted as proof that replication termination is a necessary preliminary step

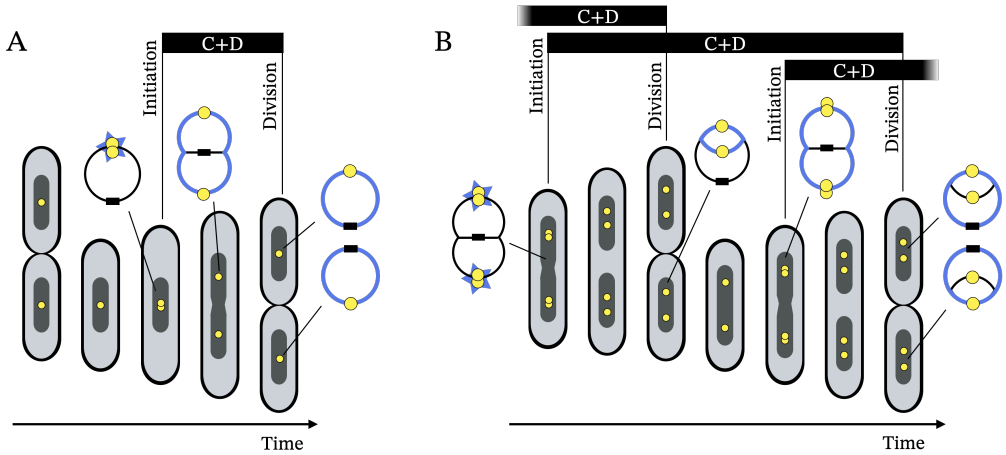


Figure 1: **The replication and division cycles in *E. coli*** (A) In slow growth conditions the replication of the chromosome is framed between the birth and the division. The period spanning from the initiation of replication to the division is called the C+D period. (B) In faster growth conditions, the initiation of replication may occur in previous generations leading to overlaps of C+D periods. The yellow circles depict origins of replication, and sister chromosomes are colored in blue.

to division. Another set of experiments [33, 29] demonstrated that increasing cell width reduced the correlation between replication and division cycles, suggesting that replication alone may not suffice to trigger division, thus challenging division-agnostic models and supporting the existence of an additional parallel regulatory mechanism.

The conflicting findings from recent studies have created uncertainty in the understanding of the mechanisms of replication-division cycles coordination. While data reveal correlations between these two processes, some results challenge their true coupling [17, 32]. Moreover, although complementary results suggest that replication and division are coordinated according to a model from the third category, few results enable to fully characterize the precise mechanisms involved. Adding to this complexity is the variability in observations across growth conditions. For instance, in slow growth conditions, the division adder model fails to describe cell division [15], and replication appears to play a central role while it becomes less preponderant in fast growth conditions [30, 31]. All of these findings call for a consensus on a model capable of capturing the coordination between replication and division cycles across a wide range of growth conditions.

In this paper, we aim to bring clarifications to current uncertainties surrounding the replication-division cycles coordination through rigorous mathematical and statistical analyses of models. In a first part, we introduce a unified mathematical framework for the systematic investigation of these models and study four models: a replication-agnostic model proposed by [17]; a division-agnostic model introduced in [18]; and two third-category models, one proposed by [27]; and a second introduced here for the first time. Within this framework, we derive theoretical results and show the necessary and sufficient conditions on replication-agnostic and division-agnostic models to capture key physiological behaviors in *E. coli*. In a second part, we then conduct a statistical analysis that goes beyond the autocorrelation approach extensively used to date. Specifically, we

compare models based on their ability to reproduce the conditional distribution of division volume, given the birth volume, initiation volume, and cell elongation rate. Through this analysis, we propose a unified comparison method that applies consistent evaluation criteria across models. Importantly, we infer the distribution of the latent variables involved in third-category models, which has never been made to date. Our results support the hypotheses underlying third-category models. Notably, our findings lend support to a new model that reconciles previously conflicting evidence regarding the interdependence between replication and division cycles. Overall, this work provides both a methodology and a model, opening new avenues for the development of a universal model of replication-division coordination.

Results

Unified mathematical framework for replication-division coordination

General assumptions and observed quantities

Multiple mathematical models have been proposed in order to describe the replication-division coordination [16, 27, 28, 17, 18]. In order to compare and to capture the mathematical differences between these models, we introduce a general framework based on assumptions and observed quantities. This framework reproduces the monitoring of a cell lineage in a Mother Machine device [4]. At each generation, a mother cell is randomly replaced by one of its two daughters. In this framework the following assumptions are imposed:

Assumptions:

1. The division is perfectly symmetric, i.e. the volume of the daughter cell at birth is equal to half of the volume of the mother at division.
2. At each replication initiation event, every origin of replication in the cell fires a new round of replication at the same time, thus doubling their number.
3. The origins of replication of a mother cell are equally partitioned into the daughters at division.
4. Over the lineage, the cell volume grows exponentially according to a growth rate function of time $\lambda(t)$, that is positive and locally integrable.

In this framework, two main quantities are studied: the cell volume, and the volume per origin of replication. According to Assumptions 1 and 4 the cell volume trajectory is continuous except at division events, when it is halved. According to Assumptions 2 and 4 the volume per origin of replication is continuous except at initiation events, when it is halved, and remains continuous at division events according to Assumptions 1 and 3. In consequence, cell volume is essentially governed by the growth rate and the division cycles while cell volume per origin is essentially governed by the growth rate and the rounds of replication initiation. The study of these quantities will be the core of the following work since their coupling will directly express the coupling between the division cycles and the replication cycles. We define $(V_b^{(n)})_{n\geq 0}$ the sequence of birth volumes where for all n, $V_b^{(n)}$ represents the birth volume of the mother cell giving rise to a daughter cell with a birth volume $V_b^{(n+1)}$. Identically, we define $(\tilde{V}_i^{(n)})_{n\geq 0}$, the sequence of volumes per

origin of replication immediately after initiations of replication, where for all n, the initiation event associated with $\tilde{V}_i^{(n+1)}$ is subsequent to the initiation event associated with $\tilde{V}_i^{(n)}$ (see Mathematical Supplementary). In a context of a dynamics over time, we additionally introduce $(T_b^{(n)})_{n\geq 0}$ the dates of division giving birth to the cell of volume $V_b^{(n)}$, and $(T_i^{(n)})_{n\geq 0}$, the dates at initiation of replication of the cell with a volume per origin of replication $\tilde{V}_i^{(n)}$. The indexation of these sequences is set in such a way the sister chromosomes obtained from the replication initiated at $T_i^{(n)}$ are separated by the division occurring at $T_b^{(n+1)}$, defining the difference $T_b^{(n+1)} - T_i^{(n)}$ as a C+D period for every n. From the assumptions of the framework, and based on a particular assumption over the initialization of the sequences, one can show the following expression (see Mathematical Supplementary, Lemma 1):

$$\forall n \geq 0, \qquad V_b^{(n+1)} = \tilde{V}_i^{(n)} \exp\left(\int_{T_i^{(n)}}^{T_b^{(n+1)}} \lambda(s)ds\right). \tag{1}$$

As in bacteria multiple C+D periods may overlap [8], this equation provides valuable properties since it connects replication and division cycles without taking into account these overlaps. However, in order to complete the framework we add an indicator of this overlap. A good indicator is the number of origins of replication within the cells, since it directly relates to the number of ongoing (and thus overlapping) C+D period [8, 34]. In this framework, it is possible to follow both the number of replication origins at replication initiation and at birth. Here, we arbitrarily chose to monitor the number of origins of replication within a cell at birth. Let $(\omega^{(n)})_{n\geq 0}$ be the number of origins of replication at the n-th birth. This quantity is defined on the space $\{2^k : k \in \mathbb{Z}\}$, since it is doubled at each initiation event and it is halved at each division. Note that biologically this quantity should be greater or equal to one. However, in our framework, no condition is formulated over the sequences $(T_i^{(n)})_{n\geq 0}$ and $(T_b^{(n)})_{n\geq 0}$ to prevent accumulation of division events over initiation events leading to a number of origins less that one. Within the introduced framework, one can show under the condition that $(T_i^{(n)})_{n\geq 0}$ is non-decreasing that this quantity is essentially linked to the sequences of dates at birth and initiation of replication (see Mathematical Supplementary, Lemma 2). For every $n \geq 0$ and $k \geq -n+1$:

$$\omega^{(n)} = 2^k \quad \Longleftrightarrow \quad T_i^{(n+k-1)} \leq T_b^{(n)} < T_i^{(n+k)}, \tag{2}$$

$$\omega^{(n)} \leq 2^k \quad \Longleftrightarrow \quad T_b^{(n)} < T_i^{(n+k)}. \tag{3}$$

Three key models from the literature

Here we introduce and formulate three models from the literature that have drawn the most interest among the replication agnostic, division agnostic and the third category of model: the Double Independent Adders Model (DIAM), the Replication Double Adders Model (RDAM), the Concurrent Processes Model (CPM).

All these models assume consensually that replication initiation events are triggered according to an initiation adder [5, 18, 17, 16]. The initiation adder states that the initiation of replication is triggered once the cell lineage has accumulated a critical volume per origin of replication since the last initiation (Fig. 2A). This can be formulated mathematically as follows. Let $(\Delta_i^{(n)})_{n\geq 0}$ be an

independent and identically distributed (i.i.d.) sequence of positive random variables that admit a density f_{Δ_i}, modeling the added volume per origin between the n-th and $n+1$-th initiations (Table 1). The sequence $(\tilde{V}_i^{(n)})_{n\geq 0}$ satisfies according to the initiation adder,

$$\forall n \geq 0, \qquad 2\tilde{V}_i^{(n+1)} = \tilde{V}_i^{(n)} + \Delta_i^{(n)}. \tag{4}$$

The **DIAM**, proposed by [17], supposes that the cell cycles are headed by a double adder mechanism, assumed to be independent: the initiation adder and the division adder (Fig. 2B, E). This decoupling between replication and division makes $(V_b^{(n)})_{n\geq 0}$ and $(\tilde{V}_i^{(n)})_{n\geq 0}$ independent processes. Similarly as the initiation adder, one may introduce $(\Delta_d^{(n)})_{n\geq 0}$, the i.i.d. random variables drawn from a density f_{Δ_d}, that model the added volumes over cell generations (Table 1). The couple $(V_b^{(n)}, \tilde{V}_i^{(n)})_{n\geq 0}$ satisfies according to the DIAM:

$$\begin{pmatrix} V_b^{(n+1)} \\ \tilde{V}_i^{(n+1)} \end{pmatrix} = \begin{pmatrix} 2^{-1}(V_b^{(n)} + \Delta_d^{(n)}) \\ 2^{-1}(\tilde{V}_i^{(n)} + \Delta_i^{(n)}) \end{pmatrix}. \tag{5}$$

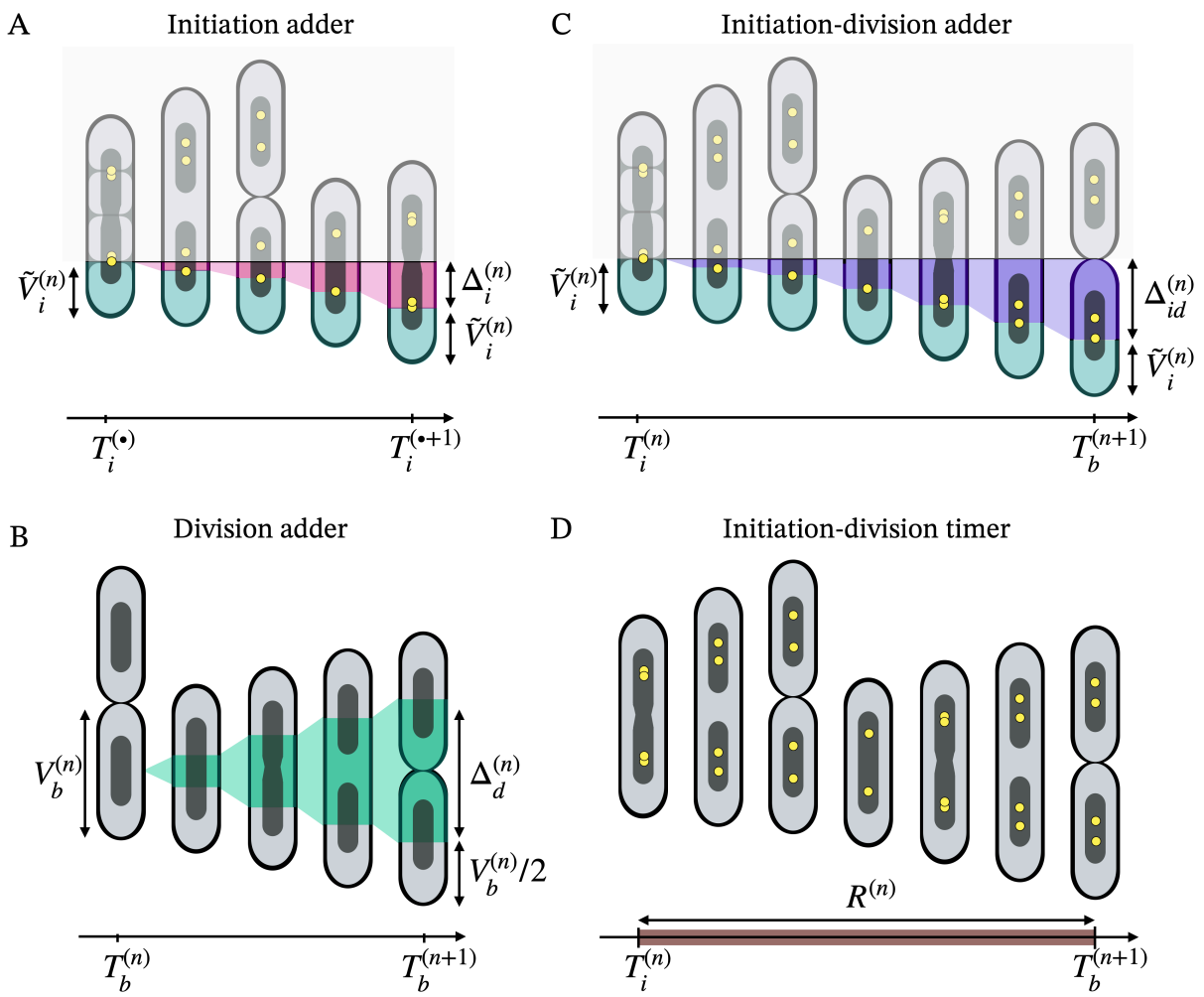
The **RDAM**, first proposed in *Mycobacterium smegmatis* [35], and then adapted in *E. coli* by [18], suggests that replication is central to the control of division. This model assumes an initiation-division adder such that division is triggered once the cell lineage has added a critical volume over the C+D period per origin newly formed at the initiation (hereafter referred to simply as the added volume per origin over the C+D period) (Fig. 2C, E). Let $(\Delta_{id}^{(n)})_{n\geq 0}$ be an i.i.d. sequence of positive random variables of density $f_{\Delta_{id}}$ that models the added volumes over the C+D periods (Table 1). The couple $(V_b^{(n)}, \tilde{V}_i^{(n)})_{n\geq 0}$ satisfies according to the RDAM:

$$\begin{pmatrix} V_b^{(n+1)} \\ \tilde{V}_i^{(n+1)} \end{pmatrix} = \begin{pmatrix} \tilde{V}_i^{(n)} + \Delta_{id}^{(n)} \\ 2^{-1}(\tilde{V}_i^{(n)} + \Delta_i^{(n)}) \end{pmatrix}. \tag{6}$$

The **CPM** proposed by [27, 28] assumes that the division is set by the slowest of two independent concurrent processes: a replication and a division-related processes [27, 29, 30]. According to this model, a cell divides after adding a critical volume since birth and completing chromosome replication and segregation. In this model, replication and segregation are supposed to follow an initiation-division timer (Fig. 2D, E). From the assumptions of the model, we can write the sequence of volume at birth $(V_b^{(n)})_{n\geq 0}$ as the maximum between a timer over the C+D period modeled by the i.i.d. sequence $(R^{(n)})_{n\geq 0}$ drawn from the density f_R, and a division adder modeled by the i.i.d. sequence $(\Delta_d^{(n)})_{n\geq 0}$ drawn from the density f_{Δ_d} (Table 1):

$$\begin{pmatrix} V_b^{(n+1)} \\ \tilde{V}_i^{(n+1)} \end{pmatrix} = \begin{pmatrix} \max\left(\tilde{V}_i^{(n)} e^{\Lambda^{(n)} R^{(n)}}, 2^{-1}(V_b^{(n)} + \Delta_d^{(n)})\right) \\ 2^{-1}(\tilde{V}_i^{(n)} + \Delta_i^{(n)}) \end{pmatrix}, \tag{7}$$

with $\Lambda^{(n)} := \frac{1}{R^{(n)}} \int_{T_i^{(n)}}^{T_i^{(n)}+R^{(n)}} \lambda(s)ds$. As cell elongation has been shown to be exponential (λ constant over time) [9, 36, 37], in the following we will do the approximation that $(\Lambda^{(n)})_{n\geq 0}$ and $(R^{(n)})_{n\geq 0}$ are independent.



E

	Replication control	Division control
DIAM	Initiation adder	Division adder
RDAM	Initiation adder	Initiation-division adder
CPM	Initiation adder	Divison adder AND Initiation-division timer
CAM	Initiation adder	Division adder AND Initiation-division adder

Figure 2: **Summary of the studied models.** (A) The initiation adder model assumes that the initiation of replication event setting the initiation volume per origin $\tilde{V}_i^{(n+1)}$ is triggered once the cell has added a volume per origin of replication $\Delta_i^{(n)}$ assumed to be independent of $\tilde{V}_i^{(n)}$, since the last initiation event. (B) The division adder assumes that the division occurring at $T_b^{(n+1)}$ is set once the cell has added a volume $\Delta_d^{(n)}$ since birth, assumed to be independent to $V_b^{(n)}$. (C) The initiation-division adder assumes that the division occurring at $T_b^{(n+1)}$ is set once the cell has added a volume per origin $\Delta_{id}^{(n)}$ since the beginning of the C+D period, assumed to be independent to $\tilde{V}_i^{(n)}$. (D) The initiation-division timer assumes that the division occurring at $T_b^{(n+1)}$ is set after an amount of time $R^{(n)}$ after the beginning of the C+D period, assumed to be independent to $\tilde{V}_i^{(n)}$. Table (E) displays the control mechanisms of replication cycles and division cycles for each model studied in this work.

Every model introduced so far supposes $(V_b^{(n)}, \tilde{V}_i^{(n)})_{n\geq 0}$ to be an homogeneous Markov chain, with the state of the chain at $n+1$ depending on the state at n, and some i.i.d. random variables. Note that an implicit assumption is made by assuming these random variables to be identically distributed, that is, that the growth conditions are constant and that no aging process is taken into account.

Among the models considered, the DIAM and RDAM both involve two i.i.d. random variables while the CPM incorporates three i.i.d. random variables, introducing an additional degree of complexity. This added degree of complexity has challenged the comparison of these models. Specifically, one difficulty resides in inferring the distribution of the latent random variables Δ_d and R, in the CPM: because of the maximum operator, the measured added volumes do not necessarily correspond to observations of Δ_d and similarly, the measured C+D periods do not necessarily correspond to observations of R. This makes the distributions of Δ_d and R to not be directly deduced from observations of added volumes and C+D periods. This inability to infer the distribution of Δ_d and R has deeply challenged the assessment of the CPM in the literature.

The CAM: a new model combining all three models in one

Along to these models we introduce a novel model called the Concurrent Adders Model (CAM) that combines hypotheses assumed in the previously introduced models. The CAM is a CPM-like model that incorporates replication-based division control via an initiation-division adder as in the RDAM rather than an initiation-division timer (Fig. 2 C, E). In line with the other models, we assume that the CAM also involves an initiation adder. Let $(\Delta_d^{(n)})_{n\geq 0}$ and $(\Delta_{id}^{(n)})_{n\geq 0}$ be i.i.d. sequences of positive random variables with respective densities f_{Δ_d} and $f_{\Delta_{id}}$ (Table 1). The couple $(V_b^{(n)}, \tilde{V}_i^{(n)})_{n\geq 0}$ satisfies according to the CAM:

$$\begin{pmatrix} V_b^{(n+1)} \\ \tilde{V}_i^{(n+1)} \end{pmatrix} = \begin{pmatrix} \max\left(\tilde{V}_i^{(n)} + \Delta_{id}^{(n)},\ 2^{-1}(V_b^{(n)} + \Delta_d^{(n)})\right) \\ 2^{-1}(\tilde{V}_i^{(n)} + \Delta_i^{(n)}) \end{pmatrix}. \tag{8}$$

Similarly to the CPM, the random variables Δ_d and Δ_{id} are latent and thus their distribution can not be directly inferred from the observations of inter-division added volume and added volume per origin over the C+D period.

Conditions for the validity of the DIAM and RDAM

Stating that division and replication are decoupled processes as proposed by the DIAM, implies that the sequences $(T_i^{(n)})_{n\geq 0}$ and $(T_b^{(n)})_{n\geq 0}$ are independent and thus events such as $T_b^{(n+1)} \leq T_i^{(n)}$ – interpreted as negative C+D periods – could occur if no additionnal assumptions are made. Similarly, if cell division is solely regulated by replication, as proposed by the RDAM, this implies that the sequence $(T_b^{(n)})_{n\geq 0}$ depends exclusively on $(T_i^{(n)})_{n\geq 0}$ and $(\tilde{V}_i^{(n)})_{n\geq 0}$, without explicitly considering the timing of the previous divisions. Consequently, events such as $T_b^{(n+1)} \leq T_b^{(n)}$, where a daughter cell divides before its mother, may not necessarily be prevented in a stochastic model with no additional assumptions. Here, we discuss that by accounting solely replication-based or division-based control of division may lead to inconsistencies without prior assumptions,

Table 1: **Random variables involved in the models and their meaning.**

Variables	Models	Meaning of the variables
$\Delta_d^{(n)}$	DIAM, CPM, CAM	Added volume from birth to the onset of replication-independent division trigger
$\Delta_i^{(n)}$	All	Added volume per origin between two subsequent initiation of replication events
$\Delta_{id}^{(n)}$	RDAM, CAM	Added volume per origin from the initiation of replication starting the C+D period to the onset of replication-dependent division trigger
$R^{(n)}$	CPM	Elapsed time from the initiation of replication starting the C+D period to the onset of replication-dependent division trigger

and we provide necessary and sufficient conditions for the DIAM and the RDAM to avoid these irregularities within the presented framework.

Intuitively, by considering division and replication as independent processes, as proposed by the DIAM, it is likely that a cell lineage may accumulate more division events than initiation events, which would lead to the division of a cell with only one origin of replication. Such event could biologically be interpreted as leading to DNA "guillotining" [38, 39] or the formation of an anucleated cell [40]. In the context of our framework, this event corresponds to the birth of a cell with less than a half of an origin of replication which coincides with an event $T_b^{(n+1)} \leq T_i^{(n)}$ by (3), which corresponds to a negative C+D period. Moreover, if replication and division are decoupled one can imagine to an excessive accumulation of initiation events over division events, resulting in an infinite number of replication origins.

In this aim, one can study the long-time behavior of the DIAM, and specifically the long-time behavior of the number of origins of replication. In Proposition 1 and 2 (Mathematical Supplementary) we show that the sequences $(V_b^{(n)})_{n\geq 0}$ and $(\tilde{V}_i^{(n)})_{n\geq 0}$ both converge to stationary distributions. Then, in Proposition 3 (Mathematical Supplementary), we show using (3) that thanks to these convergences in distribution, the number of origins of replication at birth converges in distribution to a limit random variable denoted by ω_∞ whose distribution satisfies:

$$\forall k \in \mathbb{Z}, \qquad \mathbb{P}\left[\omega_\infty \leq 2^k\right] = \mathbb{P}\left[V_b^\infty < 2^{k+1}\tilde{V}_i^\infty\right], \tag{9}$$

with V_b^∞ and $\tilde{V}_i^\infty$ the random variables limit in distribution of $(V_b^{(n)})_{n\geq 0}$ and $(\tilde{V}_i^{(n)})_{n\geq 0}$ respectively.

This first result shows that despite the decoupling of replication and division cycles, the DIAM ensures an homeostasis of the number of origins of replication. One can notice that in order to ensure a number of origin at birth greater or equal to one with probability one, the distributions of V_b^∞ and $\tilde{V}_i^\infty$ must satisfy $\mathbb{P}[\tilde{V}_i^\infty \leq V_b^\infty] = 1$. Theorem 1 (Mathematical Supplementary) provides that this condition is verified if and only if

$$\exists \sigma > 0, \quad \text{such that} \quad \Delta_i < \sigma, \quad \text{and} \quad \Delta_d > \sigma \quad \text{almost surely.}$$

This strong result highlight the necessary and sufficient conditions to apply to the DIAM to prevent the division of a cell with one origin of replication. It is straightforward to check wether this condition is satisfied from experimental data, as this condition is not satisfied if the empirical

distributions of Δ_d and Δ_i overlap. Fig. 3 shows the empirical distributions of Δ_d and Δ_i obtained in slow growth conditions in [17, 18]. As highlighted, the supports of the distributions of Δ_i and Δ_d overlap which suggests that if the cell cycles are ruled by the DIAM, in these particular conditions, there is a positive probability for a cell to divide when it has one origin of replication. From these distributions it is possible to simulate the distributions of V_b^∞ and $\tilde{V}_i^\infty$ and estimate this probability as it is equal to $\mathbb{P}[V_b^\infty < \tilde{V}_i^\infty]$ (see Fig. 3, Methods). Through a Monte-Carlo method, this probability has been computed to be of the order of 10^{-3}, and 10^{-4} according to the data of [17] in MG1655 acetate and NCM3722 arginine respectively, and 10^{-3} in glycerol according to the data of [18]. In other growth conditions, this probability turns out to be highly negligible.

Here we have discussed the conditions of validity of he DIAM by assessing whether it allows division of a cell with one origin of replication. But this condition is not sufficient in order to prevent defects in genetic material partitioning. A mother cell must also have two complete chromosomes in order split equally its chromosomes into its daughters. This involves a stronger condition than the one we tested, meaning that the probability of such event is greater than the ones we computed.

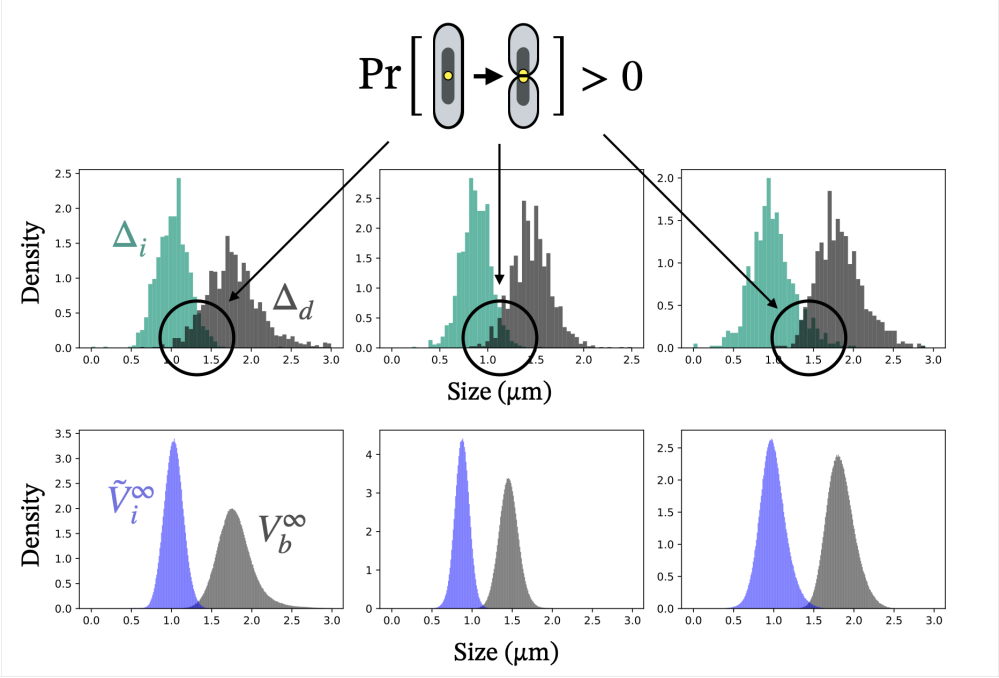


Figure 3: **Probability of one-origin-cell division according to the DIAM** (Top) Empirical distributions of Δ_i, and Δ_d in MG1655 acetate, NCM3722 arginine and BW27378 glycerol from left to right, taken from [17, 18]. The overlap of the distributions of Δ_i, and Δ_d indicates a positive probability for cells to divide with one origin of replication. (Bottom) Simulated distributions of $\tilde{V}_i^\infty$ and V_b^∞ from Monte Carlo method, which allowed to compute the probability of one-origin-cell division (see Methods).

Then, according to the RDAM, due to the stochastic nature of the added volume per origin over a C+D period Δ_{id}, it is possible for one C+D period to be nested within another. In other words, a C+D period may start and finish within the duration of a preceding C+D period. This scenario

would imply that the $(n+1)$-th birth could occur before the n-th birth, meaning that a daughter cell could divide before its mother. In order to prevent such event, the sequence of birth times, $(T_b^{(n)})_{n\geq 0}$, must be non-decreasing. We show in Proposition 4 that $(T_b^{(n)})_{n\geq 0}$ is non-decreasing if and only if

$$\exists(\sigma_1,\sigma_2)\in\mathbb{R}_+^2\setminus\{(0,0)\},\quad \text{such that}\quad \Delta_i\geq\sigma_2,\quad \text{and}\quad \sigma_1\leq\Delta_{id}\leq 2\sigma_1+\sigma_2\quad \text{a.s.}$$

These conditions involve restrictions over the supports of the densities of Δ_i and Δ_{id}. If a daughter cell may divide before its mother with positive probability, one can estimate this probability from the observations of Δ_i and Δ_{id} since this probability is precisely equal to $\mathbb{P}[2\Delta_{id}^{(n+1)}+\Delta_i^{(n)}\leq\Delta_{id}^{(n)}]$ (see Methods). Based on datasets published in [17] and [18] we show that this probability is positive in every condition while being of the order of 10^{-4} for the strains MG1655 [17] and BW27378 [18] and reaching 10^{-8} in the strain NCM3722 [17].

Distribution inference and systematic comparison of replication-division coordination models

To date, the test and the comparison of replication-division coordination models has extensively been led by assessing the ability of models to reproduce correlations in the data. From these analyses of the correlations, independence between cell quantities have been concluded from zero-correlations [17, 18, 26, 32, 31]. Such deduction holds only if the variables of interest are normally distributed, however, Shapiro–Wilk normality tests indicate that division, birth, and initiation sizes deviate significantly from normal distributions (see Table S1). Then, correlation analysis gives little indication of the performance of the CPM, as the latter is based on non-linear relationships between cell variables and has so far only been studied using linear approximations [27, 28, 29, 31]. Here, we go beyond correlations and evaluate models according to their ability to reproduce division volume distribution conditioned on covariates that we have defined as the volume at birth, the volume per origin at replication initiation starting the C+D period, and the growth rate over this C+D period. To this end, we analyze 17 datasets from [17, 18, 30], which report measurements obtained using microfluidic Mother Machine devices. These datasets include information on cell length at birth and division, cell length per origin at initiation and C+D period duration. While the previous framework focused on cell volumes we hereafter use cell size as a proxy of cell volume.

Conditional distributions of division volume

Here, we define the conditional distribution of division volume given birth and initiation volume and cell elongation rate over the C+D period. In the context of the framework, this coincides with the study of the distribution of $2V_b^{(\bullet+1)}|V_b^{(\bullet)},\tilde{V}_i^{(\bullet)},\Lambda^{(\bullet)}$. As shown in the Mathematical Supplementary, Proposition 5, for the different models, the conditional distribution of the volume at division $2V_b^{(\bullet+1)}$, given that $V_b^{(\bullet)}=v_b$, $\tilde{V}_i^{(\bullet)}=\tilde{v}_i$, and $\Lambda^{(\bullet)}=\lambda$ satisfies for the different models the following expressions for any $v>0$:

- The DIAM:

$$p(v|v_b,\tilde{v}_i,\lambda)=f_{\Delta_d}(v-v_b), \tag{10a}$$

- The RDAM:

$$p(v|v_b,\tilde{v}_i,\lambda)=\frac{1}{2}f_{\Delta_{id}}\Big(\frac{v}{2}-\tilde{v}_i\Big), \tag{10b}$$

- The CPM:

$$p(v|v_b, \tilde{v}_i, \lambda) = f_{\Delta_d}(v - v_b) F_R\left(\frac{1}{\lambda}\log\left(\frac{v}{2\tilde{v}_i}\right)\right) + \frac{1}{\lambda v} f_R\left(\frac{1}{\lambda}\log\left(\frac{v}{2\tilde{v}_i}\right)\right) F_{\Delta_d}(v - v_b), \quad (10c)$$

- The CAM:

$$p(v|v_b, \tilde{v}_i, \lambda) = f_{\Delta_d}(v - v_b) F_{\Delta_{id}}\Big(\frac{v}{2} - \tilde{v}_i\Big) + \frac{1}{2} f_{\Delta_{id}}\Big(\frac{v}{2} - \tilde{v}_i\Big) F_{\Delta_d}(v - v_b), \quad (10d)$$

with F_{Δ_d}, $F_{\Delta_{id}}$ and F_R being the cumulative distribution functions of Δ_d, Δ_{id}, and R respectively.

Distribution inference of the models

In the following we assume that the i.i.d. random variables involved in all models follow parametric densities. Specifically, we assume that these follow Generalized Gamma distributions. This three-parameters distribution gathers a broad family of unimodal probability distributions supported on $\mathbb{R}_+$, such as the Gamma, Weibull, Log-Normal, inverse-Gamma, inverse-Weibull distribution and others [41]. As demonstrated, this parametric distribution provides a good fit for the distribution of inter-division added volume, as previously shown in [42], as well as that of the C+D period, and the added volume per origin over the C+D period (see Figure S1). This motivates the choice of this distribution for Δ_d, Δ_{id} and R.

By considering a dataset $(v_d^k, v_b^k, \tilde{v}_i^k, \lambda^k)_{1\leq k\leq N}$ containing the measurements for N cell cycles of the division volume ($2V_b^{(\bullet+1)}$), birth volume ($V_b^{(\bullet)}$), initiation volume per origin ($\tilde{V}_i^{(\bullet)}$), and cell elongation rate over the C+D period ($\Lambda^{(\bullet)}$), one may define the likelihood of each model for a parameter vector θ:

$$L^N(\theta) = \prod_{k=1}^{N} p_\theta(v_d^k, |v_b^k, \tilde{v}_i^k, \lambda^k), \quad (11)$$

with p_θ the conditional distribution of division volume defined by either (10a)–(10d), with θ containing the parameters of the distributions of the i.i.d. random variables.

The distribution parameters can be estimated by maximizing this likelihood. As already mentioned, Δ_d and Δ_{id} are directly observable in the DIAM and RDAM, making their distributions identifiable from the data, and the maximization of their likelihood feasible. In contrast, in the CPM and CAM, the variables Δ_d, R, and Δ_{id} are latent, and thus are not as trivially identifiable. In Proposition 6 (Mathematical Supplementary) we show for a general case a sufficient condition to satisfy in the data to have the identifiability of the distributions of the latent variables of the CPM and CAM. Then, the maximization of their likelihood is a challenging optimization problem. To solve this problem, we have developed an Expectation-Maximization (EM) algorithm suited for parameter estimation in models such as the CPM and CAM (see Mathematical Supplementary, Figures S2-S5).

Comparison of conditional distributions of division size using BIC: third-category models best predict division size given cell covariates

After finding θ^* maximizing the likelihood (equation 11) for each model, one can compare these by computing the Bayesian Information Criterion (BIC) for every model defined as $BIC = -\log(L^N(\theta^*))+$

$\rho \log(N)/2$, with ρ being the number of parameters ($\rho = 3$ for the DIAM and RDAM; $\rho = 6$ for the CPM and CAM). The BIC is a criterion that allows to compare models based on an approximation of their posterior probability given the data. Intuitively, a lower BIC value indicates a greater probability that a model actually explains the data [43]. Compared to the Akaike Information Criterion (AIC), the BIC has a stronger penalty term, which implies that, according to the latter criterion, the CPM and CAM are penalized more severely.

Over the 17 datasets the BIC has bas been computed for each model (Table S2). The BIC scores show that either the CPM or the CAM best predict the division size conditionally to the birth and initiation size and cell elongation rate, over a strong majority of datasets. Our results indicate that the performance of the models is growth-dependent. For instance, as shown in Fig. 4, the RDAM, the CPM and the CAM consistently exhibit lower BIC scores than the DIAM under slow growth conditions, specifically under a critical average cell elongation rate around 0.5 h^{-1}. Given that the DIAM relies solely on birth volume, this suggests that in slow growth conditions, the initiation volume per origin brings strong advantage in predicting the volume at division. This observation coincides with latest results that argue that division is strongly controlled by replication in slow growth conditions [29, 30, 31]. Conversely, the RDAM under performs in predicting the division size in fast growth conditions, highlighting a significative role of the birth size in the division control. Then, our results suggest that replication-based control of division itself is growth-dependent. Specifically, the CAM and the RDAM, show lower BIC scores than the CPM under slow growth conditions ($< 0.3\ h^{-1}$). Finally, the performance of the models in division volume prediction appears to be strain dependent, as evidenced by the CAM yielding lower BIC scores than the CPM for the MG1655 strain in every growth conditions tested. Although, the RDAM and DIAM yield the best performance in the BW27783 AlaTrE and BW27783 GluCas datasets, respectively, this outcome is not supported by the AIC scores, according to which neither model achieves the best performance in any dataset (see Table S2, Figure S6).

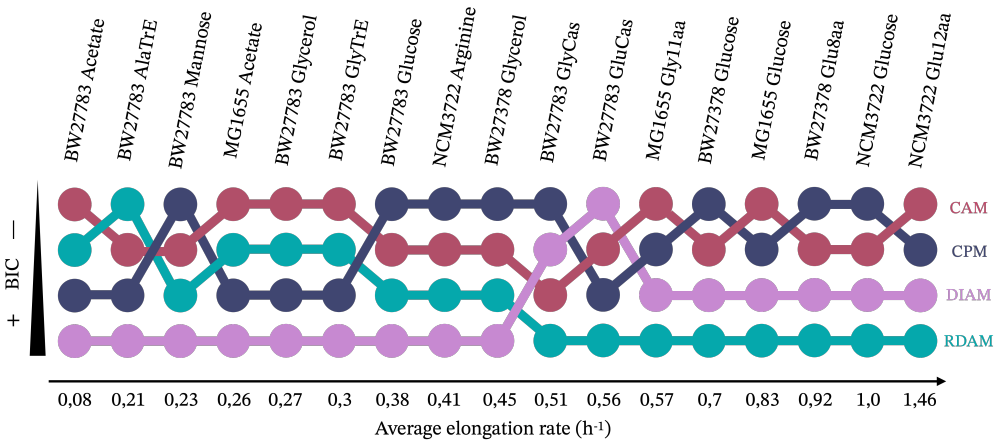


Figure 4: **Comparison of the performances of the models based on the BIC** The BIC scores have been computed for each model (DIAM in pink, RDAM in cyan, CPM in dark blue and CAM in dark red) for multiple datasets of four strains. Here are shown the ranks of the models based on their BIC score, with a lower BIC score indicating a greater probability that a model actually explains the data.

Comparison of initiation-division sizes joint distributions using Wasserstein distance: CAM best reconstructs initiation-division sizes dependence

The previous results show that the CPM and the CAM are the models most likely to capture replication-division coordination. Although these models assume distinct mechanisms of replication-dependent division control the current results do not clearly favor one mechanism over the other. To address this, we compared their ability to reproduce the empirical joint distribution of initiation and division volumes which, beyond correlations, provides comprehensive information on their dependency. By simulating division sizes conditionally on measured cell covariates, one can assess which model between CPM and CAM best captures the joint distribution of initiation and division sizes. To illustrate this, in Figure 5B, we display the heat map of the experimental joint distribution of division and initiation sizes from selected datasets and their counterparts when division size is simulated by either CPM or CAM. It appears that the CAM produces distributions that are closer to the data than the CPM in the datasets a) and c). To quantify the dissimilarity between empirical joint distributions, previous studies have relied on the average of the squared difference between the theoretical joint distributions [44, 45]. However, this method requires to know the theoretical distribution of the data, and thus, often relies on regularization of the empirical data. Here we use a metric suitable for comparing 2D empirical distributions: the Wasserstein distance. Also known as the Earth Mover's Distance, this well-studied metric is solution of a transport problem that computes the minimum average distance between paired points of two empirical distributions (see Figure 5A). This metric thus precisely measures the dissimilarity between two distributions [46, 47]. By comparing experimental and simulated data, the Wasserstein distance provides a method to assess the performance of each model in reconstructing the joint distribution of initiation and division sizes. Note that intuitively, the Wasserstein distance penalizes models whose simulated data points are far from the observed data points whereas the likelihood penalizes models that assign low probability to observed data points. While these two penalty criteria may be related, these are not equivalent and thus may lead to different conclusions. To compare the CAM and the CPM, we thus computed W, the Wasserstein distance between the experimental joint distribution of division and initiation volumes $(2V_b^{(\bullet+1)}, \tilde{V}_i^{(\bullet)})$, and its counterpart when the division volume is simulated according to either the CPM or CAM (see Figure 5A). This empirical Wasserstein distance has been computed over 100 simulations according to the CPM and CAM, and the resulted collections of Wasserstein distances have been compared through Wilcoxon-Mann-Whitney tests (see Materials).

As shown in Figure 5C, the CAM produces joint distributions of initiation and division sizes with significantly smaller Wasserstein distances from the experimental data than the CPM on a strong majority of datasets. These results indicate that the CAM more accurately reconstructs the joint distribution of division and initiation sizes, reflecting a better reproduction of the dependency between these two variables. By performing the same analysis with the joint distribution of division and birth sizes, we also found that the CAM more accurately reconstructs the dependency between these two variables (see Figure S9). From these results, strong arguments are provided in favor to the initiation-division adder as replication-based control of division under every growth condition.

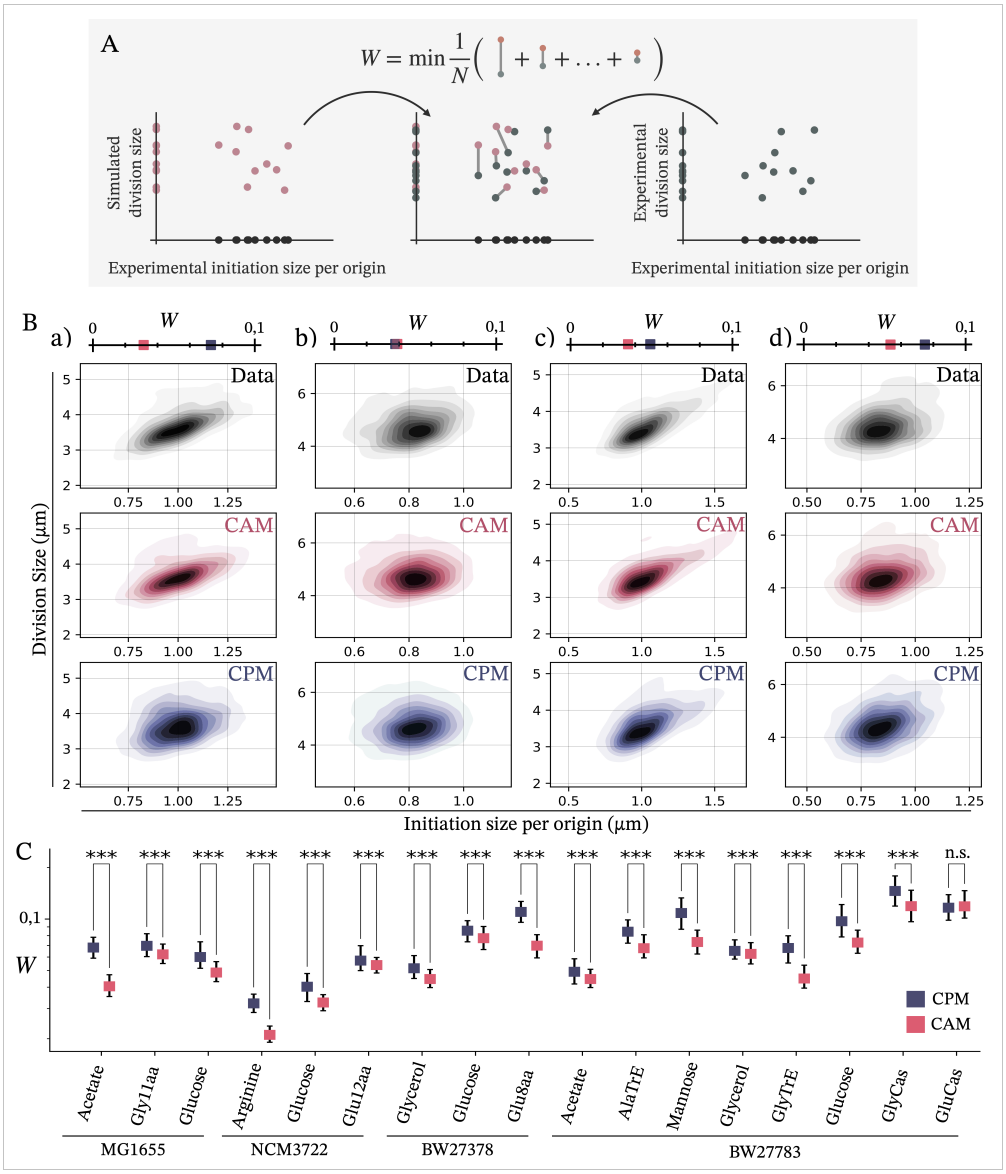


Figure 5: **CAM best reconstructs initiation-division sizes dependence** (A) The Wasserstein distance between two empirical joint distributions represents the minimum average distance between paired points. This measures the dissimilarity between two empirical distributions. Intuitively, a greater Wasserstein distance indicates a greater dissimilarity. In this work, we compute the Wasserstein distance W between the experimental joint distribution of division size and initiation size, and the corresponding distribution when the division volume is simulated according to the CPM or the CAM. (B) Examples of joint distributions between division size and initiation size from datasets ((a) MG1655 Acetate, (b) MG1655 Glucose, (c) BW27378 Glycerol, (d) BW27378 Glucose8a), and corresponding simulations of CAM, and CPM. At the top are mapped the according Wasserstein distances between the data and the simulations (CPM in blue and CAM in red). (C) Wasserstein distances between experimental and simulated joint distributions of division and initiation sizes over 100 simulations according to the CPM and CAM. Significant differences according to Wilcoxon-Mann Whitney tests are indicated: ***p-value < 0.001, n.s.= not significant (p-value > 0.05).

Discussion

In this work, we proposed a rigorous framework for the study of models of the coordination between division and replication cycles. We studied four models: the DIAM, which assumes that division is solely triggered according to a division adder; the RDAM, which places replication at the center of division triggers, assuming that cells divide according to an initiation-division adder; the CPM, which assumes that division occurs once two independent processes are completed, one corresponding to a division adder and the other to a timer over the C+D period; and finally the CAM which is a novel model firstly proposed in this paper which is similar to the CPM except that it considers the replication-based control of the division as an initiation-division adder.

We studied the conditions of validity of a division agnostic and a replication agnostic models with a rigorous mathematical point of view. Confronted with the data, our results led to the conclusion that if cells follow such models, unphysiological behavior could arise with positive probability, namely, cell division with one origin of replication for the DIAM, and daughter cells dividing before their mothers for the RDAM. While computations of the probabilities of these events would classify them as rare, this shows that these models may not be robust enough to describe finely the coordination between division and replication cycles. Beyond discussing the limits of these models, our according mathematical results provide necessary and sufficient conditions over these models in order to avoid such event. Specifically in the DIAM, we provide specific assumptions regarding the supports of the densities f_{Δ_d} and f_{Δ_i} to ensure that the number of replication origins at birth remains greater than or equal to one with probability one. This finding suggests that independent replication and division processes could maintain the number of replication origins within physiological ranges. This may have implications in the field of synthetic biology. Recent advances in this area have focused on the development of synthetic minimal cells for industrial applications [48]. The observation that independent division and initiation adders, under certain assumptions, can lead to a balanced coupling of replication and division cycles, opens the possibility of engineering cells lacking control proteins involved in the synchronization of these processes. Considering a system where division and replication are driven solely by independent molecular networks, and meet the conditions outlined in this result, could potentially be a step toward building self replicating cells.

In this work we proposed a rigorous model comparison methodology going beyond the autocorrelation analysis extensively used to date for studying coordination between division and replication cycles. Specifically, we examined the ability of the studied models to predict division volume conditioned on birth volume, initiation volume, and elongation rate. This analysis sheds light on the role of these three covariates on the onset of division. Based on this statistical analysis we highlighted that the RDAM, which does not take into account the birth volume, and the DIAM, which does not take into account the initiation volume, predict less effectively the division size in comparison to the CAM or the CPM which take into account those two covariates in the conditional distribution of division volume. This conclusion also holds according to our Wasserstein distance analysis showing the CAM yielding to better reconstructions of joint distributions of initiation-division and birth-division joint distributions than the DIAM and RDAM (see Figure S9). Our results thus highlight that birth volume and initiation volume are both determinant covariates in cell volume at division under all growth conditions. This conclusion apparently contradicts the experimental results of [17]. In the latter publication, an experiment showed that oscillations of DnaA gene expression

lead to the oscillations of the volume at initiation while the volume at birth seems to be barely affected by these oscillations. This experiment led to the conclusion that the replication cycles and the division cycles rely on two independent mechanisms, motivating the DIAM. To settle this contradiction with our results, we simulated perturbations in the initiation adder and observe how these perturbations affects the volume at birth according to the CAM. As shown in Figure 6, our simulations exhibit similar behavior to the one observed in the experiment: the initiation volume oscillates, while the birth volume is minimally affected. Our simulations show that the CAM is consistent with the results of [17] while these have been considered as a proof of independence of the division cycles from the initiation cycles.

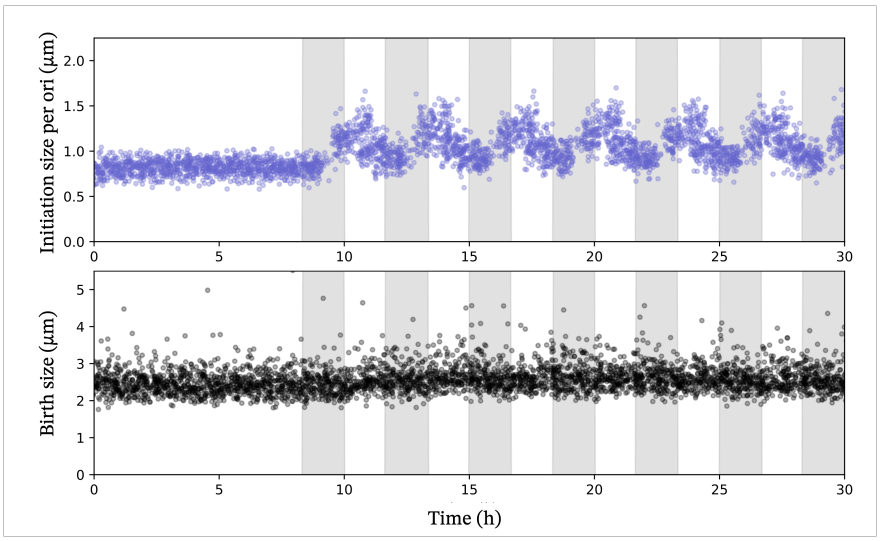


Figure 6: **Simulation of the CAM under oscillating distribution of Δ_i** In this simulation, we follow 100 lineages whose the distribution of Δ_i is increased and return to normal periodically with a period equal to 4 times the generation time (see Methods). The grey shaded areas correspond to the periods during which the Δ_i is increased. At the top, is displayed the initiation size per origin and at the bottom is displayed the birth size.

While we rule out the DIAM and the RDAM as universal models of replication-division coordination in *E. coli*, our results do not bring clear consensus to support either the CPM or the CAM. The CPM assumes that cells complete replication and segregation of the chromosomes after a fixed time, independent of the birth volume, the volume at replication initiation, and the cellular elongation rate. In contrast, the CAM posits that these processes conclude once the cell has added a specific volume per origin since replication initiation. Our results show that the CPM shows lower BIC scores than the CAM in fast growth conditions. However, according to the Wasserstein distance analysis, the CAM better reconstructs the dependence between division and initiation size under every growth condition. Interestingly, the RDAM also yields to smaller Wasserstein distances than the CPM between experimental and simulated joint distributions of initiation and division sizes in a majority of datasets (see Figure S8). This highlights that the CPM less effectively reproduces the dependence between division and initiation sizes than the RDAM which involves an initiation-division adder as replication-based control of division. Moreover, recent results [30, 31] have suggested a preponderant role of replication in division decision in slow growth conditions that

the CAM succeeds to capture whereas the CPM fails to do so (see Figure S6). Our results thus strongly supports the initiation-division adder as replication-based control of division. Although the greater BIC values of the CAM than those of the CPM call for refinements, this suggests that the CAM is close to be a universal model for coordination of replication and division cycles in *E. coli*.

Whilst our results support the CAM, to our knowledge no mechanistic intuition has yet been proposed to explain the initiation-division adder [35, 18, 49]. Although it has been suggested that this process may result from a protein-counting mechanism similarly to the division adder with FtsZ [50, 17, 51, 52, 53], we propose instead that the initiation-division adder in the CAM originates from the dynamics of nucleoid formation [54, 55, 56]. The underlying hypothesis we put forward is that a cell does not only need to replicate and segregate its two chromosomes in order to divide, but also needs to add sufficient volume from initiation to make room for the sister chromosomes and for the septum formation between segregated nucleoids. It has been shown that nucleoid elongation and segregation occur at approximately the same rate as cell elongation [57, 58, 59, 60]. Assuming that septation starts once the cell has added a critical size per sister chromosome since initiation, this gives rise, in combination with the assumption of growth-balanced nucleoid elongation and segregation, to an adder behavior over the C+D period as pictured in the CAM [17]. It is plausible that whether septation is limited by an adding size or DNA replication termination depends on growth conditions and strain variability, which could explain the heterogeneity observed in the BIC comparisons between CPM and CAM. We thus conjecture that another cell variable might be at stake in order to have a full and universal picture of the coordination between replication and division cycles. Specifically, the replication termination might be a key event involved in both division control and also initiation of replication as outlined in [31], that could be included in the CAM.

Materials and methods

The data

In this study the statistical analysis was conducted over the datasets collected by [17] found at `https://www.cell.com/current-biology/fulltext/S0960-98221930491-9#mmc3`, and by [18] found at `https://github.com/guiwitz/DoubleAdderArticle/blob/master/Data_export/Fig1_2_3.csv` and by [30] found at `https://data.mendeley.com/datasets/hwzcywscc4/draft?a=d486eddf-c4c6-40ee-b262-913f79e5d47b`.

From each datasets, four variables have been extracted: the size at division ($2V_b^{(\bullet+1)}$), the size at birth ($V_b^{(\bullet)}$), the size per origin right after the initiation of replication at the beginning of the C+D period ($\tilde{V}_i^{(\bullet)}$), and the C+D period duration ($T_b^{(\bullet+1)} - T_i^{(\bullet)}$). From these variables, the elongation rate over the C+D period ($\Lambda^{(\bullet)}$) was computed by $\log(V_b^{(\bullet+1)}/\tilde{V}_i^{(\bullet)})/(T_b^{(\bullet+1)} - T_i^{(\bullet)})$.

Maximization of the likelihoods

The likelihoods defined by equation (11) of the DIAM and RDAM have been maximizing using the Nelder Mead method implemented in the package *optimize* of SciPy in Python [61]. To maximize the likelihoods of the CPM and the CAM, an EM algorithm was implemented (see Mathematical supplementary). The maximization steps was performed, using Nelder Mead method implemented in the package *optimize* of SciPy and CMA-ES algorithm implemented in the package *cma* in Python [62]. The likelihood maximization was performed 15 times with different initial points to identify the global maximum.

Wasserstein distances computation

To compute the empirical Wasserstein distances used in this study, we proceeded as follow: for a given model M (chosen among DIAM, RDAM, CPM or CAM), we simulate the division sizes $(V_k^M)_{1\leq k\leq N}$, conditionally to the triplet $(v_b^k, \tilde{v}_i^k, \lambda^k)_{1\leq k\leq N}$ – corresponding to an observation of $(V_b^{(\bullet)}, \tilde{V}_i^{(\bullet)}, \Lambda^{(\bullet)})$. Then we compute the Wasserstein distance between the experimental joint distribution $(v_k, x_k)_k$ and its simulated counterpart $(V_k^M, x_k)_k$ where $(x_k)_k$, denotes either $(v_b^k)_k$, $(\tilde{v}_i^k)_k$ or $(\lambda^k)_k$. By using the Python package POT [63] we computed the following Wasserstein distance:

$$W\Big((V^M, x), (v, x)\Big) := \min_{\sigma\in\Sigma_N} \frac{1}{N}\sum_{k=1}^{N} \sqrt{(V_k^M - v_{\sigma(k)})^2 + (x_k - x_{\sigma(k)})^2},$$

where Σ_N represents all possible permutations over a N-uplet. The parameters used in the simulations were previously estimated from the data by maximizing the likelihood defined by (11). For each dataset and each model M, we computed the Wasserstein distances for 100 simulations. From these, we compared the Wasserstein distances of the models using the Wilcoxon-Mann-Whitney test. By denoting W_x^M the Wasserstein distance between $(V_k^M, x_k)_k$ and $(v_k, x_k)_k$ for one simulation, according the model M, we tested:

$$\mathcal{H}_0 : \mathbb{P}[W_x^{M_1} \leq W_x^{M_2}] = \frac{1}{2}$$

$$\mathcal{H}_1 : \mathbb{P}[W_x^{M_1} \leq W_x^{M_2}] \neq \frac{1}{2}.$$

Estimation of the one-origin-cell division probability in the DIAM

For each dataset the empirical distributions of Δ_d and Δ_i have been regularized through Gaussian kernel density estimation with the package *stats* form SciPy of Python. Then, $L = 10^6$ cell lineages have been simulated over 100 generations. The final collection of volumes at birth $(v_b^k)_k$ and volumes per origin at initiation $(\tilde{v}_i^k)_k$ have been used to compute the probability by counting the number of couples that satisfy $v_b < \tilde{v}_i$:

$$\hat{p} = \frac{1}{L^2}\sum_{n=1}^{L}\sum_{k=1}^{L} \mathbb{1}_{v_b^n < \tilde{v}_i^k}$$

Estimation of the probability of a daughter cell dividing before its mother in the RDAM

The probability that a daughter cell divides before its mother in the RDAM has been computed from measurements of Δ_i and Δ_{id}, noted $(\hat{\Delta}_i^k)_{1\leq k\leq N}$ and $(\hat{\Delta}_{id}^k)_{1\leq k\leq N}$ respectively, by counting the number of times the inequality $2\hat{\Delta}_{id}^j + \hat{\Delta}_i^k < \hat{\Delta}_{id}^l$ is true for any $1 \leq j, k, l \leq N$. This gives the following empirical probability:

$$\hat{p} = \frac{1}{N^3}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} \mathbb{1}_{2\hat{\Delta}_{id}^j + \hat{\Delta}_i^k < \hat{\Delta}_{id}^l}.$$

Simulation of the CAM under oscillating Δ_i

For this simulation, we generated birth and initiation volumes of 100 lineages, initializing each with $T_i^{(0)}$ distributed uniformly over $[-\tau, 0]$, where τ denotes the average generation time. To closely replicate the conditions of the DnaA underexpression experiment conducted by [17], we simulated the CAM using parameters inferred from the MG1655 Glucose dataset. The timeline was divided into intervals of normal expression and intervals of DnaA underexpression. Assuming Δ_i follows a Generalized Gamma distribution with parameters *a priori* estimated from the same dataset, we modeled the effect of DnaA underexpression

by multiplying $\Delta_i^{(n)}$ by a factor $f > 1$, when $T_i^{(n)}$ falls in an interval with DnaA underexpression. As in the original experiment, the perturbation and non-perturbation intervals follow one another and each last 2τ, resulting in an oscillatory pattern with a period of 4τ. In Fig 6, the simulation was performed using a scaling factor $f = 1,75$.

Code availability

The codes used in this work can be found at `https://github.com/Alex-PERRIN/Coordination-division-replication-E.coli`.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

Acknowledgments

The authors acknowledge Luce Breuil, Oskar Girardin, and Valentin Pesce for their fruitful discussions. This work has been supported by the Chair "Modélisation Mathématique et Biodiversité" of Veolia Environnement - Ecole Polytechnique - Museum National d'Histoire Naturelle - Fondation X. Funded by the European Union (ERC, SINGER, 101054787). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] K. Nordstrom, R. Bernander, and S. Dasgupta, "The escherichia coli cell cycle: one cycle or multiple independent processes that are co-ordinated?," *Molecular microbiology*, vol. 5, no. 4, pp. 769–774, 1991.

[2] E. Boye and K. Nordström, "Coupling the cell cycle to cell growth: A look at the parameters that regulate cell-cycle events," *EMBO reports*, vol. 4, no. 8, pp. 757–760, 2003.

[3] L. Willis and K. C. Huang, "Sizing up the bacterial cell cycle," *Nature Reviews Microbiology*, vol. 15, no. 10, pp. 606–620, 2017.

[4] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, "Robust growth of escherichia coli," *Current biology*, vol. 20, no. 12, pp. 1099–1103, 2010.

[5] A. Adiciptaningrum, M. Osella, M. C. Moolman, M. Cosentino Lagomarsino, and S. J. Tans, "Stochasticity and homeostasis in the e. coli replication and division cycle," *Scientific reports*, vol. 5, no. 1, p. 18261, 2015.

[6] S. Jun, F. Si, R. Pugatch, and M. Scott, "Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review," *Reports on Progress in Physics*, vol. 81, no. 5, p. 056601, 2018.

[7] S. Sanders, K. Joshi, P. A. Levin, and S. Iyer-Biswas, "Beyond the average: An updated framework for understanding the relationship between cell growth, dna replication, and division in a bacterial system," *PLoS genetics*, vol. 19, no. 1, p. e1010505, 2023.

[8] S. Cooper and C. E. Helmstetter, "Chromosome replication and the division cycle of escherichia coli br," *Journal of molecular biology*, vol. 31, no. 3, pp. 519–540, 1968.

[9] M. Campos, I. V. Surovtsev, S. Kato, A. Paintdakhi, B. Beltran, S. E. Ebmeier, and C. Jacobs-Wagner, "A constant size extension drives bacterial cell size homeostasis," *Cell*, vol. 159, no. 6, pp. 1433–1446, 2014.

[10] A. Amir, "Cell size regulation in bacteria," *Physical review letters*, vol. 112, no. 20, p. 208102, 2014.

[11] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, "Cell-size control and homeostasis in bacteria," *Current biology*, vol. 25, no. 3, pp. 385–391, 2015.

[12] I. Soifer, L. Robert, and A. Amir, "Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy," *Current Biology*, vol. 26, no. 3, pp. 356–361, 2016.

[13] W. D. Donachie, "Relationship between cell size and time of initiation of dna replication," *Nature*, vol. 219, pp. 1077–1079, 1968.

[14] N. S. Hill, R. Kadoya, D. K. Chattoraj, and P. A. Levin, "Cell size and the initiation of dna replication in bacteria," *PLoS genetics*, vol. 8, no. 3, p. e1002549, 2012.

[15] M. Wallden, D. Fange, E. G. Lundius, Ö. Baltekin, and J. Elf, "The synchronization of replication and division cycles in individual e. coli cells," *Cell*, vol. 166, no. 3, pp. 729–739, 2016.

[16] P.-Y. Ho and A. Amir, "Simultaneous regulation of cell size and chromosome replication in bacteria," *Frontiers in microbiology*, vol. 6, p. 662, 2015.

[17] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, "Mechanistic origin of cell-size control and homeostasis in bacteria," *Current Biology*, vol. 29, no. 11, pp. 1760–1770, 2019.

[18] G. Witz, E. van Nimwegen, and T. Julou, "Initiation of chromosome replication controls both division and replication cycles in e. coli through a double-adder mechanism," *Elife*, vol. 8, p. e48063, 2019.

[19] S. Hiraga, H. Niki, R. Imamura, T. Ogura, K. Yamanaka, J. Feng, B. Ezaki, and A. Jaffé, "Mutants defective in chromosome partitioning in e. coli," *Research in microbiology*, vol. 142, no. 2-3, pp. 189–194, 1991.

[20] O. Espéli, R. Borne, P. Dupaigne, A. Thiel, E. Gigant, R. Mercier, and F. Boccard, "A matp–divisome interaction coordinates chromosome segregation with cell division in e. coli," *The EMBO journal*, vol. 31, no. 14, pp. 3198–3211, 2012.

[21] J. Männik and M. W. Bailey, "Spatial coordination between chromosomes and cell division proteins in escherichia coli," *Frontiers in microbiology*, vol. 6, p. 306, 2015.

[22] L. Dewachter, N. Verstraeten, M. Fauvart, and J. Michiels, "An integrative view of cell cycle control in escherichia coli," *FEMS Microbiology Reviews*, vol. 42, no. 2, pp. 116–136, 2018.

[23] S. Jaramillo-Riveri, J. Broughton, A. McVey, T. Pilizota, M. Scott, and M. El Karoui, "Growth-dependent heterogeneity in the dna damage response in escherichia coli," *Molecular Systems Biology*, vol. 18, no. 5, p. e10441, 2022.

[24] L. K. Harris and J. A. Theriot, "Relative rates of surface and volume synthesis set bacterial cell size," *Cell*, vol. 165, no. 6, pp. 1479–1492, 2016.

[25] N. Ojkic, D. Serbanescu, and S. Banerjee, "Surface-to-volume scaling and aspect ratio preservation in rod-shaped bacteria," *Elife*, vol. 8, p. e47033, 2019.

[26] G. Witz, T. Julou, and E. van Nimwegen, "Response to comment on 'initiation of chromosome replication controls both division and replication cycles in e. coli through a double-adder mechanism'," *bioRxiv*, pp. 2020–08, 2020.

[27] G. Micali, J. Grilli, M. Osella, and M. Cosentino Lagomarsino, "Concurrent processes set e. coli cell division," *Science Advances*, vol. 4, no. 11, p. eaau3324, 2018.

[28] G. Micali, J. Grilli, J. Marchi, M. Osella, and M. C. Lagomarsino, "Dissecting the control mechanisms for dna replication and cell division in e. coli," *Cell reports*, vol. 25, no. 3, pp. 761–771, 2018.

[29] A. Colin, G. Micali, L. Faure, M. C. Lagomarsino, and S. van Teeffelen, "Two different cell-cycle processes determine the timing of cell division in escherichia coli," *Elife*, vol. 10, p. e67495, 2021.

[30] S. Tiruvadi-Krishnan, J. Männik, P. Kar, J. Lin, A. Amir, and J. Männik, "Coupling between dna replication, segregation, and the onset of constriction in escherichia coli," *Cell reports*, vol. 38, no. 12, p. 110539, 2022.

[31] P. Kar, S. Tiruvadi-Krishnan, J. Männik, J. Männik, and A. Amir, "Using conditional independence tests to elucidate causal links in cell cycle regulation in escherichia coli," *Proceedings of the National Academy of Sciences*, vol. 120, no. 11, p. e2214796120, 2023.

[32] G. Le Treut, F. Si, D. Li, and S. Jun, "Quantitative examination of five stochastic cell-cycle and cell-size control models for escherichia coli and bacillus subtilis," *Frontiers in microbiology*, p. 3278, 2021.

[33] H. Zheng, P.-Y. Ho, M. Jiang, B. Tang, W. Liu, D. Li, X. Yu, N. E. Kleckner, A. Amir, and C. Liu, "Interrogating the escherichia coli cell cycle by cell dimension perturbations," *Proceedings of the National Academy of Sciences*, vol. 113, no. 52, pp. 15000–15005, 2016.

[34] K. Skarstad, E. Boye, and H. B. Steen, "Timing of initiation of chromosome replication in individual escherichia coli cells.," *The EMBO journal*, vol. 5, no. 7, pp. 1711–1717, 1986.

[35] M. M. Logsdon, P.-Y. Ho, K. Papavinasasundaram, K. Richardson, M. Cokol, C. M. Sassetti, A. Amir, and B. B. Aldridge, "A parallel adder coordinates mycobacterial cell-cycle progression and cell-size homeostasis in the context of asymmetric growth and organization," *Current Biology*, vol. 27, no. 21, pp. 3367–3374, 2017.

[36] S. Iyer-Biswas, C. S. Wright, J. T. Henry, K. Lo, S. Burov, Y. Lin, G. E. Crooks, S. Crosson, A. R. Dinner, and N. F. Scherer, "Scaling laws governing stochastic growth and division of single bacterial cells," *Proceedings of the National Academy of Sciences*, vol. 111, no. 45, pp. 15912–15917, 2014.

[37] S. Iyer-Biswas, G. E. Crooks, N. F. Scherer, and A. R. Dinner, "Universality in stochastic exponential growth," *Physical review letters*, vol. 113, no. 2, p. 028101, 2014.

[38] W. R. Cook and L. I. Rothfield, "Nucleoid-independent identification of cell division sites in escherichia coli," *Journal of bacteriology*, vol. 181, no. 6, pp. 1900–1905, 1999.

[39] E. C. Hendricks, H. Szerlong, T. Hill, and P. Kuempel, "Cell division, guillotining of dimer chromosomes and sos induction in resolution mutants (dif, xerc and xerd) of escherichia coli," *Molecular microbiology*, vol. 36, no. 4, pp. 973–981, 2000.

[40] H. Adler, W. Fisher, A. Cohen, and A. A. Hardigree, "Miniature escherichia coli cells deficient in dna," *Proceedings of the National Academy of Sciences*, vol. 57, no. 2, pp. 321–326, 1967.

[41] C. Cox, H. Chu, M. F. Schneider, and A. Munoz, "Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution," *Statistics in medicine*, vol. 26, no. 23, pp. 4352–4374, 2007.

[42] I. Madrid Canales, J. Broughton, S. Méléard, and M. El Karoui, "Quantitative effects of the stress response to dna damage in the cell size control of escherichia coli," *arXiv e-prints*, pp. arXiv–2409, 2024.

[43] A. E. Raftery, "Bayesian model selection in social research," *Sociological methodology*, pp. 111–163, 1995.

[44] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert, and M. Doumic, "Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism," *BMC biology*, vol. 12, pp. 1–10, 2014.

[45] M. Doumic and M. Hoffmann, "Individual and population approaches for calibrating division rates in population dynamics: Application to the bacterial cell cycle," in *Modeling and Simulation for Collective Dynamics*, pp. 1–81, World Scientific, 2023.

[46] G. Peyré, M. Cuturi, *et al.*, "Computational optimal transport: With applications to data science," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[47] C. Villani *et al.*, *Optimal transport: old and new*, vol. 338. Springer, 2008.

[48] L. Olivi, M. Berger, R. N. Creyghton, N. De Franceschi, C. Dekker, B. M. Mulder, N. J. Claassens, P. R. Ten Wolde, and J. van der Oost, "Towards a synthetic cell cycle," *Nature communications*, vol. 12, no. 1, p. 4531, 2021.

[49] A. Meunier, F. Cornet, and M. Campos, "Bacterial cell proliferation: from molecules to cells," *FEMS Microbiology Reviews*, vol. 45, no. 1, p. fuaa046, 2021.

[50] K. R. Ghusinga, C. A. Vargas-Garcia, and A. Singh, "A mechanistic stochastic framework for regulating bacterial cell division," *Scientific reports*, vol. 6, no. 1, p. 30229, 2016.

[51] D. Serbanescu, N. Ojkic, and S. Banerjee, "Nutrient-dependent trade-offs between ribosomes and division protein synthesis control bacterial cell size and growth," *Cell reports*, vol. 32, no. 12, p. 108183, 2020.

[52] C. Nieto, C. A. Vargas-García, J. M. Pedraza, and A. Singh, "Mechanisms of cell size regulation in slow-growing escherichia coli cells: Discriminating models beyond the adder," *npj Systems Biology and Applications*, vol. 10, no. 1, p. 61, 2024.

[53] J. Männik, P. Kar, C. Amarasinghe, A. Amir, and J. Männik, "Determining the rate-limiting processes for cell division in escherichia coli," *Nature Communications*, vol. 15, no. 1, p. 9948, 2024.

[54] N. E. Kleckner, K. Chatzi, M. A. White, J. K. Fisher, and M. Stouf, "Coordination of growth, chromosome replication/segregation, and cell division in e. coli," *Frontiers in microbiology*, vol. 9, p. 1469, 2018.

[55] P. G. Huls, N. O. Vischer, and C. L. Woldringh, "Different amounts of dna in newborn cells of escherichia coli preclude a role for the chromosome in size control according to the "adder" model," *Frontiers in Microbiology*, vol. 9, p. 664, 2018.

[56] A. Zaritsky, W. Vollmer, J. Männik, and C. Liu, "Does the nucleoid determine cell dimensions in escherichia coli?," *Frontiers in Microbiology*, vol. 10, p. 1717, 2019.

[57] J. K. Fisher, A. Bourniquel, G. Witz, B. Weiner, M. Prentiss, and N. Kleckner, "Four-dimensional imaging of e. coli nucleoid organization and dynamics in living cells," *Cell*, vol. 153, no. 4, pp. 882–895, 2013.

[58] W. T. Gray, S. K. Govers, Y. Xiang, B. R. Parry, M. Campos, S. Kim, and C. Jacobs-Wagner, "Nucleoid size scaling and intracellular organization of translation across bacteria," *Cell*, vol. 177, no. 6, pp. 1632–1648, 2019.

[59] F. Wu, P. Swain, L. Kuijpers, X. Zheng, K. Felter, M. Guurink, J. Solari, S. Jun, T. S. Shimizu, D. Chaudhuri, *et al.*, "Cell boundary confinement sets the size and position of the e. coli chromosome," *Current Biology*, vol. 29, no. 13, pp. 2131–2144, 2019.

[60] C. I. Amarasinghe, M.-H. Chang, J. Männik, S. T. Retterer, M. O. Lavrentovich, and J. Männik, "The role of active mrna-ribosome dynamics and closing constriction in daughter chromosome separation in escherichia coli," *bioRxiv*, pp. 2025–04, 2025.

[61] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[62] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.

[63] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, *et al.*, "Pot: Python optimal transport," *Journal of Machine Learning Research*, vol. 22, no. 78, pp. 1–8, 2021.

Mathematical supplementary

In this supplementary document are provided every mathematical proof the article relies onto. In this work, robust mathematical results are shown in the context of a defined framework. This framework reconstructs a survey of a cell lineage in a Mother Machine device. At each generation, a mother cell is randomly replaced by one of its two daughters. This framework relies on four assumptions.

Assumptions:

1. The division is perfectly symmetric, i.e. the volume of the daughter cell at birth is equal to half of the volume of the mother at division.
2. At each replication initiation event, every origin of replication in the cell fires a new round of replication at the same time, thus doubling their number.
3. The origins of replication of a mother cell are equally partitioned into the daughters at division.
4. Over the lineage, the cell volume grows exponentially according to a growth rate function of time $\lambda(t)$, that is positive and locally integrable.

In this following, we will study four observable quantities (see Figure 1. A):

- $(V_b^{(n)})_{n\geq 0}$ the sequence of birth volumes where for all n, $V_b^{(n)}$ represents the birth volume of the mother cell giving rise to a daughter cell with a birth volume $V_b^{(n+1)}$.
- $(T_b^{(n)})_{n\geq 0}$ the dates of division giving birth to the cell of volume $V_b^{(n)}$
- $(\tilde{V}_i^{(n)})_{n\in\mathbb{N}}$, the sequence of volumes per origin of replication immediately after initiations of replication, where for all n, the initiation event associated with $\tilde{V}_i^{(n+1)}$ is subsequent to the initiation event associated with $\tilde{V}_i^{(n)}$.
- $(T_i^{(n)})_{n\geq 0}$, the date at initiation of replication of the cell with a volume per origin $\tilde{V}_i^{(n)}$

The indexation of these sequences is set as such the sister chromosomes obtained through the replication initiated at $T_i^{(n)}$, are separated at the division occurring at $T_b^{(n+1)}$. This defines the interval $T_b^{(n+1)} - T_i^{(n)}$ as a C+D period, generically referred in the literature. From an experimental point of view the initialization of these sequences can be made as follow (see Figure 1. B): we start with a cell initiating a round of replication setting $T_i^{(0)} = 0$, and $\tilde{V}_i^{(0)}$. Then, the division occurring

at the end of the C+D period initiating at $T_i^{(0)}$ sets $T_b^{(1)}$ and $V_b^{(1)}$. Finally, the birth of the cell dividing at $T_b^{(1)}$ sets $T_b^{(0)}$ and $V_b^{(0)}$. In the following we will impose on the initialization of the sequences that $V_b^{(0)} = 2\tilde{V}_i^{(0)} \exp\left(\int_{T_i^{(0)}}^{T_b^{(0)}} \lambda(s)ds\right)$. In our framework, we choose $\tilde{V}_i^{(0)} > 0$ and $V_b^{(0)} > 0$ to be free variables whose definitions can thus be set arbitrarily.

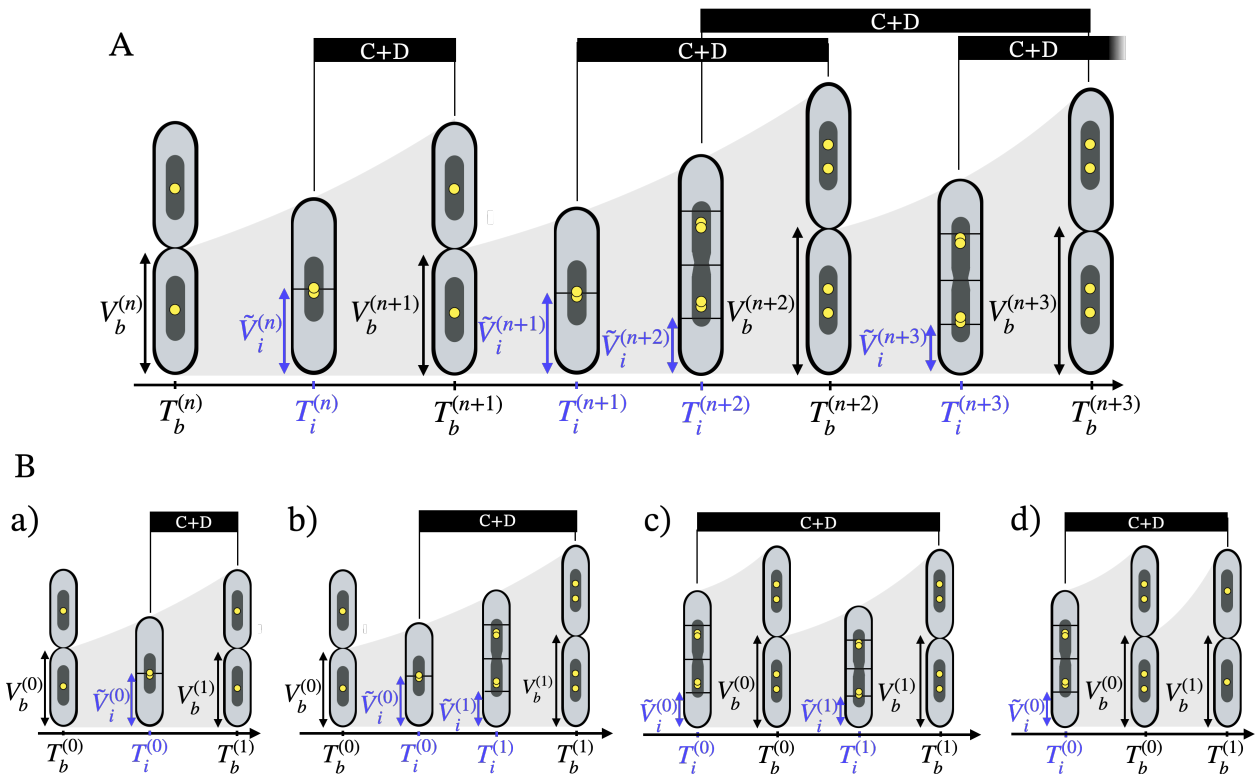


Figure 1: **The framework** We consider a cell lineage where one daughter is kept at each division. Two types of events are monitored: division events and the initiation events. The division events are modeled by the sequences $(T_b^{(n)})_{n\geq 0}$, the dates at division events, and $(V_b^{(n)})_{n\geq 0}$ the birth volume at these events. The initiation of replication events are modeled by the sequences $(T_i^{(n)})_{n\geq 0}$, the dates at initiation of replication events, and $(\tilde{V}_i^{(n)})_{n\geq 0}$ the volume per origin of replication right after these events. The order of occurrence of these events are not defined by the framework. In consequence infinitely many scenarios may set the order of occurrence of these events. (A) Illustration of a cell lineage undergoing an over-initiation, switching from a slow growth regime to a fast growth regime. (B) Illustration of possible scenarios for the initialization of the sequences. Scenario a) depicts a classical cell cycle in a slow growth regime. Scenario b) depicts an over initiation within a slow growing cell. Scenario c) depicts a classical cell cycle in a fast growth regime. Scenario d) depicts an over division of a fast growing cell. In addition to these four scenarios, the framework allows for infinite other possibilities.

Contents

1 First Mathematical expressions **4**

1.1 Relationship between $(V_b^{(n)})_{n\geq 0}$ and $(T_b^{(n)})_{n\geq 0}$ 4

1.2 Relationship between $(\tilde{V}_i^{(n)})_{n\geq 0}$ and $(T_i^{(n)})_{n\geq 0}$ 4

1.3 General formula for replication initiation-division coupling 4

1.4 The number of origins of replication at birth 5

2 Long-time behavior of the DIAM **6**

2.1 Convergence of $(V_b^{(n)})_{n\geq 0}$ and $(\tilde{V}_i^{(n)})_{n\geq 0}$ 6

2.2 Convergence of the number of origins of replication 9

3 Necessary and sufficient conditions to avoid one-origin-cell division in the DIAM **10**

4 Necessary and sufficient conditions to prevent a daughter cell from dividing before its mother in the RDAM **13**

5 Conditional densities of the models and identifiablility **14**

5.1 Expressions of the conditional densities 15

5.2 Identifiability of the conditional densities of the CPM and CAM 16

5.3 Expectation-Maximization (EM) algorithm for parameters estimation of the CPM and the CAM 18

1 First Mathematical expressions

1.1 Relationship between $(V_b^{(n)})_{n\geq 0}$ and $(T_b^{(n)})_{n\geq 0}$

According to Assumptions 1 and 4 one may find the following recurrent expression for the volume at birth $(V_b^{(n)})_{n\geq 0}$:

$$\forall n \geq 0, \qquad V_b^{(n+1)} = \frac{V_b^{(n)}}{2} \exp\left(\int_{T_b^{(n)}}^{T_b^{(n+1)}} \lambda(s)ds\right). \tag{1}$$

From this recurrent expression, one can derive by induction the following expression that will be useful in the hereafter work:

$$\forall n \geq 1, \qquad V_b^{(n)} = \frac{V_b^{(1)}}{2^{n-1}} \exp\left(\int_{T_b^{(1)}}^{T_b^{(n)}} \lambda(s)ds\right). \tag{2}$$

1.2 Relationship between $(\tilde{V}_i^{(n)})_{n\geq 0}$ and $(T_i^{(n)})_{n\geq 0}$

Similarly to the volume at birth, one may study the trajectory of the volume per origin of replication, observed at replication initiation. According to Assumptions 1 and 3, the cell volume per origin trajectory does not show discontinuity at division events. However, it shows discontinuity at replication initiation events as it is halved by two according to Assumption 2. According to Assumptions 1, 2, 3 and 4 one may find that the sequence of volume per origin at replication initiation follows the recurrence equation

$$\forall n \geq 0, \qquad \tilde{V}_i^{(n+1)} = \frac{\tilde{V}_i^{(n)}}{2} \exp\left(\int_{T_i^{(n)}}^{T_i^{(n+1)}} \lambda(s)ds\right). \tag{3}$$

This gives us by induction:

$$\forall n \geq 0, \qquad \tilde{V}_i^{(n)} = \frac{\tilde{V}_i^{(0)}}{2^n} \exp\left(\int_{T_i^{(0)}}^{T_i^{(n)}} \lambda(s)ds\right). \tag{4}$$

1.3 General formula for replication initiation-division coupling

Lemma 1. *Assuming that the sequences of birth and initiation volumes and dates are initialized in such a way that* $V_b^{(0)} = 2\tilde{V}_i^{(0)} \exp\left(\int_{T_i^{(0)}}^{T_b^{(0)}} \lambda(s)ds\right)$*, one has* $\forall n \geq 0$*:*

$$V_b^{(n+1)} = \tilde{V}_i^{(n)} \exp\left(\int_{T_i^{(n)}}^{T_b^{(n+1)}} \lambda(s)ds\right). \tag{5}$$

Proof. From Equations (2) and(4), one has:

$$\begin{aligned}
& \frac{V_b^{(n+1)}}{\tilde{V}_i^{(n)}} = \frac{V_b^{(1)}}{\tilde{V}_i^{(0)}} \exp\left(\int_{T_b^{(1)}}^{T_b^{(n+1)}} \lambda(s)ds - \int_{T_i^{(0)}}^{T_i^{(n)}} \lambda(s)ds\right) \\
\iff \quad & \frac{V_b^{(n+1)}}{V_b^{(1)}} = \frac{\tilde{V}_i^{(n)}}{\tilde{V}_i^{(0)}} \exp\left(\int_{T_i^{(n)}}^{T_b^{(n+1)}} \lambda(s)ds + \int_{T_b^{(1)}}^{T_i^{(n)}} \lambda(s)ds - \int_{T_b^{(1)}}^{T_i^{(n)}} \lambda(s)ds - \int_{T_i^{(0)}}^{T_b^{(1)}} \lambda(s)ds\right) \\
\iff \quad & \frac{V_b^{(n+1)}}{V_b^{(1)}} = \frac{\tilde{V}_i^{(n)} \exp\left(\int_{T_i^{(n)}}^{T_b^{(n+1)}} \lambda(s)ds\right)}{\tilde{V}_i^{(0)} \exp\left(\int_{T_i^{(0)}}^{T_b^{(1)}} \lambda(s)ds\right)}.
\end{aligned}$$

It follows that there exists a $K > 0$ such that for all $n \geq 0$:

$$V_b^{(n+1)} = K\tilde{V}_i^{(n)} \exp\left(\int_{T_i^{(n)}}^{T_b^{(n+1)}} \lambda(s)ds\right).$$

By taking the equality form Equation (1) $2V_b^{(1)} = V_b^{(0)} \exp\left(\int_{T_b^{(0)}}^{T_b^{(1)}} \lambda(s)ds\right)$ and the initialization assumption $V_b^{(0)} = 2\tilde{V}_i^{(0)} \exp\left(\int_{T_i^{(0)}}^{T_b^{(0)}} \lambda(s)ds\right)$, it follows that $K = 1$. □

1.4 The number of origins of replication at birth

Lemma 2. *Let $\omega^{(n)}$ be the number of origins of replication at birth at $T_b^{(n)}$. Given that $(T_i^{(n)})_{n\geq 0}$ is non-decreasing, for every $n \geq 0$ and $k \geq -n+1$:*

$$\omega^{(n)} = 2^k \iff T_i^{(n+k-1)} \leq T_b^{(n)} < T_i^{(n+k)},$$

$$\omega^{(n)} \leq 2^k \iff T_b^{(n)} < T_i^{(n+k)}.$$

Proof. We first introduce two new quantities: the number of initiation events occurred between $T_i^{(0)} = 0$ and time $t > 0$, $N_i^t := \#\{m \geq 0 : \ 0 < T_i^{(m)} \leq t\}$, and the volume per origin over time defined for $t > 0$ by $\tilde{V}_t := 2^{-N_i^t}\tilde{V}_i^{(0)} \exp\left(\int_0^t \lambda(s)ds\right)$. Note that this equation when evaluated at each date of replication initiation given by the sequence $(T_i^{(n)})_{n\geq 0}$, provides the same results as the equation (4). One may notice that the number of origins of replication at the n-th birth is precisely the ratio between the volume at the n-th birth and the volume per origin at that event. For every $n \geq 1$ and $k \geq -n+1$, one has:

$$\begin{aligned}
& \omega^{(n)} = 2^k \\
\iff \quad & V_b^{(n)} = 2^k \tilde{V}_{T_b^{(n)}} \\
\iff \quad & \tilde{V}_i^{(n-1)} \exp\left(\int_{T_i^{(n-1)}}^{T_b^{(n)}} \lambda(s)ds\right) = 2^{k-N_i^{T_b^{(n)}}} \tilde{V}_i^{(0)} \exp\left(\int_0^{T_b^{(n)}} \lambda(s)ds\right) \quad \text{by (5)}, \\
\iff \quad & 2^{-n+1}\tilde{V}_i^{(0)} \exp\left(\int_0^{T_b^{(n)}} \lambda(s)ds\right) = 2^{k-N_i^{T_b^{(n)}}} \tilde{V}_i^{(0)} \exp\left(\int_0^{T_b^{(n)}} \lambda(s)ds\right) \quad \text{by (4)}, \\
\iff \quad & N_i^{T_b^{(n)}} = n+k-1 \\
\iff \quad & T_i^{(n+k-1)} \leq T_b^{(n)} < T_i^{(n+k)} \quad \text{since } (T_i^{(n)})_n \text{ is increasing.}
\end{aligned}$$

Similarly, one may find through the same operations that for every $n \geq 1$ and $k \geq -n+1$,

$$\omega^{(n)} \leq 2^k \quad \Leftrightarrow \quad N_i^{T_b^{(n)}} \leq n+k-1 \quad \Leftrightarrow \quad T_b^{(n)} < T_i^{(n+k)}.$$

□

2 Long-time behavior of the DIAM

According to the Double Independent Adders Model (DIAM) proposed by [S12], the sequences of volume at birth and volume per origin at replication initiation follow for all $n \geq 0$:

$$2V_b^{(n+1)} = V_b^{(n)} + \Delta_d^{(n)} \tag{6}$$

$$2\tilde{V}_i^{(n+1)} = \tilde{V}_i^{(n)} + \Delta_i^{(n)} \tag{7}$$

with $(\Delta_d^{(n)})_{n\geq 0}$ and $(\Delta_i^{(n)})_{n\geq 0}$ being i.i.d. random variables of density f_{Δ_d} and f_{Δ_i} respectively. The definition of $\tilde{V}_i^{(0)}$ and $V_b^{(0)}$, can be set arbitrarily so we will assume that these are independent.

Remark: Since Δ_d and Δ_i admit densities it turns out that $V_b^{(1)}$ and $\tilde{V}_i^{(1)}$ admit densities for any $V_b^{(0)} > 0$ and $\tilde{V}_i^{(0)} > 0$. Then, by induction for all $n \geq 2$, $V_b^{(n)}$ and $\tilde{V}_i^{(n)}$ are expressed as linear combinations of two independent random variables which both admit densities, thereby ensuring that $V_b^{(n)}$ and $\tilde{V}_i^{(n)}$ themselves admit densities. Furthermore, it is immediate that $V_b^{(n)}$ and $\tilde{V}_i^{(n)}$ are positive for every n.

2.1 Convergence of $(V_b^{(n)})_{n\geq 0}$ and $(\tilde{V}_i^{(n)})_{n\geq 0}$

One of the most notable findings in the field of single-cell measurements is the narrow distribution of cell sizes and its stability across experiments. This phenomenon testifies a cell size control which has motivated numerous researches. So far, this size homeostasis is attributed to the division adder model, which has first explain deterministically the convergence of the mean of cell size to a steady state [S13, S2, S4]. However fewer works have been led to show the convergence of the distribution of cell size at birth. Gabriel and Martin (2018) [S6] demonstrated that, under the division adder model, the distribution of cell size at birth is the solution to a fixed-point problem. More recently, Madrid Canales (2023) showed that the division adder allowing asymmetric division leads to a stationary distribution for birth size that admits a density [S7].

Here, we extend these findings by showing that even with symmetric division, the volume at birth converges in distribution to a random variable which admits a density, in the condition that the added volume has a density and a finite first moment. While [S6] employed operator theory and [S7] used measure-valued processes, our approach addresses this problem within the constraints of our framework by modeling the sequence of birth volumes as a homogeneous Markov chain.

Proposition 1. *According to the DIAM, assuming that Δ_d is drawn from a density f_{Δ_d}, and*

such that $\mathbb{E}[\Delta_d] < \infty$, the sequence $(V_b^{(n)})_{n\geq 0}$ converges in law towards a random variable V_b^∞ which admits a density $f_{V_b^\infty}$.

Proof. Let us first show that $(V_b^{(n)})_n$ converges in law.
Let $\phi_{V_b^{(n)}}$ and ϕ_{Δ_d} be the characteristic functions of $V_b^{(n)}$ and Δ_d defined for every $t \in \mathbb{R}$, as $\phi_{V_b^{(n)}}(t) := \mathbb{E}[e^{itV_b^{(n)}}]$, and $\phi_{\Delta_d}(t) := \mathbb{E}[e^{it\Delta_d}]$, we have by the equation (6) :

$$\begin{aligned}\forall n \geq 1, \qquad \phi_{V_b^{(n)}}(\xi) &= \phi_{V_b^{(n-1)}}(\xi/2)\phi_{\Delta_d}(\xi/2) \\ &= \phi_{V_b^{(0)}}(\xi/2^n)\prod_{k=1}^{n}\phi_{\Delta_d}(\xi/2^k).\end{aligned}$$

As the characteristic function $\phi_{V_b^{(0)}}$ is continuous and equal to 1 at 0, $\phi_{V_b^{(0)}}(\xi/2^n)$ converges to 1 as n goes to infinity. Now we prove that the right hand term converges and is continuous at 0. For this it suffices to show that the series $\sum_{k\geq 1}|\phi_{\Delta_d}(\xi/2^k) - 1|$ is finite for all ξ (see [S1], Chapter 5: 2.2, Theorem 6), then that $\sum_{k\geq 1}\log(\phi_{\Delta_d}(\xi/2^k))$ is continuous in a neighborhood of zero.

Since we have $\mathbb{E}[\Delta_d] < \infty$, it follows that ϕ_{Δ_d} is differentiable. By applying the Taylor-Young development at 0 to the first order, one ends up for any $t \in \mathbb{R}$:

$$\begin{aligned}&\phi_{\Delta_d}(t) = \phi_{\Delta_d}(t)\Big|_{t=0} + i\mathbb{E}[\Delta_d e^{it\Delta_d}]\Big|_{t=0} t + o(t) \\ \iff \quad &\phi_{\Delta_d}(t) - 1 = i\mathbb{E}[\Delta_d]t + o(t).\end{aligned}$$

The definition of Landau's notation provides us:

$$\forall \epsilon > 0,\ \exists \eta > 0,\ \text{st } \forall x \in]-\eta, \eta[,\ |o(x)| \leq \epsilon|x|.$$

By considering $\epsilon = 1$, there exists $\eta_1 > 0$ such that $\forall x \in]-\eta_1, \eta_1[$, $|o(x)| \leq |x|$. Since for any $\xi \in \mathbb{R}$ there exists $k_\xi \geq 1$ such that $]-2^{-k_\xi}|\xi|, 2^{-k_\xi}|\xi|[\ \subseteq\ ]-\eta_1, \eta_1[$, it follows that

$$\forall x \in]-|\xi|2^{-k_\xi},\ |\xi|2^{-k_\xi}\ [,\ |o(x)| \leq |x|.$$

Since $\forall k > k_\xi$, $\xi 2^{-k} \in\ ]-|\xi|2^{-k_\xi},\ |\xi|2^{-k_\xi}\ [$, one has that $\sum_{k>k_\xi}|o(\xi/2^k)| \leq \sum_{k>k_\xi}|\xi/2^k| = 2^{-k_\xi}|\xi|$. Thus, it follows that the series $\sum_{k\geq 1}|\phi_{\Delta_d}(\xi/2^k) - 1|$ is finite for any $\xi \in \mathbb{R}$:

$$\begin{aligned}\sum_{k\geq 1}|\phi_{\Delta_d}(\xi/2^k) - 1| &\leq i\mathbb{E}[\Delta_d]|\xi| + \sum_{k\geq 1}|o(\xi/2^k)| \\ &\leq i\mathbb{E}[\Delta_d]|\xi| + \sum_{k=1}^{k_\xi}|o(\xi/2^k)| + 2^{-k_\xi}|\xi| < \infty.\end{aligned}$$

Then one can show the uniform convergence of $(\sum_{k=1}^{n}\log(\phi_{\Delta_d}(\xi/2^k)))_n$ in a neighborhood of 0. First, one may proceed to a Taylor-Young development of $\log(\phi_{\Delta_d})$ in a neighborhood of zero, in which case the principal value of the complex logarithm is well defined since ϕ_{Δ_d} evaluated on such neighborhood is defined on a neighborhood of one:

$$\begin{aligned}&\log(\phi_{\Delta_d}(t)) = \log(\phi_{\Delta_d}(t))\Big|_{t=0} + \frac{i\mathbb{E}[\Delta_d e^{it\Delta_d}]}{\phi_{\Delta_d}(t)}\Bigg|_{t=0} t + o(t) \\ &\log(\phi_{\Delta_d}(t)) = i\mathbb{E}[\Delta_d]t + o(t).\end{aligned}$$

Then, one can notice that for $\delta > 0$ fixed, $\exists k_\delta \geq 1$ such that $\forall \xi \in]-\delta, \delta[,\ 2^{-k_\delta}|\xi| \leq 2^{-k_\delta}\delta \leq \eta_1$. As previously, it follows that,

$$\exists k_\delta \geq 1,\ \forall \xi \in]-\delta, \delta[,\ \forall n \geq k_\delta, \quad \sum_{k>n} |o(\xi/2^k)| \leq \sum_{k>n} |\xi/2^k| = 2^{-n}|\xi| \leq 2^{-n}\delta.$$

Thus for $\delta > 0$ fixed, $\exists k_\delta \geq 1$, such that $\forall \xi \in]-\delta, \delta[,\ \forall n \geq k_\delta$,

$$\begin{aligned}\left|\sum_{k>n} \log(\phi_{\Delta_d}(\xi/2^k))\right| &\leq \sum_{k>n} \left|\log(\phi_{\Delta_d}(\xi/2^k))\right| \\ &\leq 2^{-n}\mathbb{E}[\Delta_d]|\xi| + \sum_{k>n} |o(\xi/2^k)| \\ &\leq 2^{-n}\delta\left(\mathbb{E}[\Delta_d] + 1\right).\end{aligned}$$

Hence for $\delta > 0$ fixed, $\exists k_\delta \geq 1$, such that $\forall \epsilon > 0$, $\exists N \geq k_\delta$ such that $\forall n \geq N$, $2^{-n}\delta\left(\mathbb{E}[\Delta_d] + 1\right) \leq \epsilon$. This concludes that the series uniformly converges in a neighbourhood of 0. Since there exists a neighbourhood of zero where ϕ_{Δ_d} is continuous, it follows that the series $\sum_{k\geq 1} \log(\phi_{\Delta_d}(\xi/2^k))$ is continuous on that neighbourhood, and therefore at zero.

Thus, by Levy's theorem (see [S5] Chapter XV: 3, Theorem 2) since $\phi_{V_b^{(n)}}$ converges pointwise to a complex function continuous at zero, $V_b^{(n)}$ converges in law to a random variable V_b^∞.

Now we prove that the random variable V_b^∞ has a density. To this end, the proof is inspired by the approach presented in the Proposition 1 of [S9].
On one hand, one has for all $n \geq 1$ and for every continuous and bounded function g:

$$\mathbb{E}[g(V_b^{(n+1)})] = \mathbb{E}[\mathbb{E}[g(V_b^{(n+1)})|V_b^{(n)}]] \qquad (*)$$

with

$$\mathbb{E}[g(V_b^{(n+1)})|V_b^{(n)} = y] = \int_0^{+\infty} g\left(\frac{x+y}{2}\right) f_{\Delta_d}(x)dx.$$

Since $y \mapsto \mathbb{E}[g(V_b^{(n+1)})|V_b^{(n)} = y]$ is continuous and bounded for any continuous and bounded function g, it follows by the definition of the weak convergence (see [S3] Chapter 1: 4.) that

$$\lim_{n\to\infty} \mathbb{E}[\mathbb{E}[g(V_b^{(n+1)})|V_b^{(n)}]] = \int_0^{+\infty}\int_0^{+\infty} g\left(\frac{x+y}{2}\right) f_{\Delta_d}(x)dx\ \pi(dy),$$

with π the distribution of V_b^∞. Then, one can get through a change of variable and the application of Fubini's theorem:

$$\begin{aligned}\int_0^{+\infty}\int_0^{+\infty} g\left(\frac{x+y}{2}\right) f_{\Delta_d}(x)dx\ \pi(dy) &= 2\int_0^{+\infty}\int_{\frac{y}{2}}^{+\infty} g(z) f_{\Delta_d}(2z-y)dz\ \pi(dy) \\ &= 2\int_0^{+\infty} g(z)\int_0^{2z} f_{\Delta_d}(2z-y)\pi(dy)\ dz.\end{aligned}$$

On another hand, one has from the definition of the convergence in law of a random variable that for every continuous and bounded function g:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[g(V_b^{(n+1)})] &= \mathbb{E}[g(V_b^\infty)] \\ &= \int_0^{+\infty} g(z)\pi(dz).\end{aligned}$$

Thus from the equality $(*)$, by passing through the limit, one can notice that the distribution of V_b^∞ is absolutely continuous with respect to the Lebesgue measure and thus admits a density.

□

Supposing that $\mathbb{E}[\Delta_i] < \infty$, with a similar proof that the sequence of volume per origin at replication initiation $(\tilde{V}_i^{(n)})_{n\in\mathbb{N}}$ converges in law to a a random variable $\tilde{V}_i^\infty$.

Proposition 2. *According to the initiation adder, assuming that Δ_i is drawn from a density f_{Δ_i}, and such that $\mathbb{E}[\Delta_i] < \infty$, the sequence $(\tilde{V}_i^{(n)})_{n\in\mathbb{N}}$ converges in law to a random variable V_b^∞ which admits a density $f_{\tilde{V}_i^\infty}$.*

2.2 Convergence of the number of origins of replication

Proposition 3. *According to the DIAM, assuming that $V_b^{(0)}$ and $\tilde{V}_i^{(0)}$ are independent, the sequence $(\omega^{(n)})_{n\in\mathbb{N}*}$ converges in law towards a random variable ω_∞ such that $\forall k \in \mathbb{Z}$,*

$$\mathbb{P}\left[\log_2(\omega_\infty) \leq k\right] = \mathbb{P}\left[V_b^\infty < 2^{k+1}\tilde{V}_i^\infty\right]. \tag{8}$$

Proof. From the result provided by Proposition 1, one has $\forall n \geq 1$ and $\forall k \geq -n$,

$$\begin{aligned}
\mathbb{P}\left[\log_2(\omega^{(n)}) \leq k\right] &= \mathbb{P}\left[T_b^{(n)} < T_i^{(n+k)}\right] \\
&= \mathbb{P}\left[\exp\left(\int_0^{T_b^{(n)}} \lambda(s)ds\right) < \exp\left(\int_0^{T_i^{(n+k)}} \lambda(s)ds\right)\right] \\
&= \mathbb{P}\left[V_b^{(1)}\exp\left(\int_0^{T_b^{(n)}} \lambda(s)ds\right) < \tilde{V}_i^{(0)}\exp\left(\int_0^{T_b^{(1)}} \lambda(s)ds\right)\exp\left(\int_0^{T_i^{(n+k)}} \lambda(s)ds\right)\right] && \text{by (5)}, \\
&= \mathbb{P}\left[V_b^{(1)}\exp\left(\int_{T_b^{(1)}}^{T_b^{(n)}} \lambda(s)ds\right) < \tilde{V}_i^{(0)}\exp\left(\int_0^{T_i^{(n+k)}} \lambda(s)ds\right)\right] \\
&= \mathbb{P}\left[V_b^{(n)} < 2^{k+1}\tilde{V}_i^{(n+k)}\right] && \text{by (2), (4)}.
\end{aligned}$$

Since $V_b^{(0)}$ and $\tilde{V}_i^{(0)}$ are independent, the previous probability relates to an inequality between two independent, and convergent random variables. It follows for every k fixed that

$$\left(V_b^{(n)}, \tilde{V}_i^{(n+k)}\right) \xrightarrow{D}_{n\to\infty} \left(V_b^\infty, \tilde{V}_i^\infty\right).$$

Then, thanks to the continuous mapping theorem (see [S3] Chapter 1: 5.), we have:

$$V_b^{(n)} - 2^{k+1}\tilde{V}_i^{(n+k)} \xrightarrow{D}_{n\to\infty} V_b^\infty - 2^{k+1}\tilde{V}_i^\infty.$$

Then, since $\tilde{V}_i^\infty$ and V_b^∞ have a density and are independent, we have thanks to the Portemanteau theorem (see [S3] Chapter 1: 4. Theorem 2.1) that

$$\lim_{n\to\infty} \mathbb{P}\left[V_b^{(n)} - 2^{k+1}\tilde{V}_i^{(n+k)} < 0\right] = \mathbb{P}[V_b^\infty - 2^{k+1}\tilde{V}_i^\infty < 0] =: p_k.$$

Since, $\forall n \in \mathbb{N}^*$ and $\forall k \geq -n$, $\mathbb{P}\left[\log_2(\omega^{(n)}) = k\right] = \mathbb{P}\left[\log_2(\omega^{(n)}) \leq k\right] - \mathbb{P}\left[\log_2(\omega^{(n)}) \leq k-1\right]$, it follows that for every $k \in \mathbb{Z}$

$$\lim_{n\to\infty} \mathbb{P}\left[\log_2(\omega^{(n)}) = k\right] = p_k - p_{k-1}.$$

One can integrate this quantity on $\mathbb{Z}$ as following:

$$\begin{aligned}\sum_{k\in\mathbb{Z}}(p_k - p_{k-1}) &= \sum_{k\in\mathbb{N}}(p_k - p_{k-1}) + \sum_{k\in\mathbb{N}^*}(p_{-k} - p_{1-k}) \\ &= \lim_{k\to\infty} p_k - \lim_{k\to\infty} p_{-k}\end{aligned}$$

Since V_b^∞ and $\tilde{V}_i^\infty$ are positive and admit densities, it follows that $\lim_{k\to\infty} p_k = \mathbb{P}[(\tilde{V}_i^\infty/V_b^\infty) > 0] = 1$ and $\lim_{k\to\infty} p_{-k} = \mathbb{P}[V_b^\infty/\tilde{V}_i^\infty \leq 0] = 0$. This concludes that $(\omega^{(n)})_{n\in\mathbb{N}*}$ converges in law towards a random variable ω_∞ distributed on $\{2^k : k \in \mathbb{Z}\}$ which has a cumulative probability function given by (8). □

3 Necessary and sufficient conditions to avoid one-origin-cell division in the DIAM

Theorem 1. *The DIAM ensures $\mathbb{P}[\omega_\infty \geq 1] = 1$ if and only if there exists $\sigma > 0$ such that:*

$$Supp(f_{\Delta_i}) \subset [0, \sigma], \qquad \textit{and} \qquad Supp(f_{\Delta_d}) \subset [\sigma, +\infty[.$$

Proof. In order to have a number of origins greater or equal to one almost surely, we aim to deduce necessary and sufficient constraints to impose to Δ_i and Δ_d, satisfying $\mathbb{P}[\omega_\infty \geq 1] = 1.$. From (8), this equality is equivalent to

$$\mathbb{P}\left[\tilde{V}_i^\infty \leq V_b^\infty\right] = 1. \qquad (*)$$

- First we show that $(*)$ is equivalent to

$$\exists \sigma > 0, \quad \mathbb{P}\left[V_b^\infty \geq \sigma\right] = 1 \quad \text{and} \quad \mathbb{P}\left[\tilde{V}_i^\infty \leq \sigma\right] = 1.$$

  - $\Leftarrow$ It is immediate that if there exists $\sigma > 0$ such that $V_b^\infty \geq \sigma$ almost surely and $\tilde{V}_i^\infty \leq \sigma$ almost surely then $\tilde{V}_i^\infty \leq \sigma \leq V_b^\infty$ almost surely.
  - $\Rightarrow$ Now we show that if $\tilde{V}_i^\infty \leq V_b^\infty$ almost surely then there exists a $\sigma > 0$ such that $\tilde{V}_i^\infty \leq \sigma \leq V_b^\infty$ almost surely. Since V_b^∞ is a positive random variable, one may introduce $\sigma^* := \inf Supp(f_{V_b^\infty}) \geq 0$, with $f_{V_b^\infty}$ the density of V_b^∞ defined by (**??**).
    Since V_b^∞ and $\tilde{V}_i^\infty$ are independent, it follows for every $\epsilon > 0$,

$$\begin{aligned}\mathbb{P}[V_b^\infty \leq \tilde{V}_i^\infty] &\geq \mathbb{P}[V_b^\infty \leq \sigma^* + \epsilon \leq \tilde{V}_i^\infty] \\ &\geq \mathbb{P}[V_b^\infty \leq \sigma^* + \epsilon]\mathbb{P}[\sigma^* + \epsilon \leq \tilde{V}_i^\infty].\end{aligned}$$

    By definition, $\mathbb{P}[V_b^\infty \leq \sigma^* + \epsilon] > 0$ for every $\epsilon > 0$. Hence, if $\mathbb{P}[V_b^\infty \leq \tilde{V}_i^\infty] = 0$, then $\mathbb{P}[\sigma^* + \epsilon \leq \tilde{V}_i^\infty] = 0$ for every $\epsilon > 0$ and therefore $\tilde{V}_i^\infty \leq \sigma^*$ almost surely. Finally, since $\tilde{V}_i^\infty$ is positive almost surely, one has necessarily $\sigma^* > 0$.

- Now, we show that in order to have $\mathbb{P}\left[V_b^\infty \geq \sigma\right] = 1$, it is necessary and sufficient to have $\mathbb{P}\left[\Delta_d \geq \sigma\right] = 1$.

– $\Leftarrow$: We suppose that $\mathbb{P}[\Delta_d \geq \sigma] = 1$. Since $\mathbb{P}[\bigcup_{k\geq 1}\{\Delta_d^{(k)} < \sigma\}] \leq \sum_{k\geq 1}\mathbb{P}[\Delta_d^{(k)} < \sigma] = 0$ it follows that $\mathbb{P}[\bigcap_{k\geq 1}\{\Delta_d^{(k)} \geq \sigma\}] = 1$. Since $\bigcap_{k\geq 1}\{\Delta_d^{(k)} \geq \sigma\} \subseteq \{\sum_{k\geq 1}\frac{\Delta_d^{(k)}}{2^k} \geq \sum_{k\geq 1}\frac{\sigma}{2^k}\}$,

$$\mathbb{P}\left[\sum_{k\geq 1}\frac{\Delta_d^{(k)}}{2^k} \geq \sigma\right] = 1.$$

As $V_b^\infty \overset{D}{=} \sum_{k\geq 1}\frac{\Delta_d^{(k)}}{2^k}$ from Proposition 2, one thus has $\mathbb{P}\left[\sum_{k\geq 1}\frac{\Delta_d^{(k)}}{2^k} \geq \sigma\right] = \mathbb{P}[V_b^\infty \geq \sigma] = 1$.

– $\Rightarrow$: By contraposition, we aim to show that $\mathbb{P}[\Delta_d < \sigma] > 0$ implies that $\mathbb{P}[V_b^\infty < \sigma] > 0$.
Note that if $\mathbb{P}[\Delta_d < \sigma] > 0$, then there exists $\epsilon > 0$ such that $\mathbb{P}[\Delta_d \leq \sigma - \epsilon] > 0$.
One may start by noticing that for any $1 \leq m < M$, and for any $\epsilon > 0$

$$\left\{\sum_{k=1}^{m}\frac{\Delta_d^{(k)}}{2^k} \leq \sigma - \epsilon\right\} \cap \left\{\sum_{k=m+1}^{M}\frac{\Delta_d^{(k)}}{2^k} < \epsilon\right\} \subseteq \left\{\sum_{k=1}^{M}\frac{\Delta_d^{(k)}}{2^k} < \sigma\right\}.$$

Since $(\Delta_d^{(n)})_{n\in\mathbb{N}^*}$ is a sequence of independent random variables, it follows that for any $1 \leq m < M$ and any $\epsilon > 0$,

$$\mathbb{P}\left[\sum_{k=1}^{M}\frac{\Delta_d^{(k)}}{2^k} < \sigma\right] \geq \mathbb{P}\left[\sum_{k=1}^{m}\frac{\Delta_d^{(k)}}{2^k} \leq \sigma - \epsilon\right]\mathbb{P}\left[\sum_{k=m+1}^{M}\frac{\Delta_d^{(k)}}{2^k} < \epsilon\right]. \qquad (**)$$

The aim in the following is to show that the left hand side of $(**)$ is positive when $M \to \infty$ if there exists $\epsilon > 0$ such that $\mathbb{P}[\Delta_d \leq \sigma - \epsilon] > 0$.

Firstly, we show that the first term of the right-hand side of $(**)$ is positive for any $m \geq 1$ when ϵ satisfies $\mathbb{P}[\Delta_d \leq \sigma - \epsilon] > 0$. Indeed, one may notice that

$$\bigcap_{k=1}^{m}\left\{\Delta_d^{(k)} \leq \sigma - \epsilon\right\} \subseteq \left\{\sum_{k=1}^{m}\frac{\Delta_d^{(k)}}{2^k} \leq \sum_{k=1}^{m}\frac{\sigma - \epsilon}{2^k}\right\} \subseteq \left\{\sum_{k=1}^{m}\frac{\Delta_d^{(k)}}{2^k} \leq \sigma - \epsilon\right\}.$$

Since $(\Delta_d^{(n)})_{n\in\mathbb{N}^*}$ is a sequence of independent random variables, if ϵ satisfies $\mathbb{P}[\Delta_d \leq \sigma - \epsilon] > 0$, then for any $m \geq 1$

$$\mathbb{P}\left[\sum_{k=1}^{m}\frac{\Delta_d^{(k)}}{2^k} \leq \sigma - \epsilon\right] \geq \mathbb{P}[\Delta_d \leq \sigma - \epsilon]^m > 0.$$

Secondly, we show that for any $\epsilon > 0$, there exists a $m \geq 1$ such that the second term of the right-hand side of $(**)$ is positive when $M \to \infty$.
First, we have by the Markov inequality for any $\epsilon > 0$ and any $1 \leq m < M$,

$$\mathbb{P}\left[\sum_{k=m+1}^{M}\frac{\Delta_d^{(k)}}{2^k} \geq \epsilon\right] \leq \frac{\mathbb{E}[\sum_{k=m+1}^{M}\frac{\Delta_d^{(k)}}{2^k}]}{\epsilon} = \mathbb{E}[\Delta_d]\frac{1 - \frac{1}{2^{M-m}}}{2^m\epsilon}.$$

It follows that for any $\epsilon > 0$, and any $m \geq 1$

$$\lim_{M\to\infty}\mathbb{P}\left[\sum_{k=m+1}^{M}\frac{\Delta_d^{(k)}}{2^k} \geq \epsilon\right] \leq \frac{\mathbb{E}[\Delta_d]}{2^m\epsilon}.$$

Since $\mathbb{E}[\Delta_d] < \infty$, for any $\epsilon > 0$, there exists $m \geq 1$ such that $\mathbb{E}[\Delta_d]/2^m\epsilon < 1$. Hence, it follows for any $\epsilon > 0$, there exists $m \geq 1$ such that

$$\lim_{M\to\infty} \mathbb{P}\left[\sum_{k=m+1}^{M} \frac{\Delta_d^{(k)}}{2^k} < \epsilon\right] > 0.$$

In summary, it follows that if there exists any $\epsilon > 0$ satisfying $\mathbb{P}\left[\Delta_d \leq \sigma - \epsilon\right] > 0$, then

$$\exists m \geq 1, \quad \lim_{M\to\infty} \left(\mathbb{P}\left[\sum_{k=1}^{m} \frac{\Delta_d^{(k)}}{2^k} \leq \sigma - \epsilon\right] \mathbb{P}\left[\sum_{k=m+1}^{M} \frac{\Delta_d^{(k)}}{2^k} < \epsilon\right]\right) > 0.$$

Regarding the limit of the left hand side of $(**)$, one has that $\sum_{k=1}^{M} \frac{\Delta_d^{(k)}}{2^k}$ converges in law to the random variable V_b^∞ defined in Proposition 2 which admits a density. This convergence in law is equivalent to the convergence of the cumulative distribution functions of $\sum_{k=1}^{M} \frac{\Delta_d^{(k)}}{2^k}$ to the continuous cumulative distribution function of V_b^∞ (see [S5] Chapter VIII: 1. Theorem 1), giving us for any σ

$$\lim_{M\to\infty} \mathbb{P}\left[\sum_{k=1}^{M} \frac{\Delta_d^{(k)}}{2^k} < \sigma\right] = \mathbb{P}\left[V_b^\infty < \sigma\right].$$

In conclusion, by passing the inequality $(**)$ to the limit when $M \to \infty$, one finally gets

$$\exists \epsilon > 0,\ \mathbb{P}\left[\Delta_d \leq \sigma - \epsilon\right] > 0 \implies \mathbb{P}\left[V_b^\infty < \sigma\right] > 0.$$

We thus showed that $\mathbb{P}\left[V_b^\infty \geq \sigma\right] = 1$ implies that $\mathbb{P}\left[\Delta_d \geq \sigma\right] = 1$ and reciprocally, which concludes the equivalence of these two assertions.

- Finally, we show that in order to satisfy $\mathbb{P}\left[\tilde{V}_i^\infty \geq \sigma\right] = 0$ it is necessary and sufficient to have $\mathbb{P}\left[\Delta_i \geq \sigma\right] = 0$.

  - $\Leftarrow$ As in the previous proof, we assume that $\mathbb{P}\left[\Delta_i < \sigma\right] = 1$. Since $\mathbb{P}[\bigcup_{k\geq 1}\{\Delta_i^{(k)} \geq \sigma\}] \leq \sum_{k\geq 1} \mathbb{P}[\Delta_i^{(k)} \geq \sigma] = 0$ it follows that $\mathbb{P}[\bigcap_{k\geq 1}\{\Delta_i^{(k)} < \sigma\}] = 1$. Since $\bigcap_{k\geq 1}\{\Delta_i^{(k)} < \sigma\} \subseteq \{\sum_{k\geq 1} \frac{\Delta_i^{(k)}}{2^k} < \sum_{k\geq 1} \frac{\sigma}{2^k}\}$, it implies that

$$\mathbb{P}\left[\sum_{k\geq 1} \frac{\Delta_i^{(k)}}{2^k} < \sigma\right] = 1.$$

    As $\tilde{V}_i^\infty \overset{D}{=} \sum_{k\geq 1} \frac{\Delta_i^{(k)}}{2^k}$, $\mathbb{P}\left[\sum_{k\geq 1} \frac{\Delta_i^{(k)}}{2^k} < \sigma\right] = \mathbb{P}\left[\tilde{V}_i^\infty < \sigma\right] = 1$.

  - $\Rightarrow$ We now show that $\mathbb{P}[\tilde{V}_i^\infty < \sigma] = 1$ implies $\mathbb{P}[\Delta_i < \sigma] = 1$.

    First, since Δ_i is positive almost surely, the sequence of events $\left(\left\{\sum_{k=1}^{m} \frac{\Delta_i^{(k)}}{2^k} \geq \sigma\right\}\right)_{m\geq 1}$ is increasing. Thus, one has for any $1 \leq m < M$ that

$$\left\{\sum_{k=1}^{m} \frac{\Delta_i^{(k)}}{2^k} \geq \sigma\right\} \subseteq \left\{\sum_{k=1}^{M} \frac{\Delta_i^{(k)}}{2^k} \geq \sigma\right\}.$$

Concurrently, one has for every $m \geq 1$ that

$$\bigcap_{k=1}^{m} \left\{ \Delta_i^{(k)} \geq \frac{\sigma}{1-\frac{1}{2^m}} \right\} \subseteq \left\{ \sum_{k=1}^{m} \frac{\Delta_i^{(k)}}{2^k} \geq \frac{\sigma}{1-\frac{1}{2^m}} \sum_{k=1}^{m} \frac{1}{2^k} \right\} = \left\{ \sum_{k=1}^{m} \frac{\Delta_i^{(k)}}{2^k} \geq \sigma \right\}.$$

Since the sequence $(\Delta_i^{(k)})_{k\geq 1}$ is independent, one thus has for every $1 \leq m < M$ that

$$\mathbb{P}\left[\Delta_i \geq \frac{\sigma}{1-\frac{1}{2^m}} \right]^m \leq \mathbb{P}\left[\sum_{k=1}^{M} \frac{\Delta_i^{(k)}}{2^k} \geq \sigma \right].$$

Then, one can study the limit of the right-hand side as $M \to \infty$. As previously, one has that the cumulative distribution function of $\sum_{k=1}^{M} \frac{\Delta_i^{(k)}}{2^k}$ converges to the continuous cumulative distribution function of $\tilde{V}_i^{\infty}$ leading for every $m \geq 1$ to

$$\mathbb{P}\left[\Delta_i \geq \frac{\sigma}{1-\frac{1}{2^m}} \right]^m \leq \mathbb{P}\left[\tilde{V}_i^{\infty} \geq \sigma \right].$$

It finally turns out that if $\mathbb{P}\left[\tilde{V}_i^{\infty} \geq \sigma\right] = 0$ then $\mathbb{P}\left[\Delta_i \geq \frac{\sigma}{1-\frac{1}{2^m}}\right] = 0$ for every $m \geq 1$. Since for every $\epsilon > 0$ there exists $m \geq 1$ such that $\sigma + \epsilon \geq \frac{\sigma}{1-\frac{1}{2^m}}$ implying that $\mathbb{P}\left[\Delta_i \geq \sigma + \epsilon\right] \leq \mathbb{P}\left[\Delta_i \geq \frac{\sigma}{1-\frac{1}{2^m}}\right]$, which concludes that

$$\mathbb{P}\left[\tilde{V}_i^{\infty} \geq \sigma\right] = 0 \implies \forall \epsilon > 0,\ \mathbb{P}\left[\Delta_i \geq \sigma + \epsilon\right] = 0.$$

□

4 Necessary and sufficient conditions to prevent a daughter cell from dividing before its mother in the RDAM

The Replication Double Adders Model (RDAM) proposed by [S15] assumes an initiation adder and thus that the volume per origin at replication initiation follows the recurrent equation (7). Then, the RDAM assumes that the sequence of volume at birth satisfies for all n

$$V_b^{(n+1)} = \tilde{V}_i^{(n)} + \Delta_{id}^{(n)}, \tag{9}$$

with $(\Delta_{id}^{(n)})_{n\geq 0}$ being a i.i.d. random variables sequence with density $f_{\Delta_{id}}$.

Remark: As $\tilde{V}_i^{(n)}$ and $\Delta_{id}^{(n)}$ are independent and admit both densities for all $n \geq 0$ it turns out that $V_b^{(n)}$ admits also a density for all $n \geq 1$. Also, from Proposition 3, it follows that $(V_b^{(n)})_{n\geq 1}$ converges in distribution to a random variable $V_b^{\infty} \overset{D}{=} \tilde{V}_i^{\infty} + \Delta_{id}$.

Proposition 4. *In the RDAM, $(T_b^{(n)})_{n\in\mathbb{N}*}$ is almost surely non-decreasing if and only if there exist $(\sigma_1, \sigma_2) \in \mathbb{R}_+^2 \setminus \{(0,0)\}$, such that*

$$Supp(f_{\Delta_{id}}) \subseteq [\sigma_1, 2\sigma_1 + \sigma_2], \qquad \text{and} \qquad Supp(f_{\Delta_i}) \subseteq [\sigma_2, \infty[.$$

Proof. We have for all $n \in \mathbb{N}$

$$\begin{aligned}
& T_b^{(n+1)} \leq T_b^{(n+2)} \quad a.s. \\
\iff \quad & V_b^{(n+1)} \leq 2V_b^{(n+2)} \quad a.s. && \text{by (2)} \\
\iff \quad & \tilde{V}_i^{(n)} + \Delta_{id}^{(n)} \leq 2(\tilde{V}_i^{(n+1)} + \Delta_{id}^{(n+1)}) \quad a.s. && \text{by (9)} \\
\iff \quad & \Delta_{id}^{(n)} \leq 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)} \quad a.s. && \text{by (7)}
\end{aligned}$$

We now show that in order to have $\Delta_{id}^{(n)} \leq 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)}$ almost surely, it is necessary and sufficient to have the existence of $(\sigma_1, \sigma_2) \in \mathbb{R}_+^2 \setminus \{(0,0)\}$ such that $\sigma_1 \leq \Delta_{id} \leq 2\sigma_1 + \sigma_2$ almost surely and $\sigma_2 \leq \Delta_i$ almost surely.

- $\Leftarrow$ Supposing that there exist $(\sigma_1, \sigma_2) \in \mathbb{R}_+^2 \setminus \{(0,0)\}$ such that $\sigma_1 \leq \Delta_{id} \leq 2\sigma_1 + \sigma_2$ and $\sigma_2 \leq \Delta_i$ almost surely, it follows immediately that

$$\Delta_{id}^{(n)} \leq 2\sigma_1 + \sigma_2 \leq 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)} \quad a.s.$$

- $\Rightarrow$ Since Δ_{id} and Δ_i are positive almost surely, there exist $\sigma_1^* := \inf Supp(f_{\Delta_{id}}) \geq 0$ and $\sigma_2^* := \inf Supp(f_{\Delta_i}) \geq 0$. We now want to show that if $\mathbb{P}[\Delta_{id}^{(n)} \leq 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)}] = 1$ therefore we have $\Delta_{id} \leq 2\sigma_1^* + \sigma_2^*$ almost surely.
  We have for every $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}[\Delta_{id}^{(n)} > 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)}] &\geq \mathbb{P}[\Delta_{id}^{(n)} > 2\sigma_1^* + \sigma_2^* + \epsilon \geq 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)}] \\
&\geq \mathbb{P}[\Delta_{id}^{(n)} > 2\sigma_1^* + \sigma_2^* + \epsilon]\mathbb{P}[\Delta_{id}^{(n)} \leq \sigma_1^* + \frac{\epsilon}{4}]\mathbb{P}[\Delta_i^{(n)} \leq \sigma_2^* + \frac{\epsilon}{4}].
\end{aligned}$$

  By definition, we have for every $\epsilon > 0$ that $\mathbb{P}[\Delta_{id}^{(n)} \leq \sigma_1^* + \frac{\epsilon}{4}] > 0$ and $\mathbb{P}[\Delta_i^{(n)} \leq \sigma_2^* + \frac{\epsilon}{4}] > 0$. In consequence, if $\mathbb{P}[\Delta_{id}^{(n)} > 2\Delta_{id}^{(n+1)} + \Delta_i^{(n)}] = 0$ then $\mathbb{P}[\Delta_{id}^{(n)} > 2\sigma_1^* + \sigma_2^* + \epsilon] = 0$ for every $\epsilon > 0$. Since Δ_{id} is almost surely positive, one necessarily has $(\sigma_1^*, \sigma_2^*) \neq (0,0)$.

In conclusion, in order to have $(T_b^{(n)})_{n \in \mathbb{N}^*}$ non-decreasing almost surely, it is necessary and sufficient to have $(\sigma_1, \sigma_2) \in \mathbb{R}_+^2 \setminus \{(0,0)\}$ such that $\sigma_1 \leq \Delta_{id} \leq 2\sigma_1 + \sigma_2$ almost surely and $\sigma_2 \leq \Delta_i$ almost surely. □

5 Conditional densities of the models and identifiablility

In this section, we present results for two other models: the Concurrent Processes Model (CPM) proposed by [S10, S11] and the Concurrent Adders Model (CAM) first proposed in this work.

For the CPM, the sequence of volume at birth $(V_b^{(n)})_{n \geq 0}$ follows $\forall n \geq 0$ the recurrent expression

$$V_b^{(n+1)} = \max\Big(\tilde{V}_i^{(n)} e^{\Lambda^{(n)} R^{(n)}},\ 2^{-1}(V_b^{(n)} + \Delta_d^{(n)})\Big), \tag{10}$$

with $(\Delta_d^{(n)})_{n \geq 0}$ and $(R^{(n)})_{n \geq 0}$ independent and i.i.d. random variables sequences, and $\Lambda^{(n)} := \frac{1}{R^{(n)}} \int_{T_i^{(n)}}^{T_i^{(n)} + R^{(n)}} \lambda(s) ds$, assumed to be independent from $(R^{(n)})_{n \geq 0}$.

Then for the CAM, the sequence of volume at birth $(V_b^{(n)})_{n \geq 0}$ follows $\forall n \geq 0$ the equation

$$V_b^{(n+1)} = \max\Big(\tilde{V}_i^{(n)} + \Delta_{id}^{(n)},\ 2^{-1}(V_b^{(n)} + \Delta_d^{(n)})\Big), \tag{11}$$

with $(\Delta_d^{(n)})_{n \geq 0}$ and $(\Delta_{id}^{(n)})_{n \geq 0}$ i.i.d. random variables sequences.

5.1 Expressions of the conditional densities

Proposition 5.

Let $p(\bullet|v_b, \tilde{v}_i, \lambda)$ be the conditional density of the division volume $2V_b^{(\bullet+1)}$ given that the triplet $(V_b^{(\bullet)}, \tilde{V}_i^{(\bullet)}, \Lambda^{(\bullet)})$ is equal to $(v_b, \tilde{v}_i, \lambda) \in \mathbb{R}_+^{*\,3}$. Let f_R, f_{Δ_d}, $f_{\Delta_{id}}$ be the densities and F_R, F_{Δ_d}, $F_{\Delta_{id}}$ the cumulative probability functions of the laws of R, Δ_d,and Δ_i respectively. For the different models, the conditional distribution of the division volume satisfies for the different models the following expressions for any $v > 0$:

- *The DIAM:*
$$p(v|v_b, \tilde{v}_i, \lambda) = f_{\Delta_d}(v - v_b), \tag{12a}$$
- *The RDAM:*
$$p(v|v_b, \tilde{v}_i, \lambda) = \frac{1}{2} f_{\Delta_{id}}\Big(\frac{v}{2} - \tilde{v}_i\Big), \tag{12b}$$
- *The CPM:*
$$p(v|v_b, \tilde{v}_i, \lambda) = f_{\Delta_d}(v - v_b) F_R\left(\frac{1}{\lambda}\log\left(\frac{v}{2\tilde{v}_i}\right)\right) + \frac{1}{\lambda v} f_R\left(\frac{1}{\lambda}\log\left(\frac{v}{2\tilde{v}_i}\right)\right) F_{\Delta_d}(v - v_b), \tag{12c}$$
- *The CAM:*
$$p(v|v_b, \tilde{v}_i, \lambda) = f_{\Delta_d}(v - v_b) F_{\Delta_{id}}\Big(\frac{v}{2} - \tilde{v}_i\Big) + \frac{1}{2} f_{\Delta_{id}}\Big(\frac{v}{2} - \tilde{v}_i\Big) F_{\Delta_d}(v - v_b), \tag{12d}$$

Proof. First, one may start from the definition of $p(\bullet|v_b, \tilde{v}, \lambda)$ for all $(v_b, \tilde{v}_i, \lambda) \in \mathbb{R}_+^{*\,3}$:

$$\forall v > 0, \quad \int_0^v p(x|v_b, \tilde{v}_i, \lambda)dx = \mathbb{P}[2V_b^{(\bullet+1)} \leq v | V_b^{(\bullet)} = v_b, \tilde{V}_i^{(\bullet)} = \tilde{v}_i, \Lambda^{(\bullet)} = \lambda].$$

One can express this quantity according to the different studied models.
For the DIAM:

$$\begin{aligned}\mathbb{P}[2V_b^{(\bullet+1)} \leq v | V_b^{(\bullet)} = v_b, \tilde{V}_i^{(\bullet)} = \tilde{v}_i, \Lambda^{(\bullet)} = \lambda] &= \mathbb{P}[\Delta_d + v_b \leq v] \\ &= F_{\Delta_d}(v - v_b).\end{aligned}$$

For the RDAM:

$$\begin{aligned}p[2V_b^{(\bullet+1)} \leq v | V_b^{(\bullet)} = v_b, \tilde{V}_i^{(\bullet)} = \tilde{v}_i, \lambda^{(\bullet)} = \lambda] &= \mathbb{P}[2(\Delta_{id} + \tilde{v}_i) \leq v] \\ &= F_{\Delta_{id}}\Big(\frac{v}{2} - \tilde{v}_i\Big).\end{aligned}$$

For the CPM:

$$\begin{aligned}\mathbb{P}[2V_b^{(\bullet+1)} \leq v | V_b^{(\bullet)} = v_b, \tilde{V}_i^{(\bullet)} = \tilde{v}_i, \lambda^{(\bullet)} = \lambda] &= \mathbb{P}[2\tilde{v}_i e^{\lambda R} \leq v, \Delta_d + v_b \leq v] \\ &= \mathbb{P}[R \leq \frac{1}{\lambda}\log(\frac{v}{2\tilde{v}_i}), \Delta_d \leq v - v_b] \\ &= F_R(\frac{1}{\lambda}\log(\frac{v}{2\tilde{v}_i}))F_{\Delta_d}(v - v_b). \quad \text{(by independence)}\end{aligned}$$

For the CAM:

$$\begin{aligned}\mathbb{P}[2V_b^{(\bullet+1)} \leq v | V_b^{(\bullet)} = v_b, \tilde{V}_i^{(\bullet)} = \tilde{v}_i, \lambda^{(\bullet)} = \lambda] &= \mathbb{P}[2(\tilde{v}_i + \Delta_{id}) \leq v, \Delta_d + v_b \leq v] \\ &= \mathbb{P}[\Delta_{id} \leq \frac{v}{2} - \tilde{v}_i, \Delta_d \leq v - v_b] \\ &= F_{\Delta_{id}}(\frac{v}{2} - \tilde{v}_i) F_{\Delta_d}(v - v_b) \quad \text{(by independence)}\end{aligned}$$

By taking the derivative of the obtained expressions according to v, one ends up with the expected results. □

5.2 Identifiability of the conditional densities of the CPM and CAM

While it is immediate that the distribution of Δ_d in the DIAM, and the distribution of Δ_{id} in the RDAM are identifiable as they are observable, it is less clear that the distributions of Δ_d, R and Δ_{id} are identifiable in the CPM and the CAM, as these are latent variables. In the following, we investigate this question and show that under specific conditions over the observed covariates in the datasets, the latent variables in the CAM and CPM are identifiable.

Let us consider the general case where we have a collection of n realizations of a random variable Y with values in $\mathcal{Y}$ that is the maximum between two random variables such that

$$Y := \max\Big(h_X(Z_1),\ g_X(Z_2)\Big), \tag{13}$$

where X is a collection of covariates that takes values in $\mathcal{X} \subseteq \mathbb{R}^d$, Z_1 and Z_2 are independent random variables distributed according to parametric cumulative distribution functions $F^1_{\theta_1}$ and $F^2_{\theta_2}$ with parameters $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$. We assume that $F^1_{\theta_1}$ and $F^2_{\theta_2}$ are positive on $(0, +\infty)$ and equal to zero on $(-\infty, 0]$ for any $(\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$, and that they belong to a class of distribution that are right-tail identifiable. We call F_θ a right-tail identifiable distribution if

$$\forall \theta, \alpha \in \Theta,\ \forall y \in \mathbb{R}, \quad \{\forall x > y,\ F_\theta(x) = F_\alpha(x) \implies \theta = \alpha\}.$$

Then, we assume that $\forall x \in \mathcal{X},\ h_x : \mathbb{R} \to \mathcal{H} \supseteq \mathcal{Y}$ and $g_x : \mathbb{R} \to \mathcal{G} \supseteq \mathcal{Y}$ are bijective, increasing, and satisfy $h_x(\mathbb{R}_+) \subseteq \mathcal{Y}$ and $g_x(\mathbb{R}_+) \subseteq \mathcal{Y}$.

For all $\theta \in \Theta_1 \times \Theta_2$, we denote $p^\theta_{Y|X}(\bullet|x)$ the conditional density of $Y|X = x$ which allows us to define the conditional likelihood over a sample of n independent realizations of Y conditionally to n realizations of X:

$$\begin{aligned} L^n : \mathcal{Y}^n \times \mathcal{X}^n \times \Theta_1 \times \Theta_2 &\longrightarrow \mathbb{R}_+ \\ (y_1, ..., y_n, x_1, ..., x_n, \theta) &\longmapsto \prod_{k=1}^n p^\theta_{Y|X}(y_k|x_k). \end{aligned}$$

We consider that the distributions of Z_1 and Z_2 are identifiable for a set of covariates defined on a space $E \subseteq \mathcal{X}^n$ if

$$\forall\ \mathbf{x} \in E,\ \forall\ \theta, \alpha \in \Theta_1 \times \Theta_2, \quad L^n(\bullet, \mathbf{x}, \theta) = L^n(\bullet, \mathbf{x}, \alpha) \implies \theta = \alpha.$$

In the following, we study a sufficient condition on E in order to have Z_1 and Z_2 identifiable.

Proposition 6. *Let* $E := \{(x_1, ..., x_n) \in \mathcal{X}^n : \ \exists\ i, j,\ \forall z > 0,\ g_{x_i}^{-1} \circ h_{x_i}(z) < g_{x_j}^{-1} \circ h_{x_j}(z)\}$ *be the observed covariates space, with* $n \geq 2$,

$$\forall\ \boldsymbol{x} \in E,\ \forall\ \theta, \alpha \in \Theta_1 \times \Theta_2, \quad L^n(\bullet, \boldsymbol{x}, \theta) = L^n(\bullet, \boldsymbol{x}, \alpha) \implies \theta = \alpha.$$

Proof. If $\forall\ \mathbf{x} := (x_1, ..., x_n) \in E,\ \forall\ \theta, \alpha \in \Theta_1 \times \Theta_2, \quad L^n(\bullet, \mathbf{x}, \theta) = L^n(\bullet, \mathbf{x}, \alpha)$ then $\forall\ (y_1, ..., y_n) \in \mathcal{Y}^n$,

$$\int_{-\infty}^{y_1} ... \int_{-\infty}^{y_n} L^n(\mathbf{x}, \mathbf{y}, \theta) d\mathbf{y} = \int_{-\infty}^{y_1} ... \int_{-\infty}^{y_n} L^n(\mathbf{x}, \mathbf{y}, \alpha) d\mathbf{y}$$

$$\iff \quad \prod_{k=1}^{n} \mathbb{P}_\theta[Y \leq y_k | X = x_k] = \prod_{k=1}^{n} \mathbb{P}_\alpha[Y \leq y_k | X = x_k]$$

$$\iff \quad \prod_{k=1}^{n} \mathbb{P}_\theta[Z_1 \leq h_{x_k}^{-1}(y_k)]\mathbb{P}_\theta[Z_2 \leq g_{x_k}^{-1}(y_k)] = \prod_{k=1}^{n} \mathbb{P}_\alpha[Z_1 \leq h_{x_k}^{-1}(y_k)]\mathbb{P}_\alpha[Z_2 \leq g_{x_k}^{-1}(y_k)],$$

by independence of Z_1 and Z_2.

We then denote $F_{\theta_1}^1$, $F_{\alpha_1}^1$, $F_{\theta_2}^2$ and $F_{\alpha_2}^2$ the cumulative distributions of Z_1 and Z_2, with $(\theta_1, \theta_2) := \theta$ and $(\alpha_1, \alpha_2) := \alpha$, and then consider the equation when it is positive, namely
$\forall\ (y_1, ..., y_n) \in \mathcal{Y}_+^n := \{(y_1, ..., y_n) \in \mathcal{Y}^n : \forall k,\ y_k > h_{x_k}(0) \vee g_{x_k}(0)\}$, in which case we have,

$$\prod_{k=1}^{n} \frac{F_{\theta_1}^1(h_{x_k}^{-1}(y_k))}{F_{\alpha_1}^1(h_{x_k}^{-1}(y_k))} = \prod_{k=1}^{n} \frac{F_{\alpha_2}^2(g_{x_k}^{-1}(y_k))}{F_{\theta_2}^2(g_{x_k}^{-1}(y_k))}.$$

To simplify, we denote $H^1 := F_{\theta_1}^1 / F_{\alpha_1}^1$ and $H^2 := F_{\alpha_2}^2 / F_{\theta_2}^2$. The objective in the following is to show that $\exists y,\ \forall x > y,\ H^1(x) = H^2(x) = 1$ which will conclude that $\theta = \alpha$ by the right-tail identifiability property of F^1 and F^2. One can rewrite the previous equation as following:

$$\prod_{k=1}^{n} H^1 \circ h_{x_k}^{-1}(y_k) = \prod_{k=1}^{n} H^2 \circ g_{x_k}^{-1}(y_k).$$

This equality is equivalent for any $1 \leq k \leq n$ to,

$$H^1 \circ h_{x_k}^{-1}(y_k) = H^2 \circ g_{x_k}^{-1}(y_k) \frac{\prod_{j \neq k} H^2 \circ g_{x_j}^{-1}(y_j)}{\prod_{j \neq k} H^1 \circ h_{x_j}^{-1}(y_j)}.$$

As this equality is true for any $(y_1, ..., y_n) \in \mathcal{Y}_+^n$ it follows that,

$$\forall\ k, \quad \forall\ y > h_{x_k}(0) \vee g_{x_k}(0), \quad H^1 \circ h_{x_k}^{-1}(y) \propto H^2 \circ g_{x_k}^{-1}(y).$$

As both H^1 and H^2 converge to one at $+\infty$, it turns out that

$$\forall\ k, \quad \forall\ y > h_{x_k}(0) \vee g_{x_k}(0), \quad H^1 \circ h_{x_k}^{-1}(y) = H^2 \circ g_{x_k}^{-1}(y). \qquad (*)$$

From $(*)$ one can deduce the following assertion:

$$\forall\ k, l, \quad \forall\ z > 0 \vee h_{x_k}^{-1} \circ g_{x_k}(0) \vee h_{x_l}^{-1} \circ g_{x_l}(0), \quad H^1(z) = H^2 \circ g_{x_k}^{-1} \circ h_{x_k}(z) = H^2 \circ g_{x_l}^{-1} \circ h_{x_l}(z).$$

Since $\forall\ (x_1, ..., x_n) \in E,\ \exists\ i, j,\ \forall z > 0,\ g_{x_i}^{-1} \circ h_{x_i}(z) < g_{x_j}^{-1} \circ h_{x_j}(z)$, it follows that for these particular i, j,

$$\forall\ z > 0 \vee g_{x_i}^{-1} \circ h_{x_i}(0), \qquad H^2(z) = H^2 \circ \varphi(z), \quad \text{with} \quad \varphi(z) := g_{x_j}^{-1} \circ h_{x_j} \circ h_{x_i}^{-1} \circ g_{x_i}(z) > z.$$

On one hand this provides us that $\forall n,\ H^2 = H^2 \circ \varphi^{\circ n}$ (with $\varphi^{\circ n} = \varphi \circ ... \circ \varphi$ n times) and on another hand that $\lim_{n \to \infty} \varphi^{\circ n}(z) = +\infty$. As H^2 converges to one at $+\infty$, this concludes that $\forall z > 0 \vee g_{x_i}^{-1} \circ h_{x_i}(0),\ H^2(z) = 1$, and equivalently $\forall z > 0 \vee h_{x_i}^{-1} \circ g_{x_i}(0),\ H^2 \circ g_{x_j}^{-1} \circ h_{x_j}(z) = H^1(z) = 1$.

□

One can derive the corollaries of this result by defining the sufficient conditions to verify in the data in order to have the distributions of the latent variables in the CAM and the CPM identifiable.

In both models, the random variable Y refers to $V_b^{(\bullet+1)}$ supported on $\mathbb{R}_+^*$, and the covariate X refers to $(V_b^{(\bullet)}, \tilde{V}_i^{(\bullet)}, \lambda^{(\bullet)}) \in \mathbb{R}_+^{*\,3}$. For the CPM, Z_1 and Z_2 may refer to Δ_d and R which thus identify $\forall (v', \tilde{v}', \lambda) \in \mathbb{R}_+^{*\,3}$,

$$\begin{aligned} h_{(v',\tilde{v}',\lambda)} : \mathbb{R} &\to \mathbb{R}_+^* \\ x &\mapsto \tilde{v}' \exp(\lambda x) \end{aligned} \qquad \begin{aligned} g_{(v',\tilde{v}',\lambda)} : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 2^{-1}(x + v'). \end{aligned}$$

Since Δ_d and R are assumed to be distributed either according to a Generalized Gamma distribution, they satisfy the conditions verified by $F^1_{\theta_1}$ and $F^2_{\theta_2}$. Thus, according to Proposition 6, the distributions of Δ_d and R are identifiable if among the data there are $(v'_1, \tilde{v}'_1, \lambda_1)$, and $(v'_2, \tilde{v}'_2, \lambda_2)$ such that

$$\forall z > 0, \quad 2\tilde{v}'_1 e^{\lambda_1 z} - v'_1 < 2\tilde{v}'_2 e^{\lambda_2 z} - v'_2.$$

For the CAM, Z_1 and Z_2 may refer to Δ_d and Δ_{id} which thus identify $\forall (v', \tilde{v}', \lambda) \in \mathbb{R}_+^{*\,3}$,

$$\begin{aligned} h_{(v',\tilde{v}',\lambda)} : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x + \tilde{v}' \end{aligned} \qquad \begin{aligned} g_{(v',\tilde{v}',\lambda)} : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 2^{-1}(x + v'). \end{aligned}$$

As for the CPM, since Δ_d and Δ_{id} are distributed either according to a Generalized Gamma distribution one can apply the Proposition 6, and state that Δ_d and Δ_{id} are identifiable if among the data there are $(v'_1, \tilde{v}'_1, \lambda_1)$, and $(v'_2, \tilde{v}'_2, \lambda_2)$ such that

$$2\tilde{v}'_1 - v'_1 < 2\tilde{v}'_2 - v'_2.$$

In practice it is easy to verify in every dataset that there exist $(v'_1, \tilde{v}'_1, \lambda_1)$, and $(v'_2, \tilde{v}'_2, \lambda_2)$ such that $v'_2 < v'_1$, $\tilde{v}'_1 < \tilde{v}'_2$, and $\lambda_1 < \lambda_2$, which implies identifiability in both CPM and CAM.

5.3 Expectation-Maximization (EM) algorithm for parameters estimation of the CPM and the CAM

We provide here a method for estimating the parameters of the distributions of Z_1 and Z_2, involving an Expectation-Maximization (EM) algorithm. To do so, we take the general cas introduced in the previous section (Equation (13)), and assume that $\forall x$ g_x^{-1} and h_x^{-1} are derivable. First, one may start by noticing that,

$$\begin{aligned} \mathbb{P}[Y \leq y, h_X(Z_1) \leq g_X(Z_2) | X = x] &= \mathbb{P}[h_x(Z_1) \leq g_x(Z_2) \leq y] \\ &= \int_0^y (g_x^{-1})'(z) f^2_{\theta_2}(g_x^{-1}(z)) \mathbb{P}[h_x(Z_1) \leq z] dz \\ &= \int_0^y (g_x^{-1})'(z) f^2_{\theta_2}(g_x^{-1}(z)) F^1_{\theta_1}(h_x^{-1}(y)) dz \end{aligned}$$

and similarly that,

$$\begin{aligned}\mathbb{P}[Y \leq y, g_X(Z_2) \leq h_X(Z_1)|X = x] &= \mathbb{P}[g_x(Z_2) \leq h_x(Z_1) \leq y] \\ &= \int_0^y (h_x^{-1})'(z) f^1_{\theta_1}(h_x^{-1}(z)) \mathbb{P}[g_x(Z_2) \leq z] dz \\ &= \int_0^y (h_x^{-1})'(z) f^1_{\theta_1}(h_x^{-1}(z)) F^2_{\theta_2}(g_x^{-1}(y)) dz.\end{aligned}$$

One then can introduce a random variable W taking values in $\{1, 2\}$, such that $\{W = 1\} = \{h_X(Z_1) \geq g_X(Z_2)\}$ and $\{W = 2\} = \{h_X(Z_1) < g_X(Z_2)\}$. By introducing $p^\theta_{Y,W|X}$ the joint distribution of (Y, W) conditionally to X, one can deduce from the previous equalities that it satisfies,

$$p^\theta_{Y,W|X}(y, w|x) = \begin{cases} (h_x^{-1})'(y)\ f^1_{\theta_1}(h_x^{-1}(y))\ F^2_{\theta_2}(g_x^{-1}(y)) & \text{if } w = 1 \\ (g_x^{-1})'(y)\ f^2_{\theta_2}(g_x^{-1}(y))\ F^1_{\theta_1}(h_x^{-1}(y)) & \text{if } w = 2. \end{cases}$$

Note that $p^\theta_{Y|X}(\bullet|x)$, the previously introduced conditional distribution of Y given $X = x$, satisfies $p^\theta_{Y|X}(\bullet|x) = p^\theta_{Y,W|X}(\bullet, 1|x) + p^\theta_{Y,W|X}(\bullet, 2|x)$. By taking L^N the likelihood previously defined, one can introduce the log-likelihood defined $\forall \mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n, \theta \in \Theta_1 \times \Theta_2$ by

$$\log L^N(\mathbf{y}, \mathbf{x}, \theta) = \sum_{k=1}^N \log p^\theta_{Y|X}(y_k|x_k).$$

Using Bayes formula, one can show for any $\theta' \in \Theta_1 \times \Theta_2$ that the log likelihood satisfies

$$\begin{aligned}\log L^N(\mathbf{y}, \mathbf{x}, \theta) &= \underbrace{\sum_{k=1}^N \mathbb{E}_{\theta'}\left[\log p^\theta_{Y,W|X}(Y, W|X)\big|X = x_k, Y = y_k\right]}_{Q(\theta;\theta')} - \underbrace{\sum_{k=1}^N \mathbb{E}_{\theta'}\left[\log p^\theta_{W|X,Y}(W|X, Y)\big|X = x_k, Y = y_k\right]}_{H(\theta;\theta')} \\ &\geq Q(\theta; \theta') - H(\theta'; \theta') \quad \text{(by Gibb's inequality).}\end{aligned}$$

with $p^\theta_{W|X,Y}(w|x, y) := \mathbb{P}_\theta[W = w|X = x, Y = y]$.

The fundamental principle of the EM algorithm is to iteratively maximize $\left(Q(\theta, \theta^{(h)})\right)_h$, such that $\theta^{(h+1)} = \arg\max_\theta Q(\theta; \theta^{(h)})$, which ensures the convergence of $(\theta^{(h)})_h$ to a local maximum of $\log L^N$ [S8]. In the following we detail $Q(\theta; \theta')$. To simplify, we denote for all k $\pi^{\theta,k}_w := p^\theta_{W|X,Y}(w|x_k, y_k)$:

$$
\begin{aligned}
Q(\theta;\theta') &= \sum_{k=1}^{N} \int \log p^{\theta}_{Y,W|X}(y_k, w|x_k)\ dp^{\theta'}_{W|X,Y}(w|x_k, y_k) \\
&= \sum_{k=1}^{N} \Big(\log p^{\theta}_{Y,W|X}(y_k, 1|x_k)\ \pi_1^{\theta',k} + \log p^{\theta}_{Y,W|X}(y_k, 2|x_k)\ \pi_2^{\theta',k} \Big) \\
&= \sum_{k=1}^{N} \Big(\ \pi_1^{\theta',k} \log \Big((h^{-1}_{x_k})'(y_k)\ f^1_{\theta_1}(h^{-1}_{x_k}(y_k)) \Big) + \pi_2^{\theta',k} \log F^1_{\theta_1}(h^{-1}_{x_k}(y_k))\ \Big) \\
&\quad + \sum_{k=1}^{N} \Big(\ \pi_1^{\theta',k} \log F^2_{\theta_2}(g^{-1}_{x_k}(y_k)) + \pi_2^{\theta',k} \log \Big((g^{-1}_{x_k})'(y_k)\ f^2_{\theta_2}(g^{-1}_{x_k}(y_k)) \Big) \ \Big) \\
&= \underbrace{\sum_{k=1}^{N} \Big(\ \pi_1^{\theta',k} \log f^1_{\theta_1}(h^{-1}_{x_k}(y_k)) + \pi_2^{\theta',k} \log F^1_{\theta_1}(h^{-1}_{x_k}(y_k))\ \Big)}_{Q^1(\theta_1;\theta')} + \sum_{k=1}^{N} \pi_1^{\theta',k} \log (h^{-1}_{x_k})'(y_k) \\
&\quad + \underbrace{\sum_{k=1}^{N} \Big(\ \pi_1^{\theta',k} \log F^2_{\theta_2}(g^{-1}_{x_k}(y_k)) + \pi_2^{\theta',k} \log f^2_{\theta_2}(g^{-1}_{x_k}(y_k))\ \Big)}_{Q^2(\theta_2;\theta')} + \sum_{k=1}^{N} \pi_2^{\theta',k} \log (g^{-1}_{x_k})'(y_k).
\end{aligned}
$$

One can observe that $Q(\theta;\theta')$ can be decomposed into the sum of three terms: a function of θ_1 denoted by $Q^1(\theta_1;\theta')$; a function of θ_2 denoted by $Q^2(\theta_2;\theta')$; and some constant. This allows us to maximize $Q(\theta,\theta')$ by independently maximizing $Q^1(\theta_1;\theta')$ with respect to θ_1 and $Q^2(\theta_2;\theta')$ with respect to θ_2. This provides the following EM algorithm to apply for the problem of interest.

Expectation step: Computation for all $1 \leq k \leq N$ of

$$
\pi_1^{\theta^{(h)},k} = \frac{p^{\theta^{(h)}}_{Y,W|X}(y_k, 1|x_k)}{p^{\theta^{(h)}}_{Y|X}(y_k|x_k)}, \qquad \pi_2^{\theta^{(h)},k} = 1 - \pi_1^{\theta^{(h)},k}.
$$

Maximization step: Computation of $\theta^{(h+1)} := (\theta_1^{(h+1)}, \theta_2^{(h+1)})$ with,

$$
\theta_1^{(h+1)} = \arg\max_{\theta_1 \in \Theta_1} Q^1(\theta_1;\theta^{(h)}), \qquad \theta_2^{(h+1)} = \arg\max_{\theta_2 \in \Theta_2} Q^2(\theta_2;\theta^{(h)}).
$$

In the context of the CPM and the CAM, the two latent variables Z_1 and Z_2 follow Generalized Gamma distributions. Considering this probability distribution in the expression of Q^1 and Q^2, prevents us from deriving an analytical solution for their maximization. However, these optimization problems remain much simpler than the original one and can be efficiently solved using numerical optimization algorithms such as Nelder-Mead (implemented in the package Scipy of python [S14]) and CMA-ES (implemented in the package cma in Python)

References

[S1] Lars Valerian Ahlfors and Lars V Ahlfors. *Complex analysis*. Vol. 3. McGraw-Hill New York, 1979.

[S2] Ariel Amir. "Cell size regulation in bacteria". In: *Physical review letters* 112.20 (2014), p. 208102.

[S3] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[S4] Manuel Campos et al. "A constant size extension drives bacterial cell size homeostasis". In: *Cell* 159.6 (2014), pp. 1433–1446.

[S5] William Feller. *An introduction to probability theory and its applications, Volume 2*. Vol. 81. John Wiley & Sons, 1991.

[S6] Pierre Gabriel and Hugo Martin. "Steady distribution of the incremental model for bacteria proliferation". In: *Networks and Heterogeneous Media* 14.1 (2019), pp. 149–171.

[S7] Ignacio Madrid. "Exponential ergodicity of a degenerate age-size piecewise deterministic process". In: *Acta Applicandae Mathematicae* 187.1 (2023), p. 5.

[S8] Geoffrey J McLachlan and Thriyambakam Krishnan. *The EM algorithm and extensions*. John Wiley & Sons, 2008.

[S9] Sylvie Méléard and Denis Villemonais. "Quasi-stationary distributions and population processes". In: (2012).

[S10] Gabriele Micali et al. "Concurrent processes set E. coli cell division". In: *Science Advances* 4.11 (2018), eaau3324.

[S11] Gabriele Micali et al. "Dissecting the control mechanisms for DNA replication and cell division in E. coli". In: *Cell reports* 25.3 (2018), pp. 761–771.

[S12] Fangwei Si et al. "Mechanistic origin of cell-size control and homeostasis in bacteria". In: *Current Biology* 29.11 (2019), pp. 1760–1770.

[S13] Sattar Taheri-Araghi et al. "Cell-size control and homeostasis in bacteria". In: *Current biology* 25.3 (2015), pp. 385–391.

[S14] Pauli Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature methods* 17.3 (2020), pp. 261–272.

[S15] Guillaume Witz, Erik van Nimwegen, and Thomas Julou. "Initiation of chromosome replication controls both division and replication cycles in E. coli through a double-adder mechanism". In: *Elife* 8 (2019), e48063.

Supplementary Figures

Table S1: **P-values of Shapiro–Wilk normality tests of cell quantities.** We performed normality tests on division, birth, and initiation sizes using the Shapiro–Wilk test. P-values greater than 0.01, indicating no significant deviation from a normal distribution, are shown in bold.

Datasets	**Division size**	**Birth size**	**Initiation size**
MG1655 acet	1.71e-30	4.17e-23	8.39e-08
MG1655 gly11aa	1.17e-25	1.64e-23	2.51e-11
MG1655 glu	3.60e-30	2.97e-29	3.08e-10
NCM3722 arg	9.37e-09	1.14e-12	9.82e-08
NCM3722 glu	1.18e-11	8.95e-12	3.21e-06
NCM3722 glu12aa	1.94e-23	2.32e-21	5.75e-20
BW27378 gly	1.03e-21	3.49e-27	1.30e-22
BW27378 glu	1.00e-25	5.80e-27	2.21e-19
BW27378 glu8a	8.81e-22	6.47e-20	1.23e-18
BW27783 acet	1.59e-07	3.09e-07	5.61e-07
BW27783 alaTrE	1.68e-06	2.06e-12	3.22e-12
BW27783 man	**8.85e-01**	1.28e-08	3.59e-09
BW27783 gly	1.45e-04	1.45e-12	1.76e-08
BW27783 glyTre	4.02e-05	8.14e-08	**6.93e-01**
BW27783 glu	**1.82e-01**	**4.23e-01**	2.95e-03
BW27783 glyCas	4.49e-05	8.21e-03	3.18e-04
BW27783 gluCas	1.32e-11	4.35e-06	6.58e-19

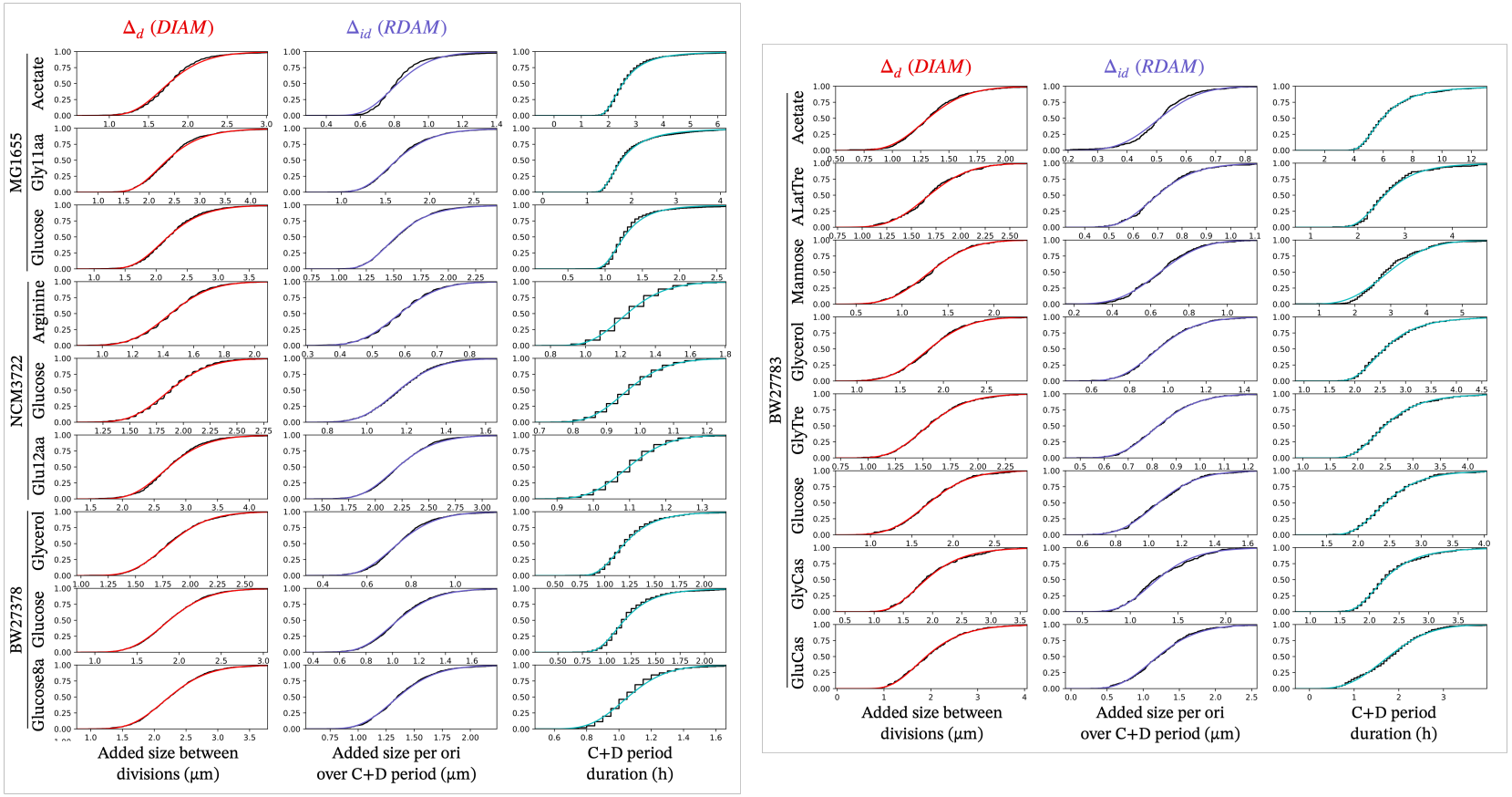
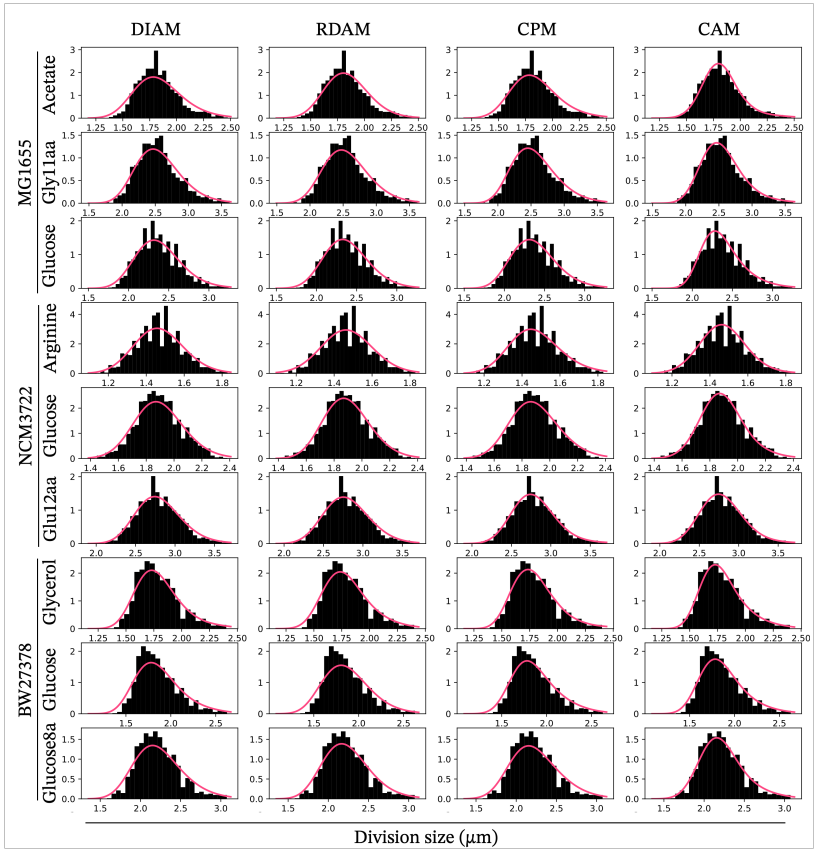


Figure S1: **Fitting Generalized Gamma distributions to inter-division added size, C+D period, and added size per origin over the C+D period** In this work we assume that the i.i.d. random variables involved in the models follow Generalized Gamma distributions. The graphs show the empirical cumulative distribution functions (in black) and their corresponding fits by likelihood maximization to generalized Gamma distributions for three cell variables: size added between divisions, size added per origin during the C+D period, and the C+D period.



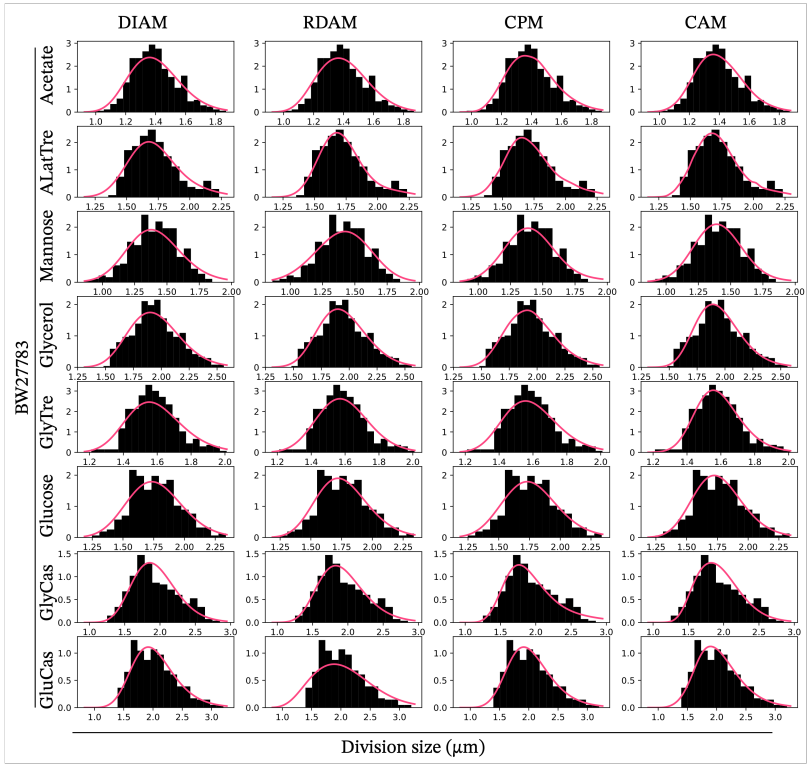


Figure S2: **Theoretical and empirical distributions of the size at division** In black are shown the empirical distributions of the division size at division. The pink curve shows the marginal distribution of the next division size according to the models. This curve have been computed as follow: $f(v) = \int p_{\theta^*}(v|v', \tilde{v}, \lambda) d\mu(v', \tilde{v}, \lambda)$ with p_{θ^*} the conditional distribution of the next volume at birth with the set of parameters θ^* maximizing the likelihood, and μ the experimental empirical measure of $(V_b^{(\bullet)}, \tilde{V}_i^{(\bullet)}, \lambda^{(\bullet)})$.

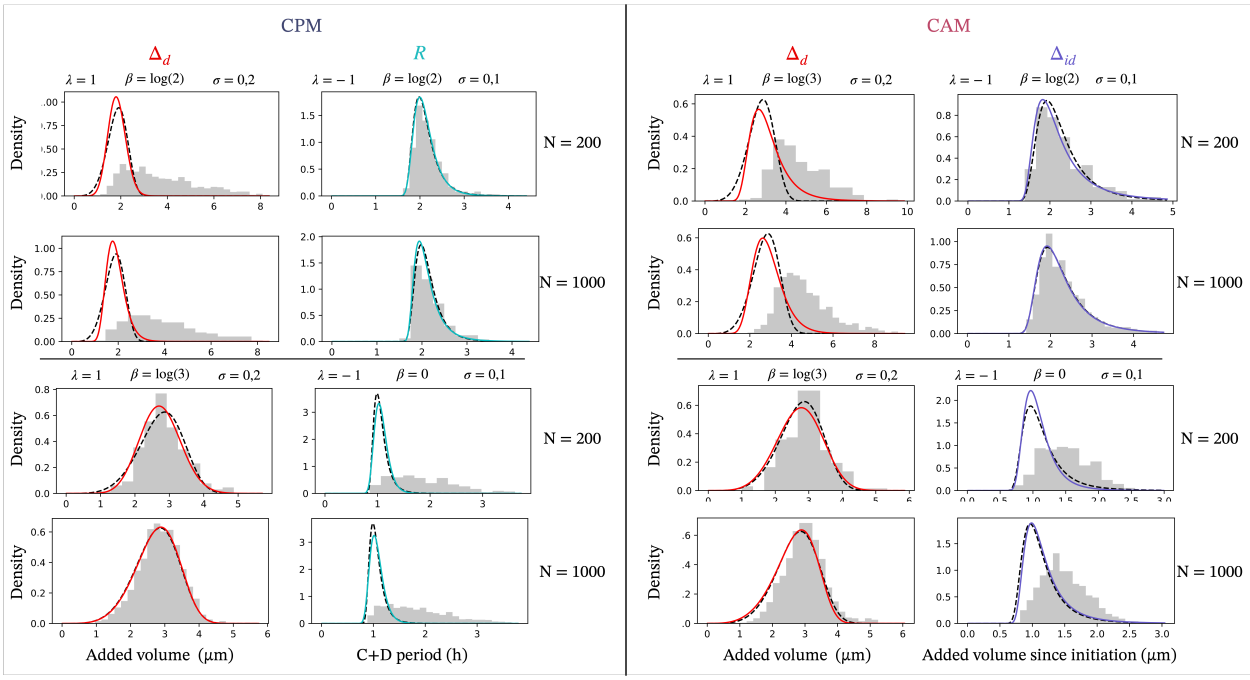


Figure S3: **Testing distribution estimation on synthetic data with known parameters** To evaluate the accuracy of our CPM and CAM distribution inference methods, we generated synthetic data using two sets of known parameters and two sample sizes. The graphs display the estimated distribution (solid colored line), the true target distribution (dashed black line), and the synthetic data (gray histogram). The parameter sets were chosen to reflect scenarios where replication has minimal or strong influence on division control: when replication- based control of division is negligible, the distribution of Δ_d coincides with that of the inter-division added size; when replication is preponderant, the distributions of R or Δ_{id} coincide with those of the C+D period and the size added per origin over the C+D period. Based on the experimental data sizes (ranging from 215 to 1806), we selected sample sizes $N = 200$ and $N = 1000$ for testing. The inferred distributions closely match the targets in these cases. However, the method may show severe limitations when replication rarely or exclusively limits division, and when the sample size is small.

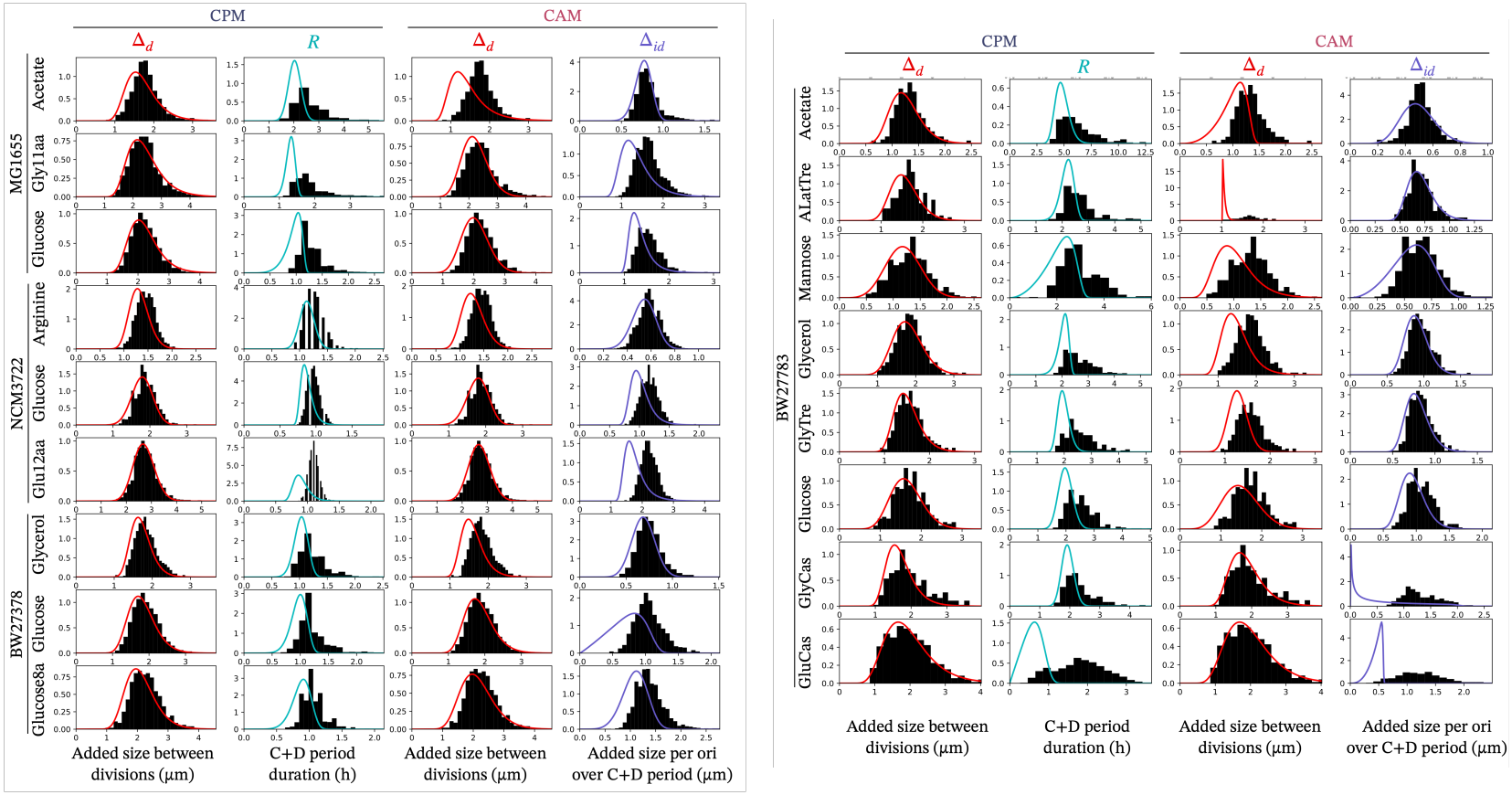


Figure S4: **Inferred distributions of the latent variables of the CPM and the CAM** The Inferred distributions of the latent variables of the CPM and the CAM are shown by the colored curves. The black histograms shows the empirical distributions of the added size between divisions, the C+D period durations, and the added size per origin over the C+D period.

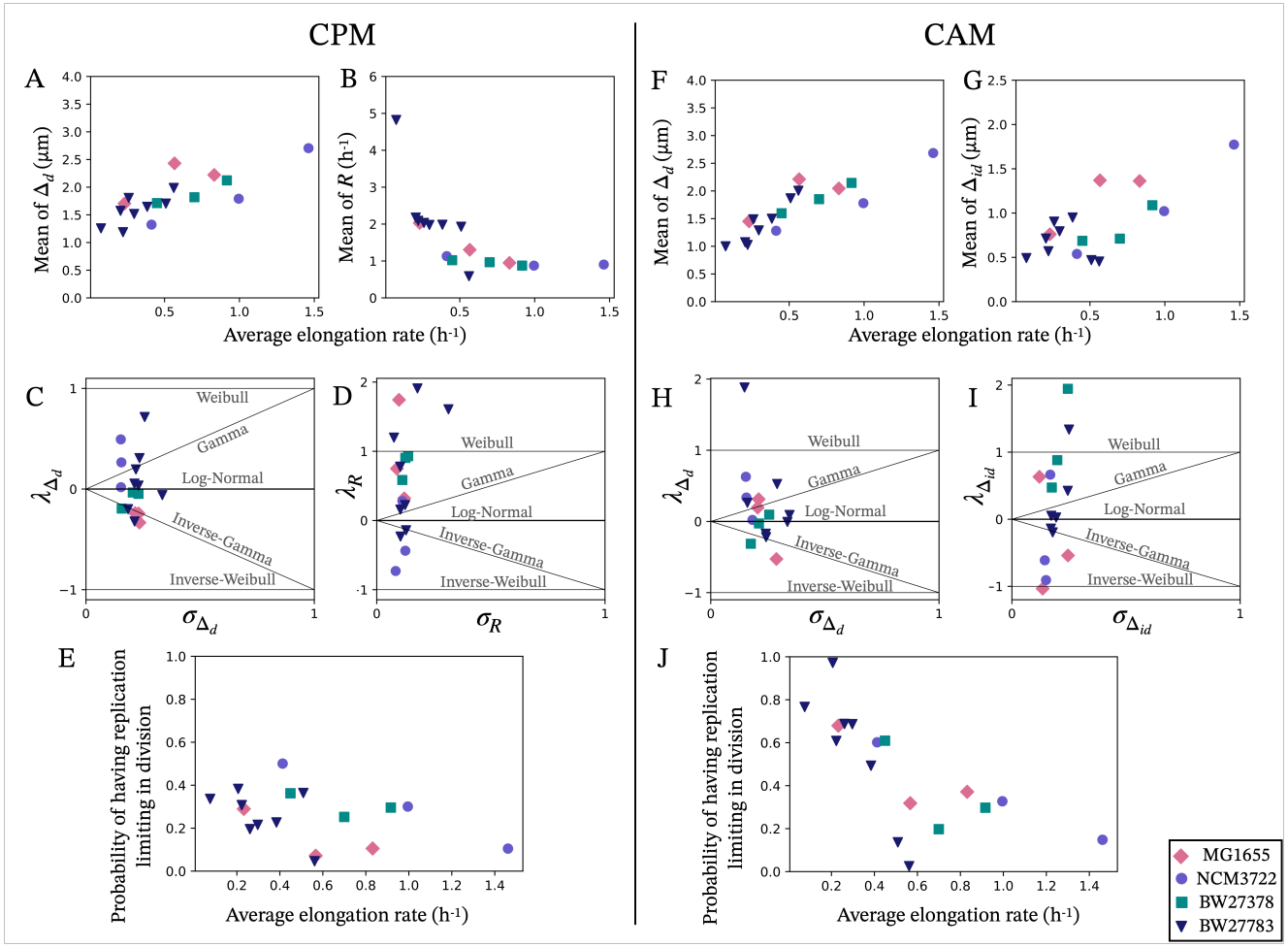


Figure S5: **Inferred parameters analysis of the CPM and CAM** In (A,B,F,G) Mean of the latent variables Δ_d, and R in the CPM and Δ_d and Δ_{id} in the CAM respectively across growth conditions. (B,C,H,I) Estimated parameters λ as a function of σ of the Generalized Gamma distribution of density $f(x) = \frac{|\lambda|}{\sigma x \Gamma(\lambda^{-2})}\left(\lambda^{-2}\left(\frac{x}{e^{\beta}}\right)^{\frac{\lambda}{\sigma}}\right)^{\lambda^{-2}} \exp\left(-\lambda^{-2}\left(\frac{x}{e^{\beta}}\right)^{\frac{\lambda}{\sigma}}\right)$. This distribution includes various probability distributions: Gamma, inverse Gamma, Log-Normal Weibull, inverse Weibull disribution [**?**]. Based on the parameters estimations, Δ_d, R and Δ_{id} in the CPM and the CAM do not seem to be consistently distributed according to one of these distributions. However, one global trend that arises is a rather constant σ. In E and J are displayed the probabilities of having the replication limiting in the division across growth conditions. For the CPM this probability is computed from: $\mathbb{P}_{\theta^*}[2^{-1}(V_b^{(\bullet)} + \Delta_d) \leq \tilde{V}_i^{(\bullet)} e^{\lambda^{(\bullet)} R}] = \int \mathbb{P}_{\theta^*}[2^{-1}(v' + \Delta_d) \leq \tilde{v}' e^{\lambda R}] d\mu(v', \tilde{v}, \lambda)$, with θ^* maximizing the likelihood, and μ the experimental empirical measure of $(V_b^{(\bullet)}, \tilde{V}_i^{(\bullet)}, \lambda^{(\bullet)})$. Similarly for the CAM, this probability is computed from $\mathbb{P}_{\theta^*}[2^{-1}(V_b^{(\bullet)} + \Delta_d) \leq \tilde{V}_i^{(\bullet)} + \Delta_{id}] = \int \mathbb{P}_{\theta^*}[2^{-1}(v' + \Delta_d) \leq \tilde{v}' + \Delta_{id}] d\mu(v', \tilde{v}, \lambda)$. As shown, the CAM predicts probabilities of having replication limiting in division close to one in slow growth conditions, which agrees with recent results (Tiruvadi et al. (2022), Kar et al. (2023))

Table S2: **BIC and AIC scores**

	BIC scores				AIC scores			
Datasets	DIAM	RDAM	CPM	CAM	DIAM	RDAM	CPM	CAM
MG1655 acet	677.42	445.54	497.82	195.65	669.39	437.52	481.77	179.60
MG1655 gly11aa	1233.28	1494.29	1219.77	1198.14	1225.32	1486.33	1203.85	1182.22
MG1655 glu	1169.80	1364.29	1147.48	1076.43	1161.55	1356.04	1130.98	1059.93
NCM3722 arg	-253.43	-358.51	-659.33	-454.58	-261.59	-366.66	-675.64	-470.90
NCM3722 glu	269.03	378.46	188.57	221.41	261.13	370.56	172.77	205.61
NCM3722 glu12aa	989.38	1241.54	978.02	976.80	981.45	1233.61	962.15	960.94
BW27378 gly	140.21	101.21	-24.41	18.88	133.18	94.17	-38.47	4.81
BW27378 glu	412.90	630.59	343.75	409.96	405.58	623.26	329.10	395.31
BW27378 glu8a	766.69	886.78	701.16	741.83	759.20	879.29	686.19	726.85
BW27783 acet	54.27	-37.54	-10.12	-42.00	48.28	-43.53	-22.10	-53.99
BW27783 alaTrE	68.10	14.18	30.77	17.57	63.04	9.12	20.66	7.46
BW27783 man	114.00	108.90	79.12	93.30	108.45	103.35	68.02	82.20
BW27783 gly	179.26	131.06	155.34	117.16	173.20	125.00	143.22	105.04
BW27783 glyTre	59.71	31.28	37.40	5.30	53.95	25.52	25.88	-6.22
BW27783 glu	125.57	123.97	116.18	120.77	120.23	118.64	105.51	110.10
BW27783 glyCas	193.60	252.14	131.86	199.31	188.32	246.86	121.30	188.74
BW27783 gluCas	359.90	471.72	365.20	363.81	354.03	465.85	353.47	352.07

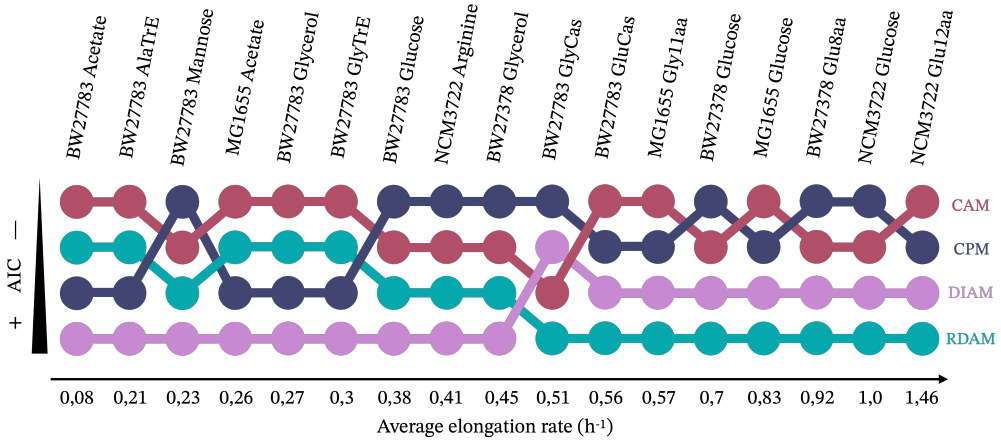


Figure S6: **Comparison of the performances of the models based on the AIC** The AIC scores have been computed for each model (DIAM in pink, RDAM in cyan, CPM in dark blue and CAM in dark red) for multiple datasets of four strains. Here are shown the ranks of the models based on their AIC score.

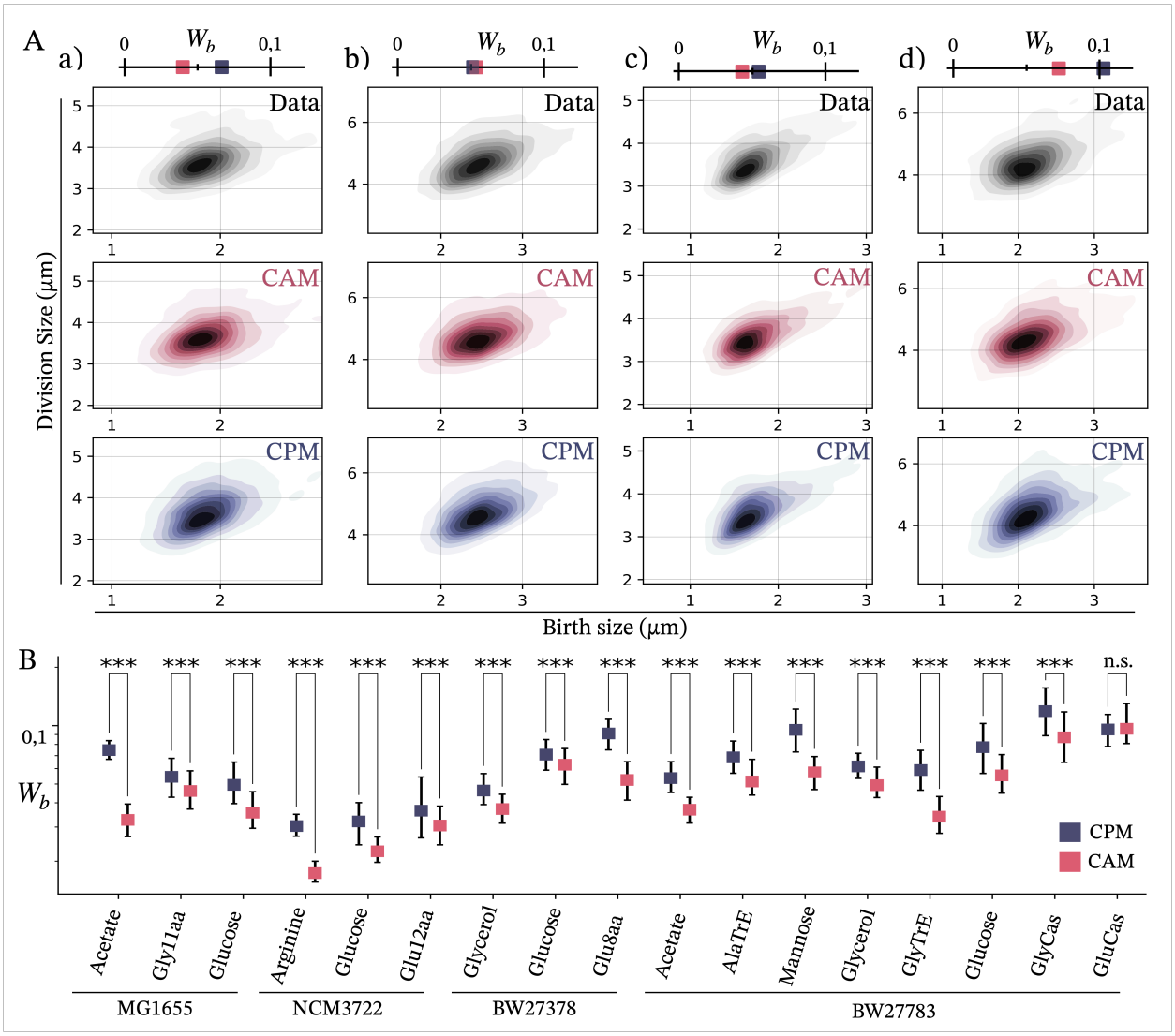


Figure S7: **Wasserstein distances between experimental and simulated joint distributions of division and birth sizes** Wasserstein distances between experimental and simulated joint distributions of division and birth sizes for 100 simulations according to the CPM (blue) and CAM (red). Significant differences according to Wilcoxon-Mann Whitney tests are indicated: ***p-value < 0.001,n.s.= not significant (p-value > 0.05).

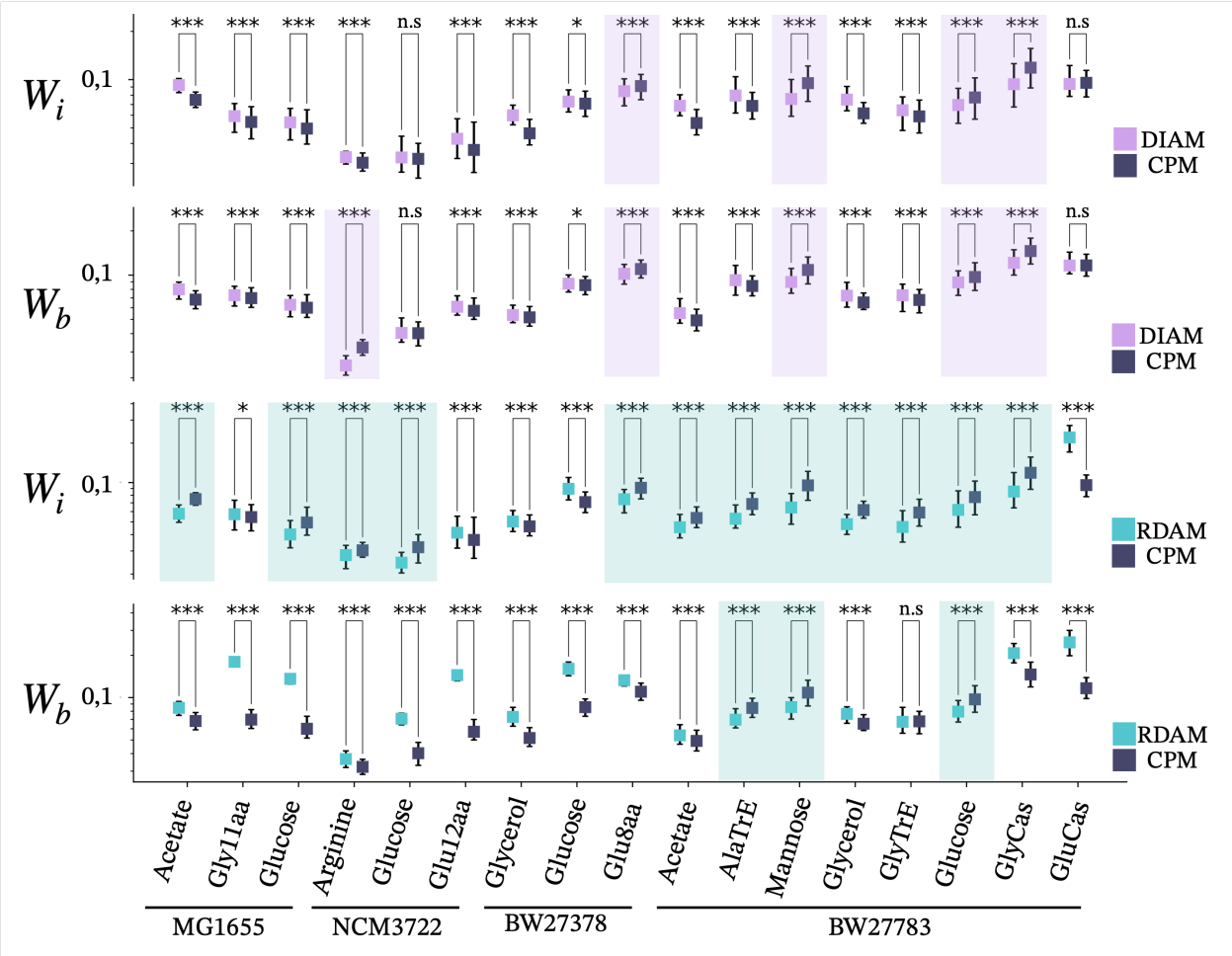


Figure S8: **Comparison of Wasserstein distances between DIAM/RDAM and CPM** The Wasserstein distances between experimental and simulated joint distributions of initiation-division sizes and birth-division sizes have been computed for the RDAM and the DIAM. In comparison with the Wasserstein distances of the CPM, the DIAM and the RDAM shows significantly smaller Wasserstein distances in some datasets for both initiation-division sizes and birth-division sizes joint distributions. The shaded areas highlight the datasets where the CPM is outperformed by either DIAM or RDAM. W_i denotes for the Wasserstein distance between experimental and simulated joint distribution of initiation-division sizes and W_b for the Wasserstein distance between experimental and simulated joint distribution of birth-division sizes. Significant differences according to Wilcoxon-Mann Whitney tests are indicated: ***p-value < 0.001, **p-value < 0.01, *p-value < 0.05, n.s.= not significant

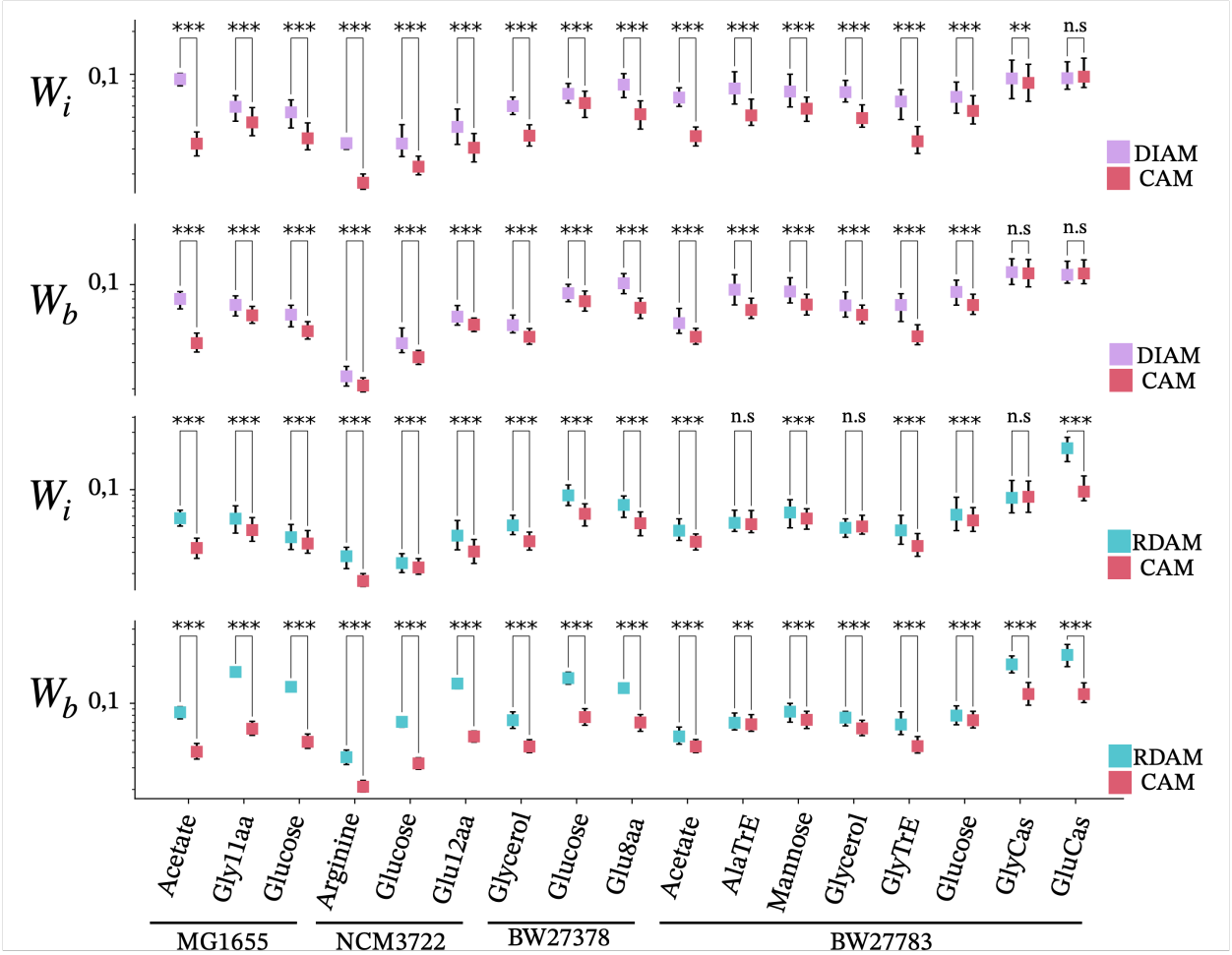


Figure S9: **Comparison of Wasserstein distances between DIAM/RDAM and CAM** Similarly to figure S8, the Wasserstein distances between experimental and simulated joint distributions of initiation-division sizes and birth-division sizes have been computed for the RDAM and the DIAM. In comparison with the Wasserstein distances of the CAM, the DIAM and the RDAM do not show significantly smaller Wasserstein distances, and conversely, the CAM yields significantly smaller Wasserstein distances than the DIAM and RDAM in a strong majority of datasets. W_i denotes for the Wasserstein distance between experimental and simulated joint distribution of initiation-division sizes and W_b for the Wasserstein distance between experimental and simulated joint distribution of birth-division sizes. Significant differences according to Wilcoxon-Mann Whitney tests are indicated: ***p-value < 0.001, **p-value < 0.01, *p-value < 0.05, n.s.= not significant