

Report

A Probabilistic Model for Cell Cycle Distributions in Synchrony Experiments

David A. Orlando¹

Charles Y. Lin¹

Allister Bernard²

Edwin S. Iversen³

Alexander J. Hartemink^{2,*}

Steven B. Haase^{1,*}

¹Department of Biology; ²Department of Computer Science; and ³Institute of Statistics and Decision Sciences; Duke University; Durham, North Carolina USA

*Correspondence to: Alexander J. Hartemink; Department of Computer Science; Box 90129; Durham, North Carolina 27708 USA; Email: amink@cs.duke.edu/
Steven B. Haase; DCMB Group; Department of Biology; Box 90338; Durham, North Carolina 27708 USA; Tel.: 919.613.8205; Fax: 919.613.8177; Email: shaase@duke.edu

Original manuscript submitted: 01/03/07

Manuscript accepted: 01/12/07

Previously published online as a *Cell Cycle* E-publication:

<http://www.landesbioscience.com/journals/cc/abstract.php?id=3859>

KEY WORDS

synchrony, cell cycle, population distribution, probabilistic model

ACKNOWLEDGEMENTS

This work was supported by ACS-IRG 83-006 from the American Cancer Society to S.B.H., and in part by a National Science Foundation CAREER award and an Alfred P. Sloan Fellowship to A.J.H.

ABSTRACT

Synchronized populations of cells are often used to study dynamic processes during the cell division cycle. However, the analysis of time series measurements made on synchronized populations is confounded by the fact that populations lose synchrony over time. Time series measurements are thus averages over a population distribution that is broadening over time. Moreover, direct comparison of measurements taken from multiple synchrony experiments is difficult, as the kinetics of progression during the time series are rarely comparable. Here, we present a flexible mathematical model that describes the dynamics of population distributions resulting from synchrony loss over time. The model was developed using *S. cerevisiae*, but we show that it can be easily adapted to predict distributions in other organisms. We demonstrate that the model reliably fits data collected from populations synchronized by multiple techniques, and can accurately predict cell cycle distributions as measured by other experimental assays. To indicate its broad applicability, we show that the model can be used to compare global periodic transcription data sets from different organisms: *S. cerevisiae* and *S. pombe*.

INTRODUCTION

To fully understand the mechanisms that regulate cell division, it would be ideal if quantitative measurements of dynamic processes could be made on a single cell as it progresses through the cell division cycle. Accurate quantitative measurements (such as transcript or protein abundance) often require the collection of material from many cells. Therefore, large initial populations of cells are often synchronized at some point in the cell cycle and subsequently released from synchrony. The characteristics of the synchronous population of cells can then be measured over time as it progresses through the cell cycle. Unfortunately, synchronization is an imprecise art: initial populations are never completely synchronous, they tend to lose synchrony over time, and their temporal progression rarely matches precisely from experiment to experiment.^{1,2} As a result, quantitative time-series measurements of a synchronized population of cells do not accurately represent the characteristics of a single cell as it progresses through the cell cycle. If the cell cycle distribution of a population of cells at various times during an experiment could be estimated, then assessments that more accurately reflect true single-cell values could be determined from population-level measurements.³⁻⁵ Here, we develop a robust mathematical model to estimate the evolving cell cycle distribution of a population of cells in a synchrony experiment.

Due to its experimental tractability, the budding yeast *S. cerevisiae* has been widely used as a model system for examining cell cycle dynamics in synchronous cell populations. A variety of methods have been established for synchronizing a yeast cell population at various points in the cell cycle, and the cell cycle position of individual yeast cells can be readily determined by monitoring landmark events.^{1,6-8} Because these events can be observed for single cells within a population, they are not subject to the convolution effects that arise in population-level measurements. For example, bud emergence, a distinct morphological landmark that correlates with the transition from G₁ into S, can be monitored by light microscopy. Cells remain budded until the completion of mitosis and cytokinesis, when cells reenter G₁. Thus, the passage of cells from G₁ (unbudded) into S (budded), or from M (budded) into G₁ (unbudded) can be measured by plotting the percentage of budded cells in a population (budding index) as a function of time. In addition to bud presence, cell cycle positions of individual cells in a population can also be determined by measuring DNA content by flow cytometry.⁹⁻¹¹

Although some information about population distributions can be learned by examining budding index and DNA content on their own, the resolving power of these methods is limited. Budding index can only tell us the fraction of cells in G_1 . DNA content can discriminate between cells in G_1 , S and G_2/M , but does not help distinguish between cells in sub-compartments of G_1 , G_2 or M (e.g., early vs. late G_1 , or G_2 vs. M). Nevertheless, when combined with a mathematical model of cell cycle progression, budding index and DNA content can be used to estimate distributions of significantly higher resolution.

Bar-Joseph et al.⁴ and Qiu et al.⁵ have developed models for learning cell cycle distributions for deconvolving population-level measurements of cell cycle regulated transcription. The model of Bar-Joseph and colleagues relies on budding index measurements to learn distributions. The model of Qiu and colleagues is based solely on the transcription dynamics of established cell cycle regulated genes. Both models assume that synchrony is lost because cells in the population progress through the cycle with variable rates. However, neither model captures asynchrony effects that may result from cell division during the time course. Niemisto et al.¹² and Lu et al.¹³ have developed approaches for deconvolution that are not based on an underlying continuous population dynamic model, but rely instead on discrete predictions of population distributions at each time point.

Several other mathematical models have been used to estimate distributions in synchronized populations of yeast cells. Most can be divided into two classes: cell ensemble models and population balance equation models (reviewed in refs. 13–15). The power of cell-ensemble models is their ability to describe processes within a single cell in great detail. However, determination of cell cycle state is dependent on a large array of measurements that are difficult to collect at every time point in each experiment. Population balance equation models describe cells with respect to their position in a population, and are therefore well suited for population-level measurements. Unfortunately, these models are partial-integro-differential equations which are very difficult to solve except under special conditions.¹⁶ Further, their solution is conditioned on a set of process parameters, dependent on environmental conditions, for which inference and evaluation are difficult.¹⁵

Multi-type branching process constructions have also been used to model cell cycle population dynamics focusing on the effects of cell division.¹⁷ These models were developed to study the asymptotic behavior of cycling populations and are not well suited for experiments aimed at investigating short-term population behavior (on the order of 1–2 cell cycles), or for inferring their underlying dynamics.

Here, we present a robust mathematical model for determining cell cycle distributions of synchronized cell populations over time. Previous studies have suggested that asymmetric cell divisions in budding yeast are a major source of synchrony loss.^{18–21} Our model explicitly incorporates reproduction within the population via a branching process construction, and thereby accounts for synchrony loss resulting from asymmetric cell division. The model is fully parametric, can be expressed in closed-form, and is conceptually easy to extend or incorporate into hierarchical models that combine data across experiments or involve multiple data types. The closed form, parametric formulation allows for full Bayesian inference; point and interval estimates of model parameters, summaries of their joint posterior distribution, and confidence sets of budding index curves are easily calculated. The model addresses some of the shortcomings of synchrony experiments by allowing the comparison of data across multiple experiments, and by learning functions that can estimate

single-cell values from population-level measurements convolved by synchrony loss.

MATERIALS AND METHODS

Yeast strains, cell growth and synchrony. The yeast strain used in this study is a derivative of BF264-15DU (*MATa ade1 his2 leu2-3 112 trp1-1 ura3Δns bar1*). Yeast cultures were grown in YEP medium (1% yeast extract, 2% Bacto-Peptone, 0.001% adenine, 0.001% uracil) supplemented with 2% sugar (dextrose or galactose). Cells were grown at 30°C except as noted. α -factor synchrony experiments were performed by adding the mating pheromone α -factor (20ng/ml) to the growth medium for 2 hr. Synchronized cell populations were then harvested by centrifugation, and released from the arrest by resuspending in cells in medium without α -factor. Elutriation synchrony experiments were performed as described previously.³¹ Cells were first grown in YEP-galactose medium, and then released into YEP-dextrose + 1M sorbitol. Sorbitol was used for osmotic support, and has a marginal effect on cell growth.

Flow cytometry. Flow cytometric analysis of DNA content was performed as described previously.⁹

Mathematical description of CLOCCS model. The CLOCCS model specifies the distribution, $\Pr(P_t|\Theta, t)$, of cell positions in the population at a given time, t , and conditional on the model's parameters, Θ , where

$$\Theta = \begin{cases} \mu_0 & \text{Length of recovery phase, Gr} \\ \sigma_0^2 & \text{Variance in recovery length} \\ \sigma_v^2 & \text{Variance in lifeline velocities} \\ \lambda & \text{Length of a normal cell cycle} \\ \delta & \text{Length of daughter specific phase, Gd} \end{cases}$$

The expression for $\Pr(P_t|\Theta, t)$ is obtained by introducing the concept of cohorts, in which cells in the population are identified by their generation g and reproductive instance r . The distribution of position given time is obtained by marginalizing the joint distribution of position and cohort given time over cohort. In particular, we write

$$\Pr(P_t|\Theta, t) = \sum_G \sum_R \Pr(P_t|\Theta, R, G, t) \Pr(G|\Theta, R, t) \Pr(R|\Theta, t).$$

The first term on the right hand side of the equation above represents the distribution of position of cells in cohort $\{g, r\}$ on the lifeline as a function of time. Position within a cohort follows a normal law as described in Results. Because daughter cells begin δ units before the start of the normal cell cycle, all cohorts other than $\{g = 0, r = 0\}$ are truncated at that point. Specifically,

$$\Pr(P_t|\Theta, g, r, t) = \begin{cases} \phi\left(\frac{P_t - (-\mu_0 + t)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right) & g = 0, r = 0 \\ I_{\{P_t \geq -\delta\}} \frac{\phi\left(\frac{P_t - (-\mu_0 + t - r \cdot \lambda - g \cdot \delta)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2} \left(1 - \Phi\left(\frac{-\delta - (-\mu_0 + t - r \cdot \lambda - g \cdot \delta)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right)\right)} & \text{else,} \end{cases}$$

where $\phi(\cdot)$ denotes the standard normal density function, $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and $I_{\{A\}}$ denotes the indicator function for the event A which evaluates to 1 if A is true and 0 if A is false.

Taken together, $\Pr(G|\Theta, R, t)$ and $\Pr(R|\Theta, t)$ specify the probability mass function of the cohorts as a function of time, t . These expressions explicitly model the increase in population size that results from cohorts branching into mothers and daughters as they progress through mitosis.

Let $M(g,r,t)$ denote the size of cohort $\{g,r\}$ at time t relative to the size of the $\{0,0\}$ cohort.

$$M(g,r,t) = \begin{cases} 1 & g=0, r=0 \\ 0 & g=0, r>0 \\ 0 & g \geq 1, r < g \\ \left(1 - \Phi\left(\frac{\mu_0 - t + r \cdot \lambda + (g-1) \cdot \delta}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right)\right) \cdot \binom{r-1}{g-1} & g \geq 1, r \geq g \end{cases}$$

Let $Q(r,t)$ denote the size of all cohorts with reproductive instance r at time t relative to the size of the $\{0,0\}$ cohort.

$$Q(r,t) = \sum_{y=0}^{r-1} \left(\left(1 - \Phi\left(\frac{\mu_0 - t + r \cdot \lambda + y \cdot \delta}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right)\right) \cdot \binom{r-1}{y} \right).$$

Finally, let $Q(t)$ denote the total size of the population at time t relative to the size of the $\{0,0\}$ cohort.

$$Q(t) = \sum_{r=0}^R Q(r,t).$$

Hence,

$$\Pr(G|\Theta, R, t) = \frac{M(G, R, t)}{Q(R, t)}$$

and

$$\Pr(R|\Theta, t) = \frac{Q(R, t)}{Q(T)}.$$

Inference given budding data. Budding is determined by a cell's position on the lifeline: the fraction of budded cells in the population at any given time is equal to the fraction of cells in S, G₂, and M phases. This quantity is determined by the CLOCCS model, which specifies the distribution of cells across the lifeline at any given time point. Let β denote the fraction of time through the cell cycle, λ , at which budding begins. Buds disappear at the end of the cell cycle when cells divide.

Let $B(t, \beta, \Theta, g, r)$ denote the expected fraction of cohort $\{g,r\}$ that is budded at time t . For cohort $\{g=0, r=0\}$,

$$B(t, \beta, \Theta, g, r) = \sum_{c=0}^C \left[\Phi\left(\frac{\delta \cdot (c+1) - (-\mu_0 + t)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right) - \Phi\left(\frac{\delta \cdot (c+\beta) - (-\mu_0 + t)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right) \right]$$

otherwise,

$$B(t, \beta, \Theta, g, r) = \sum_{c=0}^C \left[\frac{\Phi\left(\frac{\delta \cdot (c+1) - (-\mu_0 + t - r \cdot \lambda - g \cdot \delta)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right) - \Phi\left(\frac{\delta \cdot (c+\beta) - (-\mu_0 + t - r \cdot \lambda - g \cdot \delta)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right)}{\left(1 - \Phi\left(\frac{-\delta - (-\mu_0 + t - r \cdot \lambda - g \cdot \delta)}{\sqrt{\sigma_0^2 + t^2 \cdot \sigma_v^2}}\right)\right)} \right] \cdot M(g, r, t)$$

The sums over cell cycle, c , account for the fact that a given cohort could be distributed across multiple cell cycles and therefore have budded members in discontinuous regions. We account for this by summing the fraction of the cohort in the budded region for a large number of cell cycles C .

The expected fraction of cells budded at time t is

$$B(t, \beta, \Theta) = \frac{\sum_{g=0}^G \sum_{r=g}^R B(t, \beta, \Theta, g, r)}{Q(t)}.$$

We assume a Bernoulli sampling model for the budding measurements made at each time period, and further assume that measurements are independent from time period to time period with success probability $B(t, \beta, \Theta)$. Let $X_{it}=1$ if the i th sampled cell at time t is budded and 0 if it is not and let n_t denote the total number of cells sampled at time t . Hence the sampling model for X , the vector of X_{it} 's, is

$$\Pr(X|\Theta, \beta) = \prod_{t=1}^T B(t, \beta, \Theta)^{\sum_i X_{it}} \times (1 - B(t, \beta, \Theta))^{n_t - \sum_i X_{it}}$$

Model implementation and fitting. The model fits reported in this paper were performed using an implementation written in C and run within the R statistical programming environment. We employed a Markov Chain Monte Carlo (MCMC) algorithm for posterior inference. Specifically, we updated individual parameters via a systematic scan random walk Metropolis sampling scheme.³² In keeping with their biological interpretations, we constrained all parameters, except for μ_0 to be positive. For *S. cerevisiae*, we placed a beta(4,16) prior on β to reflect the prior belief that β is between 5% and 35% with probability 0.95, and in no case larger than 1. We placed an inverse gamma(12,1) prior on σ_v to reflect the belief that it is centered around 0.09, which is consistent with the value reported by Bar-Joseph et al.⁴ An exponential prior with scale parameter 40 was placed on δ to avoid explaining too much asynchrony with this parameter, especially in cases where the time courses were not long and the effects due to reproduction had not had time to effect the system. We also constrained δ to be no larger than the estimate for the cell cycle time (λ).

Point estimates of the model parameters were calculated as sample averages, and interval estimates were calculated as ranging from the 2.5th to 97.5th percentile of the parameter's samples. The chain was run for a burn-in period followed by 250,000 subsequent iterations. The chain showed quick convergence and good mixing by visual inspection. In addition, we employed a battery of convergence diagnostics implemented in the R package CODA to evaluate the convergence of the model's parameters. All chains passed the Raftery and Lewis diagnostic, the Heidelberger and Welch stationary tests, and the Geweke diagnostic with each diagnostic's parameters set to their CODA default values.³³⁻³⁵ While these results are consistent with convergence of the algorithm and indicate that the sample is informative for the kind of posterior summaries that interest us, they do not guarantee that the sample is representative of the posterior distribution. However, we assumed it was and thereafter based inference on the sample.

Predicting flow cytometric data. Due to measurement noise and sample-to-sample variability in the fluorescence histograms, some preprocessing of the observed data was necessary. Histograms at each time point were aligned so that 1C and 2C peaks were registered to the same channels across all time points. Measurement noise seemed to be multiplicative, so we log-transformed fluorescence values in an attempt to normalize noise variance across all channels. After transformation, the noise was assumed to be Gaussian with a variance dependent on channel number. To estimate the variance, we examined clear 1C and 2C peaks and found that standard deviations of 0.093 and 0.073 captured the variance at those channels well when judged visually. The standard deviation for the intervening channels was assumed to be a linear interpolation between the two extremes. The predicted DNA content histogram represents a sum of normal distributions with means at the fluorescence values and variances

described above, weighted by the expected population percentages at each of those fluorescence values as calculated by the distribution over the lifeline and the mapping from lifeline position to fluorescence value.

The fraction of the cell cycle during which cells are in S was determined by analyzing log-transformed DNA histograms from asynchronous populations. We fit Gaussian distributions to the 1C and 2C peaks, representing cells either in G_1 or G_2/M . After subtracting these two populations, we were left with $\sim 25\%$ of the mass, which we assumed to be equal to the length of S in a normal cell cycle (data not shown). While budding and the start of DNA replication occur at similar times, they are not exactly coincident.¹⁸ To account for the difference, the start of DNA replication was conservatively set to 4 minutes earlier than the detection of budding ($\beta - 0.05$).

Model based alignment of *S. cerevisiae* and *S. pombe*. For the *S. cerevisiae* experiment we used budding index as observed data.²⁸ Unfortunately, budding index data was not gathered by Pramila and colleagues for the synchrony used to generate the α -30 microarray dataset. As an approximation, we used the average of wild-type budding curves shown in Figures 2 and 5 in Pramila et al. (Fig. 6A).²⁸ For *S. pombe*, we used septation index from the Elutriation 2 experiment to learn model parameters (Fig. 6B).²⁷ The resulting model parameter estimates are shown in Supplemental Table 1. The interpretation of the parameters for *S. pombe* is slightly different than for *S. cerevisiae*. The negative value for μ_0 captures the fact that elutriated *S. pombe* cells are in G_2 and are therefore already in their first cell cycle and thus the recovery time (G_r) to the first G_1 is, in a sense, negative. The meaning of σ_0 , σ_v and λ are the same across organisms. The value for δ is constrained to be zero in *S. pombe* because division is symmetric and therefore the difference between mothers and daughters should be equal to zero. The interpretation of β is also changed. Where β in *S. cerevisiae* is the percentage of the cell cycle where cells are unbudded, in *S. pombe* it captures the percentage of the cell cycle where cells are unseptated. Using the learned parameter estimates, we were able to create a mapping between the two experiments (Fig. 6C). To map *S. pombe*, we used the knowledge that *S. pombe* cells begin septating late in mitosis, corresponding to a lifeline position $\beta\lambda$ in the *S. pombe* lifeline (as opposed to *S. cerevisiae* lifeline where position $\beta\lambda$ defines the G_1/S border) and that *S. pombe* cells divide approximately halfway through S-phase, corresponding to a *S. pombe* lifeline position λ (in *S. cerevisiae*, this position would be defined as $(\beta+0.25/2)\lambda$, where S-phase is 25% of the cell cycle). Using these sets of corresponding cell cycle phase/lifeline points we were able to successfully map the two synchronies into the same lifeline space and align the observed data (Fig. 6D). Due to the large relative differences in cell cycle phase length in *S. pombe*, the mapping function has a more pronounced differential stretching from time point to lifeline position.

To make the two transcription array dataset values comparable, we took the log₂ of the values given in the *S. pombe* dataset.²⁷ This ensured that all values in either dataset were describing the log₂ fold over or under expression vs. the mean for each gene across the time course at each time point. Using the mapping described above, we were able to align corresponding time-points in the two datasets and make relevant cross-species temporal transcriptional profile comparisons (Fig. 6E).

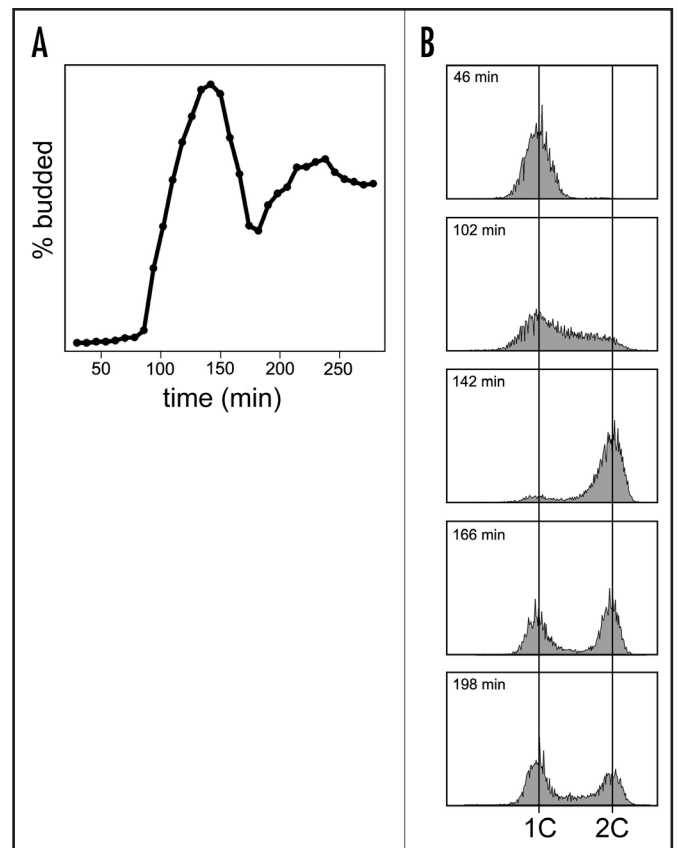


Figure 1. A representative elutriation synchrony time course. Wild-type cells were synchronized by centrifugal elutriation. Aliquots from a synchronized population were drawn at 8 min intervals. At each time point, at least 200 cells were assessed for the presence of buds (A). Corresponding aliquots were also fixed for flow cytometric analysis of DNA content (B).

RESULTS

Experimental monitoring of cell cycle progression in a synchronous population of cells. Representative data from a typical experiment in which bud presence and DNA content were monitored in a synchronized population of wild-type yeast cells are shown in Figure 1. In this experiment, a population of cells synchronized in early G_1 was collected by centrifugal elutriation, a size-selection method for extracting small, unbudded cells from an asynchronous population.⁷ Early G_1 cells were then incubated in growth medium. Cell samples were drawn and scored for the presence of buds and analyzed for DNA content at 8 min intervals.

As is common in G_1 -synchrony experiments, cells begin to bud only after a significant lag (~ 80 min in this experiment, Fig. 1A). This extended unbudded period is likely to reflect two conditions. First, elutriation preferentially selects the smallest cells from the asynchronous population (early G_1). A strong connection between cell growth and cell cycle progression has been documented, and it has been proposed that the initiation of DNA replication is dependent on cells reaching a critical size.^{8,10,18-21} Thus, small elutriated cells are likely to undergo a significant delay before a sufficient size is reached. Second, cells collected by elutriation need time to recover from the effects of the elutriation procedure itself. During the procedure, cells are subjected to changes in growth medium and osmolarity, as well as temperature shifts. Each of these treatments likely perturb cell cycle progression, and probably contribute to the delay observed early in

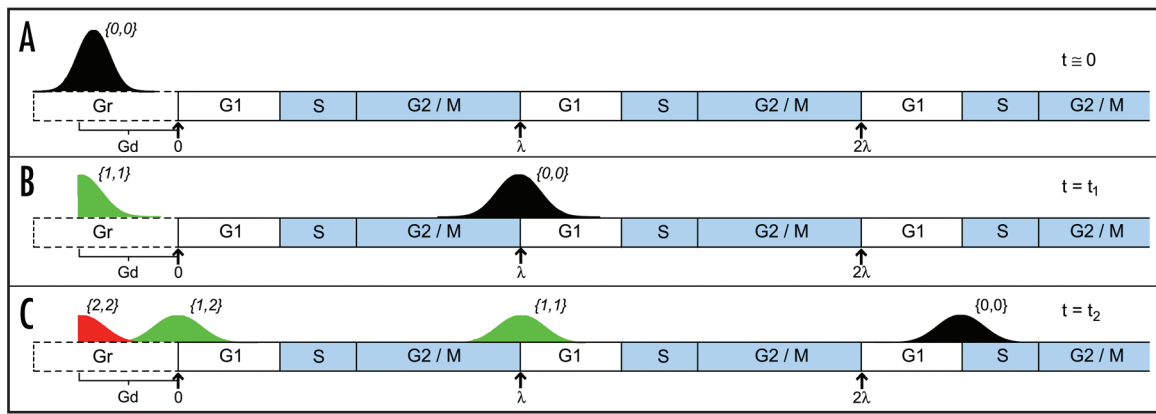


Figure 2. CLOCCS lifeline representation. (A) The initial synchronized population of cells, cohort $\{0,0\}$, begins at the far left of the lifeline, in the Gr region. As time progresses, this population passes through the different phases of the cell cycle (labeled boxes). The lifeline also indicates the unbudded region of the cell cycle (unshaded), and the budded region (shaded blue). (B) As the population divides at the M/G₁ transition, cohort $\{0,0\}$ continues down the lifeline, but as individual cells divide, a new cohort $\{1,1\}$ branches off and appears as a truncated normal at the start of Gd. (C) Later in the time course, both the initial population, cohort $\{0,0\}$, and the new daughter cohort $\{1,1\}$ will divide, producing two new daughter cell cohorts. Cohort $\{1,2\}$ arises from the second round of division of cohort $\{0,0\}$, and cohort $\{2,2\}$ arises from the first round of division of cohort $\{1,1\}$.

the time course.^{22,23} Nevertheless, despite potential perturbations resulting from elutriation, we observed that cells did correctly trigger events associated with passage through G₁ (e.g., activation of the G₁ transcriptional program, data not shown).

After 140 min, the percentage of budded cells declined, indicating that cells were completing mitosis and cytokinesis. However, the budding index never returned to 0%, suggesting that some cells in the population had begun to bud again in the second cycle before others had completed cytokinesis. Cell counts at each time point indicated that nearly all of the cells completed cell division (data not shown), and thus it is unlikely that the curve reflects a subpopulation of cells that never underwent cell division. At 175 min, the percentage of budded cells began to rise again, indicating that the bulk of the cells in the population were entering the budded phase of their second cycle.

Flow cytometric measurements of DNA content (Fig. 1B) were used as an independent measure of cell cycle progression in the synchrony experiment. Flow cytometry confirmed findings of previous studies that S-phase entry and bud emergence occur at similar times (see Fig. 1A and B, 102 min).^{10,24} DNA content measurements predicted that the bulk of the population completed mitosis and cytokinesis between 142 and 166 min, exactly when a significant portion of the population became unbudded (Fig. 1A). Flow cytometric data also demonstrates that cells in the population were broadly distributed across multiple cell cycle phases by 198 min.

Our observations suggest that multiple sources of synchrony loss ultimately contribute to population asynchrony. From the slope of the budding index curve during its initial rise, it is clear that the population was already not perfectly synchronous (Fig. 1A). Perfectly synchronous cells would be expected to transition from 0 to 100% budded instantaneously, but the population took approximately 60 min to become fully budded. Asynchrony in the population at this early stage is also evident in the flow cytometric measurements of DNA content (Fig. 1B, 102 min). These observations suggest that the initial population was not fully synchronous, or that the population had already lost synchrony during transit through G₁, or both.

The slope of the budding curve as the cells entered the budded phase of the second cycle is significantly lower than on the first cycle, indicating that the population had lost synchrony between these two points.

A significant portion of the synchrony loss that is observed in the second cycle is likely due to the asymmetric nature of cell division in *S. cerevisiae*. Several studies demonstrate that the cell cycle period for daughter cells is significantly longer than for mother cells.^{18,19,21,25} Following cell division, daughter cells are smaller than mother cells, and mechanisms that are not yet well understood are thought to delay the daughter cells in early G₁ until they achieve a critical cell size.²⁶ Mother cells however, are often already above the critical size threshold, and progress more rapidly through G₁.

CLOCCS: A probabilistic model for synchrony loss during cell cycle progression. To describe synchrony loss during the cell cycle, we created a probabilistic model framework, CLOCCS (characterizing loss of cell cycle synchrony). As suggested by previous findings and as illustrated by our initial observations, mathematical models must capture multiple sources of synchrony loss in order to accurately predict the distributions of cells over time. In our framework, we visualize the movement of a synchronized population of cells using a linear graph we term a “lifeline” (Fig. 2). Although the cell cycle is often graphically depicted in a circular form, a linear representation more easily allows us to consider events that occur only once in the lifetime of each cell, such as effects related to the synchrony procedure, or to the daughter-specific delay in early G₁. In our lifeline representation, the initial synchronized population progresses through a period of recovery from synchrony (Gr). In addition, when new daughter cells are born, they are placed near the beginning of a lifeline and progress through a daughter-specific early G₁ period (Gd). In either case, cells then move through cell cycle phases representing the “normal” cycles experienced by mother cells (Fig. 2A–C).

Our mathematical formulation begins with a simple premise: that the position of any cell i on the lifeline at time t (P_{ti}) is a function of its initial position (P_{0i}) and its velocity (V_i) such that $P_{ti} = P_{0i} + V_i t$.

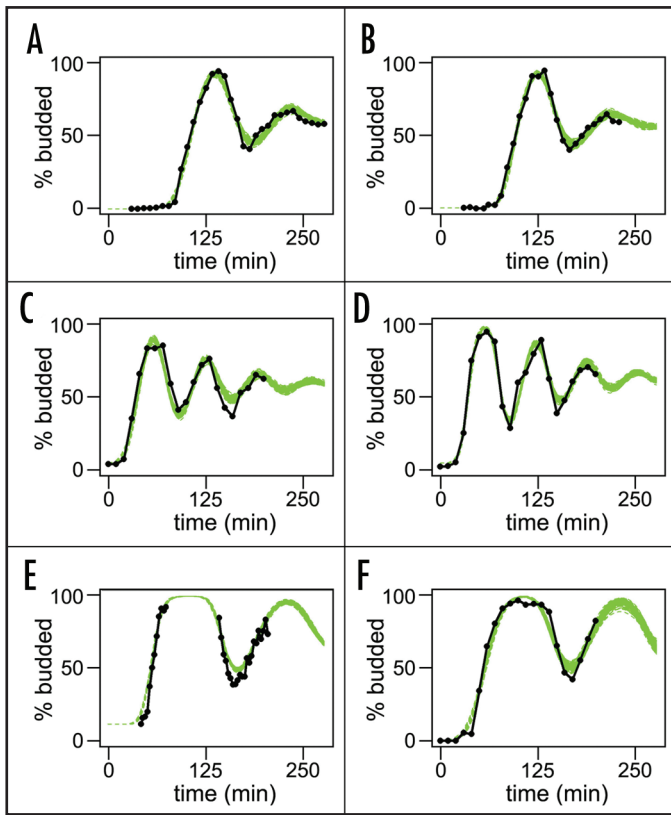


Figure 3. Model fits for multiple experiments. Observed budding index curves (black lines) for cells synchronized by elutriation (A and B) or α -factor (C and D) and then incubated at 30°C. Experiments synchronized by α -factor and then incubated at 21°C are shown (E and F). One hundred random realizations from the Markov chain used to fit each experiment were used as parameterizations for the model, and the resulting predicted budding curves are shown (green lines). The width of the band created by the Markov chains is indicative of the effect of uncertainty in the model parameters. The model parameterizations from the Markov chains for each of the experiments are shown in Table 1.

The model can explicitly capture three mechanisms of synchrony loss: (1) Variability in the initial population. Because we cannot measure landmark events in early G_1 , we have no direct evidence of variability in the initial population. However, variability in the size distribution of early G_1 cells has been reported, and could easily account for variance in the starting population.^{8,21} We assume the positions of cells in the initial population (P_{0i}) can be represented by a normal distribution with unknown mean, capturing the length of time required for recovery from arrest, and unknown variance, capturing the initial variability: $N(-\mu_0, \sigma_0^2)$. (2) Variability in the velocity of individual cells as they move through the cycle. The model assumes that each cell progresses through each phase of the cell cycle at a constant velocity chosen from a normal distribution $N(1, \sigma_v^2)$ with unknown variance. (3) Variability related to asymmetric cell division. Because cell division in *S. cerevisiae* gives rise to distinct populations of mother and daughter cells that progress through the cell cycle with different kinetics, the population is modeled as a mixture of the original synchronized population and subpopulations (cohorts) that arise from cell division (Fig. 2).

The model accounts for the new generations that arise during the course of an experiment by indexing each population cohort and using those indices to appropriately place the cohorts on the lifeline.

After each cell division, daughter cells are placed at the beginning of Gd on the lifeline, while mother cells continue into the next G_1 phase on the lifeline (Fig. 2B and C). The model assumes that the length of Gd (δ) is constant, although evidence from single cell studies suggests that the length of Gd may be variable.^{18,21,25} The cohorts are indexed by a latent integer valued variable g which, when multiplied by the length of Gd, captures generation-dependent asynchrony arising from the daughter-specific delay. However, because the lifeline representation allows each cell to undergo multiple rounds of reproduction, new members can be born into the same generational cohort at different rounds of reproduction. This aspect of the model poses a significant problem when trying to determine the marginal probability of g , $\Pr(g)$, which is required to write the model in closed form. To solve this problem we subdivide the generational cohorts based on the round of reproduction from which their individual members arose, thereby indexing them with another integer valued latent variable, r . The variable r , when multiplied by the cell cycle period λ , ensures that different cohorts of the same generation, g , are correctly placed and separated on the lifeline. Each cohort can now be identified by a unique bivariate index, $\{g, r\}$. As the initial population (Fig. 2A, $\{0, 0\}$) divides at the end of the first mitosis, a new cohort of daughter cells (Fig. 2B, $\{1, 1\}$), branches off and appears at the beginning of Gd, while the mother cell cohort continues down the lifeline (Fig. 2B, $\{0, 0\}$). As time progresses, both the new population $\{1, 1\}$ and the original mother population $\{0, 0\}$ undergo cell division (Fig. 2C) to produce two new daughter populations ($\{0, 0\}$ gives rise to $\{1, 2\}$ and $\{1, 1\}$ gives rise to $\{2, 2\}$). This arrival of new members to the population is modeled by what is known in the probability literature as a branching process. Thus, the distribution of the entire population can be represented as the sum of distributions of the unique cohort populations.

We model these cohort-specific position distributions as the sum of two independent, normally distributed effects and a cohort-specific fixed effect. In particular, we write $P_{ti} = P_{0i} + V_i t - \lambda r_i - \delta g_i$ where the position of cell i at time $t = 0$, denoted P_{0i} , is chosen from the distribution describing asynchrony in the initial population, $N(-\mu_0, \sigma_0^2)$. The constant velocity V_i of cell i as it progresses down the lifeline is chosen from $N(1, \sigma_v^2)$. Cohort indices g and r are respectively multiplied by δ and λ to ensure correct placement on the lifeline as described above. Hence, the position of a cell in cohort $\{g, r\}$ at time t is normally distributed with mean $-\mu_0 + t - \lambda r - \delta g$ and variance $\sigma_0^2 + t^2 \sigma_v^2$. Details can be found in Materials and Methods.

We introduce the parameter β to represent the fraction of time cells spend in the unbudded phase of the normal mother cell cycle. On the lifeline, bud emergence coincides with the onset of S-phase at the points $\lambda\beta$, $\lambda(\beta+1)$, $\lambda(\beta+2)$, etc., and buds disappear after cells complete cytokinesis at points λ , 2λ , 3λ , etc. (Fig. 2, shaded regions). Details can be found in Materials and Methods.

Fitting the model using budding index measurements. Observed data regarding the budding status of cells in a population consist of counts of budded cells b_t among the total number of cells sampled n_t at each sampling time t . We assume these counts to be independent binomial random variables with a success probability $\Pr(b_t | \mu_0, t, \lambda, \delta, \sigma_0, \sigma_v)$ and number of trials n_t .

We employed a random walk Markov chain Monte Carlo algorithm for parameter inference. Each chain was run for 250,000 iterations after a burn-in period, and each passed a battery of convergence diagnostics (see Materials and Methods). Parameters were estimated for multiple independent synchrony experiments using both elutriation and α -factor synchrony protocols (Fig. 3). Budding

Table 1 **Model estimates for multiple experiments**

	Elutriation			α -Factor		
	30°C			30°C	21°C	
	A	B	C	D	E	F
μ_0	95.8 (92.7,98.6)	85.1 (80.7,88.3)	20.3 (14.2,24.5)	18.7 (14.7,23.1)	38.1 (31.8,45.9)	37.7 (33.2,43.4)
λ	77.1 (72.8,81.7)	73.4 (68.5,80.2)	62.9 (58.3,69.6)	63.6 (58.7,67.8)	121.5 (111.0,128.9)	124.0 (116.9,129.4)
δ	41.4 (36.2,46.1)	42.3 (35.3,47.4)	13.6 (1.5,21.3)	7.8 (0.4,16.6)	13.1 (0.6,29.2)	6.9 (0.3,18.5)
σ_0	15.3 (14.0,16.6)	13.0 (10.3,14.7)	12.1 (11.0,13.2)	8.4 (7.4,9.4)	9.2 (8.2,10.3)	17.6 (16.1,19.0)
σ_v	0.06 (0.04,0.08)	0.07 (0.05,0.11)	0.08 (0.05,0.11)	0.09 (0.06,0.10)	0.09 (0.06,0.10)	0.07 (0.05,0.10)
β	0.14 (0.12,0.17)	0.16 (0.13,0.20)	0.28 (0.23,0.34)	0.27 (0.22,0.31)	0.15 (0.10,0.19)	0.16 (0.12,0.19)

Each column corresponds to a different experiment, synchronized by either centrifugal elutriation or α -factor arrest, and grown at either 30°C or 21°C. The observed budding curves, as well as the model predictions for each experiment, are shown in Figure 3. Each cell contains the mean value for that row's parameter, given 250,000 iterations of the Markov chain used to fit the column's experiment. Beside each mean is the 95% confidence range for that parameter in that experiment.

index curves for each of the experiments are shown, and overlaid on those curves (green) are the predictions from models parameterized by one hundred realizations of the Markov chain. These reflect the degree of posterior uncertainty in the budding curve and can be interpreted as forming a confidence band for the curve based on the uncertainty in the parameters.

Estimates of posterior means of the parameters as well as the 2.5th and 97.5th percentiles for each parameter in multiple experiments are detailed in Table 1. The best-fit values for each of the parameters are remarkably similar for experiments performed under the same conditions. Notably, little change is observed in the variance in velocity of cells as they progress through the cycle (σ_v), even across experiments performed under significantly different conditions. The small values observed for σ_v suggest that once cells enter the cycle, they progress at very similar rates, and that much of the observed loss of synchrony is not explained by this factor alone.

As expected, significant differences are observed for some parameters when experimental conditions are changed. The length of the cell cycle λ is extended as expected when α -factor synchrony experiments are run at lower temperatures (see Table 1, 30°C vs. 21°C). We also observed that the values for μ_0 are significantly larger in elutriated populations than in α -factor synchronized cells (Table 1). The value of μ_0 reflects the amount of additional time required for the initial cells to complete their first cycle, over and above the normal λ . Unlike cells synchronized by α -factor, elutriated populations were exposed to an osmotic shock resulting from their release into growth medium containing 1M sorbitol (see Materials and Methods), and pathways that transiently delay cell cycle progression in response to osmotic stresses are expected to increase μ_0 .²³

The increase in estimated values of δ in elutriated populations (Table 1) suggests that the daughter-specific delay in G_1 (Gd) is significantly longer than in populations synchronized with α -factor. This observation could be explained by distinct differences in the synchrony procedures. The elutriation process preferentially collects small daughter cells, but α -factor-treated cells continue to increase in size during the arrest. Thus, upon release, small elutriated cells delay entry into the cell cycle until they reach a minimum size requirement, while large α -factor-arrested cells are already above the minimum budding size and enter the cell cycle rapidly. Additionally, large mother cells resulting from α -factor arrest have been shown to produce large daughter cells.⁸ Consequently, large cells in the $\{0,0\}$ cohort are likely to produce large buds that become large daughters in the $\{1,1\}$ cohort, and these large daughters are likely to experience

a shorter delay in Gd, than the small daughters produced from elutriation-synchronized cells.

Predicting cell cycle distributions. To test the capabilities of the model, we used parameter values fit by the model to predict some measurements that were not used in fitting the parameters. The measurement of DNA content by flow cytometry is a common method for determining cell cycle distributions (Fig. 1B), so we

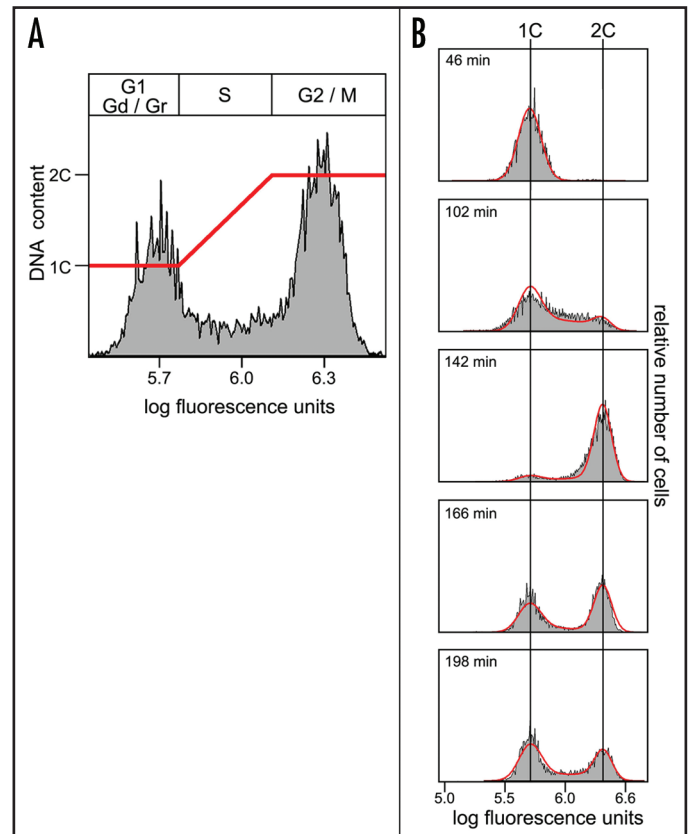


Figure 4. Model prediction of budding index and DNA content. (A) The relationship between cell cycle position from the lifeline (top bar), DNA content (red line), and observed fluorescence values from flow cytometric analysis (gray region). (B) Observed fluorescence histograms at multiple time points from the synchrony experiment shown in Figure 1 (shaded gray), overlaid with model predictions for fluorescence histograms (red line) using parameter values fitted to budding index data.

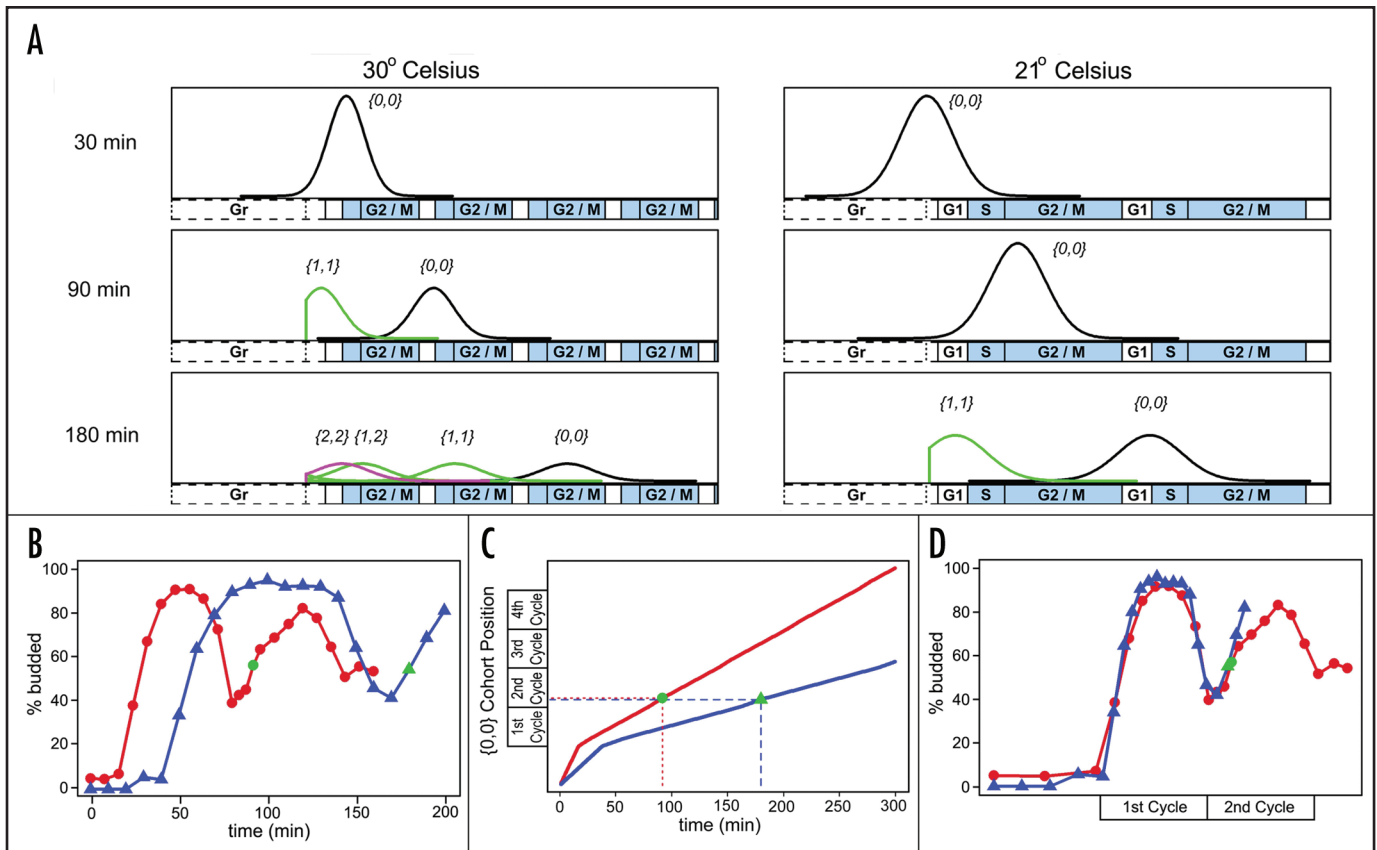


Figure 5. Model-based alignment of time course data from synchrony experiments. (A) The lifeline distributions at three time points for cells synchronized with α -factor incubated at either 30°C (left side) or 21°C. (B and C) Alignment of budding curves for the two experiments (red line = 30°C, blue line = 21°C), based on clock time. Time points at which the two populations exhibit similar distributions (see A) are highlighted (green dots; 90 min for 30°C, 180 min for 21°C). (D) Alignment of budding curves to the lifeline using the position of the mean of the $\{0,0\}$ cohort for mapping.

asked if our model, fit only by budding data, could predict the DNA content of populations at each time point. A visual examination of flow cytometric data (Fig. 1B) suggests that noise associated with measurement of DNA content results in distributions around the fluorescence values for the 1C and 2C peaks (corresponding to cells in G_1 and G_2 , respectively), and likely across intermediate values (corresponding to cells in S). We modeled measurement noise as normal distributions around true values with a variance linearly dependent on measured fluorescence. Using this assumption we were also able to estimate the length of S phase (25%) as a fraction of the total cell cycle (see Materials and Methods).

With corrections for DNA content measurement errors and an estimate for the length of S phase, we used the budding data from the experiment shown in (Fig. 1A) and its associated model parameters (Table 1, Column A) to predict DNA content at each time point. Cells mapping to Gr, Gd, and G_1 were assumed to have a fluorescence value corresponding to a DNA content of 1C. Similarly, cells mapping to G_2 or M were assumed to have a fluorescence value equal to a DNA content of 2C. For cells in S, we used a simple linear mapping, so that a cell 50% through S is mapped to a fluorescence value corresponding to a DNA content of 1.5C (Fig. 4A).

At each of the measurement time points, we found a good correlation between the observed DNA content (Fig. 4B, Supplemental Movie 1, gray filled), and the corresponding predictions by model estimated parameters (Fig. 4B, Supplemental Movie 1, red line). Our findings indicate that the model, fit solely on budding index

data, can accurately predict the distribution of cells with respect to an independent measure of cell cycle position.

Comparing quantitative measurements from experiments performed under different growth conditions. Qualitatively, measurements of dynamic processes during the cell cycle are generally reproducible. However, even when experimental conditions are tightly controlled, the populations in synchrony experiments rarely progress into and through the cell cycle with identical kinetics. Several parameters often vary from experiment to experiment including recovery time (μ_0) and cell cycle length (λ), and daughter-specific delay (δ). Even small variations these times can have profound effects when attempting to compare measured values from time point to time point across different experiments.

Rather than directly comparing values as a function of clock time, we can exploit the model's ability to predict cell cycle distributions for different experiments, and use these distributions to align data based on population distribution state rather than clock time. As a proof of principle experiment, we began by using the model to predict lifeline distributions for two α -factor synchrony experiments performed at different temperatures (Fig. 5). It is immediately apparent that the cell cycle distributions are very different between the two experiments at the same clock times (Fig. 5A) and that aligning the budding data from these two experiments based on clock time (Fig. 5B) does not allow for accurate biological inference.

We map values from one experiment into the other, using the lifeline position of the mean of the $\{0,0\}$ cohort, normalized by cell

cycle length. For example a value of 1.5 would describe a population where the mean of the $\{0,0\}$ cohort is halfway through its second cell cycle, whereas a value of 0.75 would describe a population only 75% through its first cell cycle. The cell cycle distributions of two populations are comparable at time points where the location of the $\{0,0\}$ cohort, and therefore the values for the descriptor, are the same. The ability to accurately align experiments based on the position of the $\{0,0\}$ cohort diminishes as the estimates for δ begin to vary significantly between experiments; as δ diverges, the correlation between $\{0,0\}$ position and overall population distribution between two experiments becomes worse, especially at later time points. Nevertheless, we constructed curves for α -factor synchrony experiments performed at 30°C and 21°C, defining the $\{0,0\}$ position as a function of clock time (Fig. 5C). These curves illustrate how two different experiments can have the same cell cycle distributions at very different clock times (Fig. 5C, dashed lines). Using these curves, we realigned the budding data (Fig. 5D) so that relevant comparisons could be made based on lifeline position, rather than clock time.

Comparing time-series transcription data from *S. cerevisiae* and *S. pombe*. We have described the CLOCCS model in the context of learning cell cycle distributions based on the budding status of *S. cerevisiae*. However, the flexibility of the model allows it to be fit using any cell cycle event that can be measured on single cells, and where we know when and for how long in the cell cycle this event occurs.

We fit the model to time series data collected from the unrelated yeast species, *S. pombe*. Despite a high degree of conservation in the cell cycle machinery, cell division in *S. pombe* is fundamentally different from *S. cerevisiae*. First, *S. pombe* does not bud, but reproduces by binary fission. Thus, divisions are not asymmetric. The cell cycle time is somewhat longer than *S. cerevisiae* under standard growth conditions, and unlike *S. cerevisiae*, *S. pombe* cells spend the majority of their time in the G_2 phase of the cell cycle. Nonetheless, cell cycle position in *S. pombe* can be monitored by measuring septation index. *S. pombe* cells begin to form a septum, which is a precursor to the cytokinetic ring, at the end of mitosis. Septation persists through the short G_1 , and is lost part way into S-phase when cells separate. We fit CLOCCS model parameters to septation data from a time series experiment where genome-wide transcription was measured.²⁷ Details of the model fitting for *S. pombe* can be found in Materials and Methods.

We also fit CLOCCS parameters to budding data from recently published time-series transcription experiments in *S. cerevisiae*.²⁸ Model fits for budding in *S. cerevisiae* experiments and septation in *S. pombe* experiments are shown in (Fig. 6A and B) (respectively) and parameter estimates are in Supplemental Table 1. Budding and septation curves were then mapped onto the same cell cycle lifeline as described above, with blue curves representing budding in *S. cerevisiae*, and red curves representing septation in *S. pombe* (Fig. 6C and D). Because budding and septation occur in different phases of the cell cycle, these curves are not expected to align.

Using the model based alignment, we were able to directly compare the transcription curves for several cell cycle-regulated orthologs. By visually comparing transcription data from *S. cerevisiae* and *S. pombe*, published reports suggested that several orthologs were transcribed in similar cell cycle phases.^{27,29,30} Our results demonstrate a good correlation between the expression of orthologs in G_1 (RNR1/CDC22 and CLB6/CIG2), S-phase (HHF1/HHF1 and HHT1/HHT1), and late mitosis (SIC1/RUM1) (Fig. 6E). However, our alignments suggest that genes expressed in G_2 (CDC20/SLP1 and CDC5/PLO1) as well as some members of the late mitotic cluster (DBF2/SID2) are transcribed earlier in *S. cerevisiae* than in *S. pombe*.

DISCUSSION

The study of dynamic processes driving cell division often requires synchronization of cell populations, and we set out to obtain an accurate description of how cells in a synchronized population are distributed in the cell cycle at times following synchrony release. We have developed a closed-form, parametric, mathematical model, CLOCCS, that describes how an initially semi-synchronous population loses synchrony over time. The model captures three distinct sources of asynchrony: asynchrony in the initial population, asynchrony caused by variance in rates of progression of individual cells through the cell cycle, and asynchrony caused by asymmetric cell divisions. Our linear lifeline accommodates population effects related to reproduction, and allows for offspring to be separated into distinct cohorts.

That a population of cells is made up of a collection of distinct subpopulations, or cohorts, is a crucial aspect of our model. The cell cycle period is extended for daughter cells as compared to mother cells, introducing population asynchrony as a result of reproduction.^{18,19} Separating the population into cohorts such that lifeline distribution of cells in each cohort can be determined, and describing the reproductive dynamics that control the relative contribution of each cohort to the entire population, allows us to include these reproduction-based effects when determining the distribution of the whole population over the lifeline. To retain a closed-form model, we introduced two latent variables, g and r , that index distinct cohorts. The model assumes that, apart from the daughter-specific G_d period, cell cycles are the same for each cohort of cells as they move from one cycle to the next along the lifeline. In support of this assumption, previous observations suggest that first-time mothers and multi-time mothers progress through G_1 with similar kinetics.^{18,19,21,25}

We have shown that the model accurately fits data collected from experiments that employ different synchrony methods or growth conditions, and that the values for parameters within an experiment type are very similar. Furthermore, the model, fit only on budding data, can accurately predict the distribution of cells based on DNA content, an independent measure of cell cycle position (Supplemental Movie 2 and Supplemental Table 2). In the future we might simultaneously use several independent measures of cell cycle position for more robust parameter estimation.

Biological significance of model parameters. The parameters for the model were chosen to capture sources of asynchrony stemming from well-established biological processes. All of the parameters are mathematically identifiable, and the estimates are consistent from experiment to experiment under the same conditions. In the case of experiments run at two different temperatures, the model correctly accommodates the slower growth rate of cells at lower temperatures by extending the cell cycle time λ (Table 1). Parameter estimates suggest that a daughter cell cycle is almost 50% longer than a mother cell cycle (δ/λ , Table 1A and B). In agreement with this model prediction, previous time-lapse studies on individual cells demonstrated that a daughter cell cycle was approximately 40–50% longer than a mother cell cycle.¹⁸

Values for δ are significantly lower for populations synchronized by α -factor, but as discussed above, these lower values likely reflect the very large cell size of the initial population. The initial assumption that δ is constant is probably not entirely accurate for populations synchronized by α -factor; it is likely that as these populations undergo multiple cell divisions and return to normal cell sizes, the values for δ will eventually increase to those observed for elutriated cells.⁸

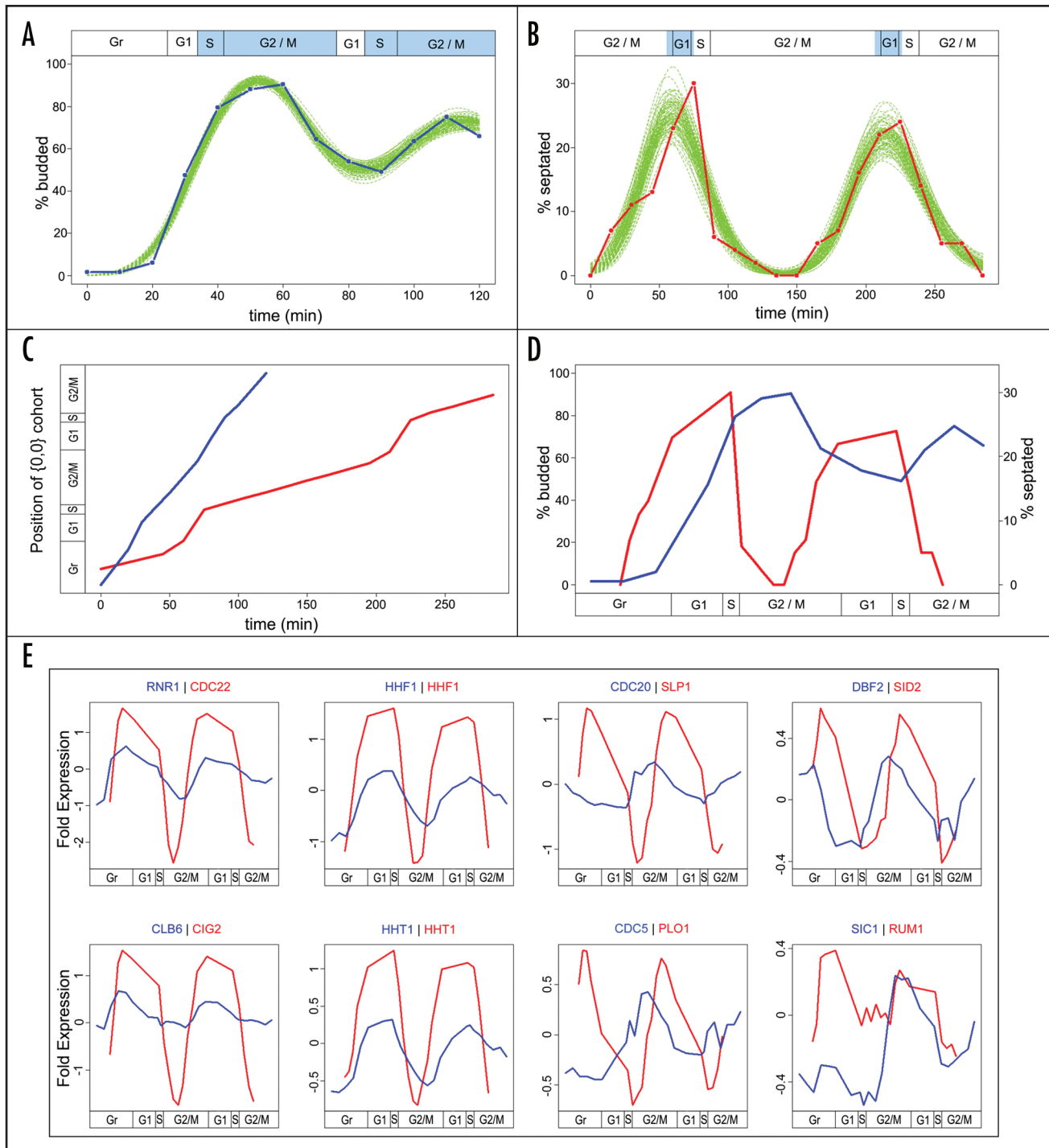


Figure 6. CLOCCS based alignment across organisms. Model estimates were generated based on the budding index curve for *S. cerevisiae*²⁷ (A, blue line) and the septation index curve for *S. pombe*²⁸ (B, red line) (see Materials and Methods for details). The light blue regions in the lifelines above the curves show the areas of the cell cycle where budding and septation occur, respectively. Using the mapping from time to {0,0} position for both experiments (C) the two curves were aligned (D). Since budding and septation do not occur at the same points in the cell cycle we do not expect these two curves to overlap. Using the same mapping function (C), we align the mean-normalized log₂-transformed transcriptional profiles for a set of gene homologs with the *S. cerevisiae* gene name on the left and the expression profile shown in blue, and the *S. pombe* gene name on the right and the expression profile shown in red.

Estimated parameter values appear to match expectations based on previous experimental observations, providing support for the ability of the model to describe the behavior of cells in a population. However, because the parameters may also capture unrecognized sources of asynchrony, the precise biological significance of their values should be interpreted with caution.

Relative contributions of different sources of asynchrony. Since the values for variance in velocity (σ_v^2) are relatively small, our parameter estimates suggest that initial asynchrony (σ_0^2) and asynchrony related to daughter-specific delays in G₁ (δ) are likely to be the most significant contributors to population asynchrony. This prediction is in agreement with single cell studies that demonstrate

that the variance in the cell cycle time of mothers is relatively small with respect to the variance observed between the cell cycle periods of mother and daughter cells.^{18,19,21,25}

These findings highlight the importance of including synchrony loss due to asymmetric cell division in mathematical models. For example, the only source of asynchrony in the model of Bar-Joseph and colleagues is variation in the velocities of cells progressing through the cycle.⁴ The increased complexity of our model, especially the inclusion of explicit reproduction dynamics, allows us to capture the underlying process better than is possible with a single asynchrony source model.

Applications of the model. As discussed above, because synchrony experiments are not precisely reproducible, populations in different experiments are not identically distributed across the cell cycle at each point in time. Thus, values for measured cell cycle parameters at the same time point over multiple experiments are not necessarily comparable, and statistical analysis cannot be readily applied to data from multiple experiments. To circumvent this limitation, values are often reported from a single experiment, with qualitative validation from duplicate experiments. However, by mapping time points onto a lifeline using our CLOCCS model (Fig. 5), relevant values can be compared across multiple experiments, and the statistical significance of the measurements across experiments can be assessed.

We have also shown that the CLOCCS model can be used to compare data sets across different species. The formulation of the CLOCCS model is flexible, so that it can be conditioned on other cell cycle measurements such as septation index in *S. pombe*. Once the model parameters have been fit, quantitative measurements can be aligned and compared directly. Previous studies reported correlations between the cell cycle transcription patterns of *S. cerevisiae* and *S. pombe*, although these crude correlations were made by a simple visual examination of the data sets.^{27,29,30} Our findings indicate a high degree of similarity between orthologs that are expressed in G₁ and S-phase of the cell cycle. But they also suggest that some *S. cerevisiae* orthologs expressed in G₂/M are transcribed relatively earlier than their *S. pombe* counterparts. This difference in the pattern of expression timing may reflect the fact that *S. pombe* cells have an extended G₂ period, and thus there is a discontinuity between the locations of the G₂/M border on the *S. cerevisiae* and *S. pombe* lifelines. Our findings support the notion that at least some facets of the cell-cycle-regulated transcriptional program are conserved between *S. cerevisiae* and *S. pombe*, although a more thorough examination of these data sets in the future may reveal interesting differences.

The very fact that synchronized populations lose synchrony over time also prevents the accurate measurement of cell cycle processes that occur in discrete cell cycle phases. Because populations are often distributed across multiple cell cycle phases at each time point, the measured value of any parameter represents a convolution of both the true value for cells at a given cell cycle position, along with values associated with cells distributed in other cell cycle phases. By learning the distribution of cells at each time point, the convolution kernel can be determined. With that kernel in hand, it is possible to derive estimates for any process at distinct cell cycle positions by employing a deconvolution algorithm. The CLOCCS model represents a practical advance in that it can easily be incorporated into such deconvolution algorithms, and its closed-form will help the time complexity of such algorithms. Thus, our work represents a significant step towards the accurate quantification of dynamic cell cycle processes assayed by synchrony protocols.

Notes

Supplemental material can be found at
www.landesbioscience.com/supplement/OrlandoCC6-4-Sup1.pdf
www.landesbioscience.com/supplement/OrlandoCC6-4-Sup2.mpg
www.landesbioscience.com/supplement/OrlandoCC6-4-Sup3.mpg

References

- Shedden K, Cooper S. Analysis of cell-cycle gene expression in *Saccharomyces cerevisiae* using microarrays and multiple synchronization methods. *Nucleic Acids Res* 2002; 30:2920-9.
- Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* 2002; 13:1977-2000.
- Lu R, Nakorchevskiy A, Marcotte EM. Expression deconvolution: A reinterpretation of DNA microarray data reveals dynamic changes in cell populations. *PNAS* 2003; 100:10370-75.
- Bar-Joseph Z, Farkash S, Gifford DK, Simon I, Rosenfeld R. Deconvolving cell cycle expression data with complementary information. *Bioinformatics* 2004; 20(Suppl 1):123-130.
- Qiu P, Wang ZJ, Liu JR. Polynomial model approach for resynchronization analysis of cell-cycle gene expression data. *Bioinformatics* 2006; 22:959-66.
- Hartwell L. Genetic control of the cell division cycle in yeast. *Experimental Cell Res* 1971; 69:265-76.
- Gordon CN, Elliott SC. Fractionation of *Saccharomyces cerevisiae* cell populations by centrifugal elutriation. *J Bacteriol* 1977; 129:97-100.
- Johnston GC, Pringle JP, Hartwell LH. Coordination of growth with cell division in the yeast *S. cerevisiae*. *Exp Cell Res* 1977; 105:75-98.
- Haase SB, Reed SI. Improved flow cytometric analysis of the budding yeast cell cycle. *Cell Cycle* 2002; 1:132-6.
- Slater ML, Sharrow SO, Gart JJ. Cell cycle of *Saccharomyces cerevisiae* in populations growing at different rates. *Proc Natl Acad Sci USA* 1997; 74:3850-4.
- Tobey RA, Crissman HA. Unique techniques for cell analysis utilizing mithramycin and flow microfluorometry. *Exp Cell Res* 1975; 93:235-9.
- Niemisto A, Nykter M, Aho T, Jalovaara H, Marjanen K, et al. Distribution estimation of synchronized budding yeast population. Proceedings of the Winter International Symposium on Information and Communication Technologies 2004. Cancun, 2004:243-48.
- Arino O. A survey of structured cell population dynamics. *Acta Biotheor* 1995; 43:3-25.
- Henson MA. Dynamic modeling of microbial cell populations. *Curr Opin Biotechnol* 2003; 14:460-7.
- Sidoli FR, Mantalaris A, Asprey SP. Modelling of mammalian cells and cell culture processes. *Cytotechnology* 2004; 44:27-46.
- Liou JJ, Srien F, Fredrickson AG. Solutions of population balance models based on a successive generations approach. *Chem Eng Sci* 1997; 52:1529-40.
- Alexanderson M. On the existence of the stable birth-type distribution in a general branching process cell cycle model with unequal cell division. *J Appl Probab* 2001; 38:685-95.
- Hartwell L, Unger M. Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *J Cell Biol* 1977; 75:422-35.
- Lord PG, Wheals AE. Asymmetrical division of *Saccharomyces cerevisiae*. *J Bacteriol* 1980; 142:808-18.
- Lord PG, Wheals AE. Variability in individual cell cycles of *Saccharomyces cerevisiae*. *J Cell Sci* 1981; 50:361-76.
- Woldringh CL, Huls PG, Vischer NO. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* α as determined by image cytometry. *J Bacteriol* 1993; 175:3174-81.
- Rowley A, Johnston GC, Butler B, Werner-Washburne M, Singer RA. Heat shock-mediated cell cycle blockage and G₁ cyclin expression in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 1993; 13:1034-41.
- Belli G, Gari E, Aldea M, Herrero E. Osmotic stress causes a G₁ cell cycle delay and downregulation of cln3/cdc28 activity in *Saccharomyces cerevisiae*. *Mol Microbiol* 2007; 39:1022-35.
- Barford JR, Hall RJ. Estimation of the length of cell cycle phases from asynchronous cultures of *Saccharomyces cerevisiae*. *Exp Cell Res* 1976; 102:276-84.
- Bean JM, Siggia ED, Cross FR. Coherence and timing of cell cycle start examined at single-cell resolution. *Mol Cell* 2006; 21:3-14.
- Jorgensen P, Tyers M. How cells coordinate growth and division. *Curr Biol* 2004; 14:R1014-27.
- Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, et al. Periodic gene expression program of the fission yeast cell cycle. *Nat Genet* 2004; 36:809-17.
- Pramila T, Wu W, Miles S, Noble WS, Breeden LL. The forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev* 2006; 20:2266-78.
- Sherlock G. Starting to recycle. *Nat Genet* 2004; 36:795-6.
- Peng X, Karuturi RK, Miller LD, Lin K, Jia Y, et al. Identification of cell cycle-regulated genes in fission yeast. *Mol Biol Cell* 2005; 16:1026-42.
- Haase SB, Reed SI. Evidence that a free-running oscillator drives G₁ events in the budding yeast cell cycle. *Nature* 1999; 401:394-7.
- Gilks WR, Richardson S, Spiegelhalter DJ eds. Markov Chain Monte Carlo in Practice. London: Chapman and Hall, 1995.
- Rafferty AE, Lewis SM. Implementing MCMC. Markov Chain Monte Carlo in Practice. London: Chapman and Hall, 1996:115-30.
- Heidelberg P, Welch PD. Simulation run length control in the presence of an initial transient. *Oper Res* 1983; 31:1109-44.
- Geweke J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. Bayesian Statistics 4. Oxford, UK: Clarendon Press, 1992.