

Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing

Scott Cheng-Hsin Yang^{1,*}, Nicholas Rhind² and John Bechhoefer¹

¹ Department of Physics, Simon Fraser University, Burnaby, British Columbia, Canada and ² Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA

* Corresponding author. Department of Physics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6.
Tel.: +1 778 782 5924; Fax: +1 778 782 3592; E-mail: scotty@sfu.ca

Received 25.11.09; accepted 16.7.10

Microarrays are powerful tools to probe genome-wide replication kinetics. The rich data sets that result contain more information than has been extracted by current methods of analysis. In this paper, we present an analytical model that incorporates probabilistic initiation of origins and passive replication. Using the model, we performed least-squares fits to a set of recently published time course microarray data on *Saccharomyces cerevisiae*. We extracted the distribution of firing times for each origin and found that the later an origin fires on average, the greater the variation in firing times. To explain this trend, we propose a model where earlier-firing origins have more initiator complexes loaded and a more accessible chromatin environment. The model demonstrates how initiation can be stochastic and yet occur at defined times during S phase, without an explicit timing program. Furthermore, we hypothesize that the initiators in this model correspond to loaded minichromosome maintenance complexes. This model is the first to suggest a detailed, testable, biochemically plausible mechanism for the regulation of replication timing in eukaryotes.

Molecular Systems Biology 6: 404; published online 24 August 2010; doi:10.1038/msb.2010.61

Subject Categories: simulation and data analysis; cell cycle

Keywords: DNA replication program; genome-wide analysis; microarray data; replication-origin efficiency; stochastic modeling

This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial No Derivative Works 3.0 Unported License, which permits distribution and reproduction in any medium, provided the original author and source are credited. This license does not permit commercial exploitation or the creation of derivative works without specific permission.

Introduction

The kinetics of DNA replication in eukaryotic cells are carefully controlled, with some parts of the genome replicating early and others replicating later. Patterns of replication timing correlate with gene expression, chromatin structure, and subnuclear localization, suggesting that replication timing may have an important function organizing the nucleus and regulating its function (Goren and Cedar, 2003). The timing of DNA replication is regulated largely by the timing of replication-origin activation. Although the biochemical steps of origin firing are increasingly well understood, the regulation that leads to defined patterns of replication timing is still a mystery (Scalfani and Holzen, 2007).

Over the past 15 years, the development of two new technologies has led to significant progress in the description of DNA replication kinetics. The first development, molecular combing, is a single-molecule technique in which one labels replicated and unreplicated DNA and then stretches out the fibers to detect replication patterns (Herrick and Bensimon, 1999; Herrick *et al.*, 2000). The technique yields 'snapshots' of the replication state of individual fragments of chromosomes.

Advantages include high spatial resolution (≈ 2 kb) and data from individual DNA molecules, rather than ensemble averages (Patel *et al.*, 2006; Czajkowsky *et al.*, 2008). The main disadvantage is that the size of combed fragments (100 kb to 1 Mb) is small relative to the typical size of full chromosomes, which limits the information available about patterns of replication timing on larger scales and complicates analysis (Zhang and Bechhoefer, 2006).

The second development applies microarrays to probe ensemble averages of replication extent across the genome (Raghuraman *et al.*, 2001; Yabuki *et al.*, 2002; Feng *et al.*, 2006; Heichinger *et al.*, 2006). Compared with molecular combing, the main disadvantages of microarrays are the loss of information about cell-to-cell variability and the need for complicated post-signal processing. On the other hand, the main advantages are the experiment's high throughput and complete coverage of the genome that result in both temporally and spatially resolved data. The availability of high-resolution, genome-wide data allows one to measure and model variations in firing time and origin efficiency at a level of detail that exceeds that provided by past studies.

The common feature of both technologies is their ability to generate large data sets for the analysis of replication. With large data sets comes the potential to construct a much more detailed picture of how replication occurs and is regulated (Hyrien and Goldar, 2010). For molecular-combing experiments, previous work has shown that quantitative information such as replication-origin firing rates and replication-fork velocities can be extracted (Yang *et al*, 2009). Such information has led to an appreciation of the function of stochastic effects in initiation (Herrick and Bensimon, 1999), to a greater understanding of the ‘random-completion problem’ for embryonic replication (Yang and Bechhoefer, 2008), to models highlighting searching and binding kinetics in initiation timing (Gauthier and Bechhoefer, 2009), and to suggestions that initiation patterns may be universal across species (Goldar *et al*, 2009).

A prevailing picture of eukaryotic replication is that origins are positioned at defined sites and initiated at preprogrammed times (Donaldson, 2005). In *Saccharomyces cerevisiae*, origin positions are defined, in part, by 11–17 base pair autonomous replicating sequence (ARS) consensus motifs, which bind the origin recognition complex (ORC) (Sclafani and Holzen, 2007). The timing of origin initiation is more controversial: although microarray experiments have generally been interpreted based on the assumption of a deterministic temporal program (Raghuraman *et al*, 2001), recent molecular-combing experiments suggest that origin initiation is stochastic (Patel *et al*, 2006; Czajkowsky *et al*, 2008). A more subtle issue is that the consequences of the spatiotemporal connection between origin initiation and fork progression in multiple-origin systems are often not taken into account. The replication extent of an origin site is sometimes implicitly assumed to be solely from the origin itself (Eshaghi *et al*, 2007), even though the locus can also be replicated by nearby origins. Our thesis is that a more rigorous analysis of microarray replication data based on the same type of stochastic models used to understand combing experiments can yield greater insight into the replication process and contribute to forming a more accurate picture of the replication program. We argue that both microarray and combing experiments are compatible with stochastic origin initiation and that the apparent disagreement is resolved after performing a more sophisticated analysis of the microarray data.

The analytical model of DNA replication we present extends previous work that focused on replication in *Xenopus laevis* (Jun *et al*, 2005). The present version includes defined origin position, variable-elongation rates, and probabilistic initiation. Our model is sufficiently general to describe both deterministic and stochastic replication-kinetics scenarios and assumes only that initiation events are not correlated. To show the kinds of questions our model can address, we fit the model to a recent set of time course microarray data on *S. cerevisiae* (McCune *et al*, 2008) and extracted origin positions, the firing-time distribution for each origin, origin efficiencies, and a global fork velocity. We found that the trend of origin firing embedded in the data is incompatible with a naive deterministic model. Based on this finding, we propose a stochastic-initiator model where temporal patterns in replication can arise in the absence of an explicit mechanism for controlling the timing of origin initiation. The details of this

model support a specific molecular mechanism for the regulation of replication timing, in which the number of minichromosome maintenance (MCM) complexes loaded at an origin regulates, in part, the origin firing time.

More generally, the model can be used to reconstruct the complete spatiotemporal program quantitatively, as reflected by its ability to fit genome-wide microarray data. It is less biased than current empirical models, as it accounts for the effects of passive replication (Raghuraman *et al*, 2001; Eshaghi *et al*, 2007). As it is analytic, it is also faster than simulation-based models (Lygeros *et al*, 2008; Blow and Ge, 2009; de Moura *et al*, 2010). For these reasons, we believe the model introduced here to be a powerful tool for analyzing spatio-temporally resolved DNA replication data. Moreover, the probabilistic nature of the model allows analysis of a broad range of dynamics, from deterministic to completely random timing, and thus offers new ways to look at replication. In particular, we show that reproducible replication patterns do not necessarily indicate a temporal program where the initiation involves time-measuring mechanisms; rather, temporal order can emerge as a consequence of stochastic initiations.

Results

The time and precision of origin firing are correlated

To investigate the regulation of replication kinetics, we have developed a general mathematical model of replication and used it to analyze the recent budding yeast replication time course data of McCune *et al* (2008). Such microarray experiments yield the fraction of cells in a population that have a specific site replicated after a given time in S phase (see Supplementary Material Section A for more experimental details). The positions of origins are determined by peaks in the graphs of replication fraction, as an origin site is replicated before its neighboring sites (Figure 1). Following standard analytical methods, we define a time t_{rep} as the time by which half of the cells show replication of the origin locus (Raghuraman *et al*, 2001). Implicit in this picture are the assumptions that the variation of firing times of an origin, t_{width} , is small and that t_{rep} is independent of t_{width} .

A simple analysis of the data suggests a more complex picture. We first analyzed the data by fitting a sigmoid through the time course associated with each origin (Figure 1). Each sigmoid yields a t_{rep} and t_{width} (defined as the time for the origin locus to go from 25 to 75% replicated). In contrast to the deterministic timing scenario that assumes t_{width} is much less than t_{rep} , the extracted t_{width} approximately equals t_{rep} (Figure 1C). The apparent variability in origin timings suggests that stochasticity is important to an accurate description of the replication kinetics. Moreover, in a model where the timing of origin firing is regulated by external triggers (Goren and Cedar, 2003), one expects no correlation between t_{rep} and t_{width} . In other words, the t_{width} points would scatter about a horizontal line in Figure 1C, implying that variations in t_{width} are independent of t_{rep} . Instead, we see a strong correlation between t_{width} and t_{rep} , suggesting the two are mechanistically related.

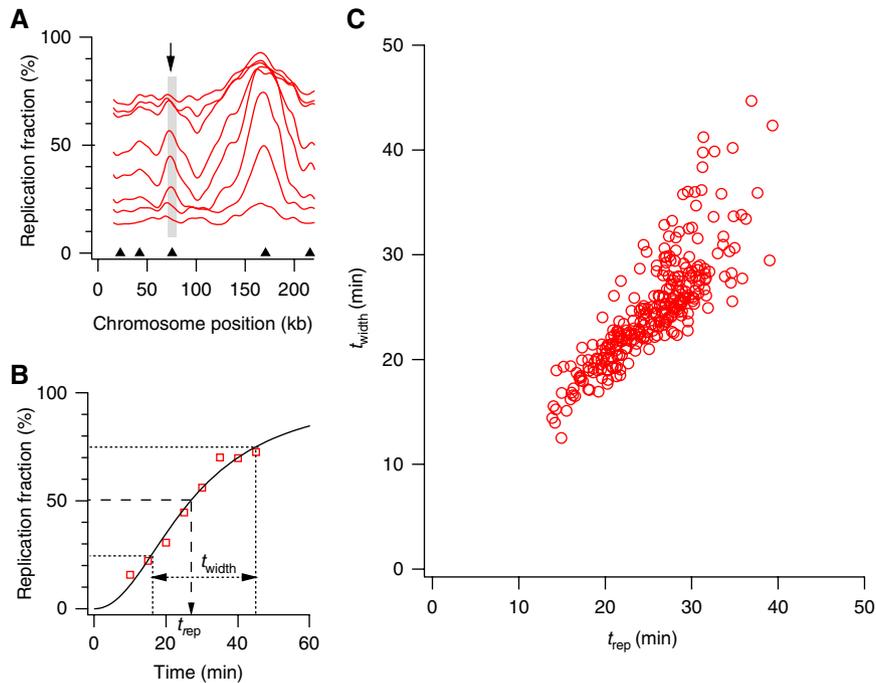


Figure 1 (A) Smoothed time course microarray data for chromosome I. Triangles are origins identified in Alvino *et al* (2007). The arrow indicates the origin analyzed in (B). (B) Replication fraction of the indicated origin versus time. Fitting the data with Equation (1) gives the median time t_{rep} and the 25–75% time width t_{width} . (C) Time widths versus median times for all 275 origin loci identified in Alvino *et al* (2007).

An analytical model for replication kinetics

The discrepancies between a naive picture of origin firing control and the data motivate a more detailed approach. We have developed an analytical model that can generate genome-wide time course data sets comparable to those from microarray-based replication experiments (see Materials and methods). The model takes into account probabilistic initiation and the effects of passive replication. Its parameters can be separated into a ‘local’ group that describes the properties of each origin and a ‘genome-wide’ group that describes quantities that are roughly constant across the genome. Limitations of the model and the microarray data used are discussed in the Discussion.

We first introduce the ‘sigmoid model’ (SM), which describes each origin with three local parameters: the origin position, the median time of the firing-time distribution ($t_{1/2}$), and the width of the firing-time distribution (t_w , defined as $t_{0.75} - t_{0.25}$). The last two parameters define a sigmoidal curve for the cumulative firing-time distribution. These two parameters $t_{1/2}$ and t_w differ from t_{rep} and t_{width} . The former describes the *firing time* of an origin, whereas the latter describes the *replication time* of an origin site, which is replicated by both the origin on the site and nearby origins. The fit to chromosome XI is shown in Figure 2 to demonstrate that the model captures the replication process well. Fits to all chromosomes can be found in Supplementary Figure 1. A detailed statistical analysis of the fits is presented in Supplementary Material Section B.

Extracted origins and fork velocity

The initial list of origins consisted of 732 positions from the OriDB database that had previously been identified using a

variety of methods (Nieduszynski *et al*, 2007). The results do not depend sensitively on this initial list, as we allowed the positions to vary in the fit. After eliminating origins according to the criteria described in Materials and methods, the SM gave 342 origins (origin parameters tabulated in Supplementary Table I). Out of the 342 we identify, 236 colocalize with the 275 origins identified previously using a similar data set (Alvino *et al*, 2007). The remaining 106 origins were not previously recognized by Alvino *et al*, mostly because they are not associated with apparent peaks in the microarray data. We found that 75% of the 106 colocalize (within 5 kb) with origins in the OriDB database (Nieduszynski *et al*, 2007).

The SM can be modified to include variable-fork rates across the genome (see Supplementary Material Section C). Fitting the data with a model that uses spatially varying fork velocities, we found that both the constant-velocity SM and variable-velocity SM fit the data well (Supplementary Figures 6 and 7). Both capture >98% of the variance in the data (see Supplementary Material Section C). This suggests that a spatially constant velocity is plausible. The genome-wide fork velocity we extracted from the SM is 1.9 kb/min. Our statistical analysis of the fits suggests an error of 0.2 kb/min on v (see Supplementary Material Section B). Consistent with this conclusion, recent work using ChIP-chip to monitor the movement of GINS, an integral member of replication forks, shows a fork progression rate of about 1.6 ± 0.3 kb/min that does not vary significantly across the genome (Sekedat *et al*, 2010). Our conclusion contrasts with that of Raghuraman *et al* (2001), where variations in slope from peak to trough were interpreted as fork-velocity variations. In our model, these variations can be mostly accounted for by different levels of passive replication from origins with different distributions of

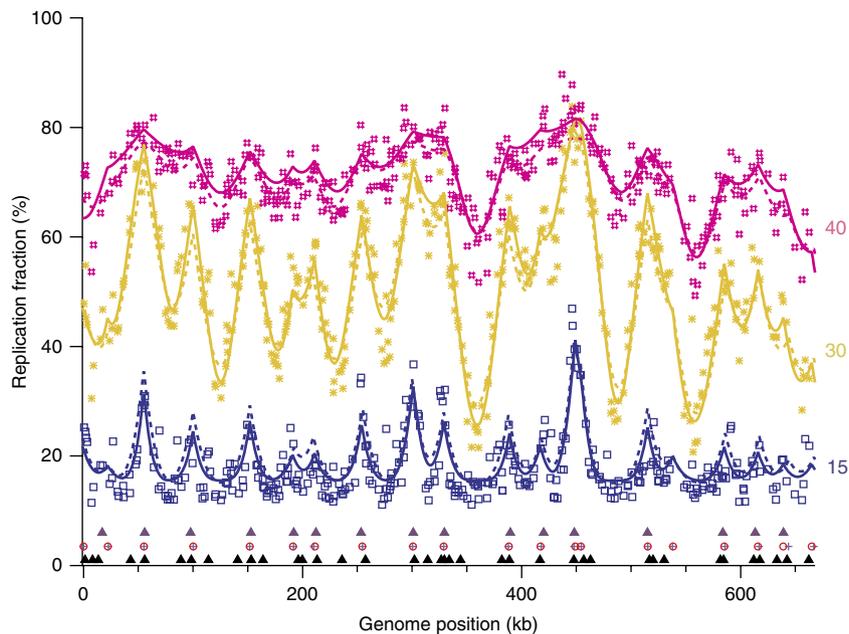


Figure 2 Chromosome-XI section of the genome-wide fits. Markers are data, solid lines are fits from SM, and dotted lines are fits from MIM. Upper row of solid triangles at the bottom denote origin positions identified in Alvino *et al* (2007). In the middle row, open circles are estimated origin positions in the SM, whereas crosses are from the MIM. The lower row of triangles correspond to origins in the OriDB database (Nieduszynski *et al*, 2007). The three curves from bottom to top correspond to the replication fraction $f(x)$ at 15, 30, and 40 min after release from the *cdc7-1* block. The data set covers the genome at ≈ 2 kb resolution.

firing times, including the above-mentioned origins that lack distinct peaks.

The SM recapitulates observed firing-time distributions and initiation rates

The least-understood aspect of the replication process is its temporal program (Sclafani and Holzen, 2007). One reason is that there has been no direct way of visualizing both the spatial and temporal aspects of replication at high resolution. Another reason is that the implications of temporal stochasticity in initiations have often been neglected. We extracted the firing-time distributions of the 342 origins identified in the SM (Figure 3A). The widths are comparable to the length of S phase, confirming that stochastic effects have an important function. The kinetic curves of the origins extracted from the model also show that t_w increases with $t_{1/2}$ (Figure 3B and D; compare with Figure 1). Further, we show that four representative, previously studied origins (ARS413, ARS501, ARS606, and ARS1114.5) are typical origins that follow this global $t_{1/2}$ - t_w relationship (Figure 3).

The rate of initiation, defined as the number of initiations per time per unreplicated length, is a crucial parameter in kinetics (see Materials and methods). It has been proposed that an increasing rate of initiation later in S phase would lead to robust completion of replication, even if origin firing is stochastic (Hyrien *et al*, 2003). To investigate the initiation rate in the SM, we plotted initiation rates averaged over the genome and over individual chromosomes (Figure 3C). Our results show the initiation rate rising for most of S phase and then declining in late S phase. A similar pattern has been described in a number of organisms (Goldar *et al*, 2009). However, the

genome-wide-averaged initiation rate we extracted does not decay to zero before S phase ends as it did in the previous analysis. In Goldar *et al* (2009), the use of t_{rep} rather than the full distribution of firing times of an origin leads to an underestimation of origin initiation at late times. Nonetheless, as the proposed universality of the initiation rate across species was for *scaled* initiation rates, that conclusion may well survive reanalysis of the data with a stochastic model that would modify the extracted average rates and scalings.

Geometric effect from passive replication affects variability in origin efficiencies

Understanding origin efficiency is important because these efficiencies determine the replication completion time and the robustness of the replication program (Lygeros *et al*, 2008; Blow and Ge, 2009). The efficiency of an origin is closely related to its geometry—its location relative to other origins. Imagine two highly efficient origins placed near each other; only one of the two origins will fire in any given cell because the initiation of one origin will passively replicate the other origin. Placing an efficient origin next to an inefficient one should decrease the firing of each by different amounts. For an isolated origin, one expects that its efficiency would be unaffected by passive replication.

In our analysis, we distinguish between ‘efficiency,’ which is traditionally defined as the probability that an origin fires during normal replication, taking into account passive replication effects, and ‘potential efficiency.’ The latter term is defined to be the probability that an origin would have initiated before the end of S phase if there were no passive replication (Rhind *et al*, 2010). We will occasionally use

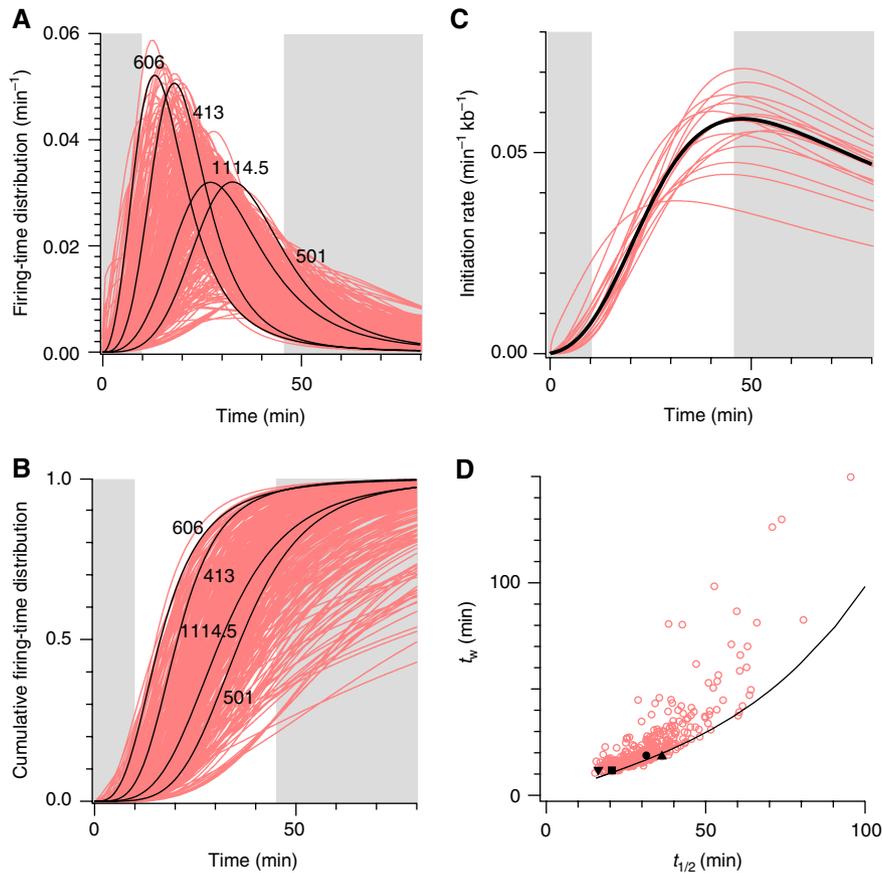


Figure 3 (A) Firing-time distribution for the 342 origins extracted using the SM. Darker curves are representative distributions chosen to illustrate origins with different values of $t_{1/2}$. Numbers denote their ARS number in the OriDB database (Nieduszynski *et al*, 2007). Gray background corresponds to times not included in the data ($t < 10$ and $t > 45$ min); curves in these domains are extrapolated. The length of S phase is roughly 60 min. (B) Cumulative firing-time distributions for the 342 origins. Each trace is the kinetic curve that would be measured for the corresponding origin site if there were no neighboring origins. Darker curves correspond to darker curves in (A). The potential efficiencies of the origins are the curves' values at $t=60$ min. (C) Averaged initiation rates. The heavy dark curve is the genome-wide averaged rate, whereas the lighter curves are chromosome-averaged rates. (D) Trend of $t_{1/2}$ versus t_w . Markers are results from the SM. The curve is the result from the multiple-initiator model (MIM) using $\Phi_o(t)=1/[1+(76/t)^3]$. The values 76 and 3 are the resulting parameters of the genome-wide MIM fit. Solid square, triangle, inverted triangle, and circle correspond to ARS413, 501, 606, and 1114.5, respectively.

'observed efficiency' instead of just 'efficiency' to emphasize the effect of passive replication. Experimentally, techniques that hinder fork progression to avoid passive replication have been applied to extract the potential efficiencies of origins (Heichinger *et al*, 2006). However, such approaches provide only rough estimates of potential efficiency because they block origin firing in late S phase. Thus, to determine whether the observed efficiency of origins is due primarily to their potential efficiency or to their proximity to neighboring origins, we investigated the relationship between efficiency and potential efficiency.

We used the extracted origin positions and firing-time distributions to calculate both the observed efficiencies (via Equation 11) and potential efficiencies of all identified origins (via evaluating $\Phi_i(t=t_{\text{end}})$, the cumulative firing probability distribution in our model, see Materials and methods). Here, t_{end} is the length of S phase, and we estimate it to be 60 min from flow-cytometric determination of replication progress (see Supplementary Material Section A). We found that more than half of the origins have high potential efficiency (>0.9), but the observed efficiencies of origins vary much more

(compare Figure 4A and B). Furthermore, we found that the relation between observed efficiency and potential efficiency can be approximately accounted for by a mean-field argument (see Supplementary Material Section D), where all neighboring origins are replaced by an average neighbor that has the genome-wide-averaged firing-time distribution (Figure 4C). Thus, the efficiency of origins in *S. cerevisiae* can largely be explained by geometric effects because of neighboring origins.

Late-firing origins are less efficient

As seen in Figure 4D, later-firing origins have lower potential efficiency. The monotonic decrease in efficiency with increasing $t_{1/2}$ is a consequence of the t_w -versus- $t_{1/2}$ relationship mentioned above. The larger t_w associated with later-firing origins imply that the chance for them to initiate within S phase is less than that of earlier-firing origins.

A previous experiment reported that although the ARS501 origin is late firing, its kinetics and efficiency resemble that of an early-firing origin (Ferguson *et al*, 1991). The ARS501 kinetics was used to support a scenario where origin activation

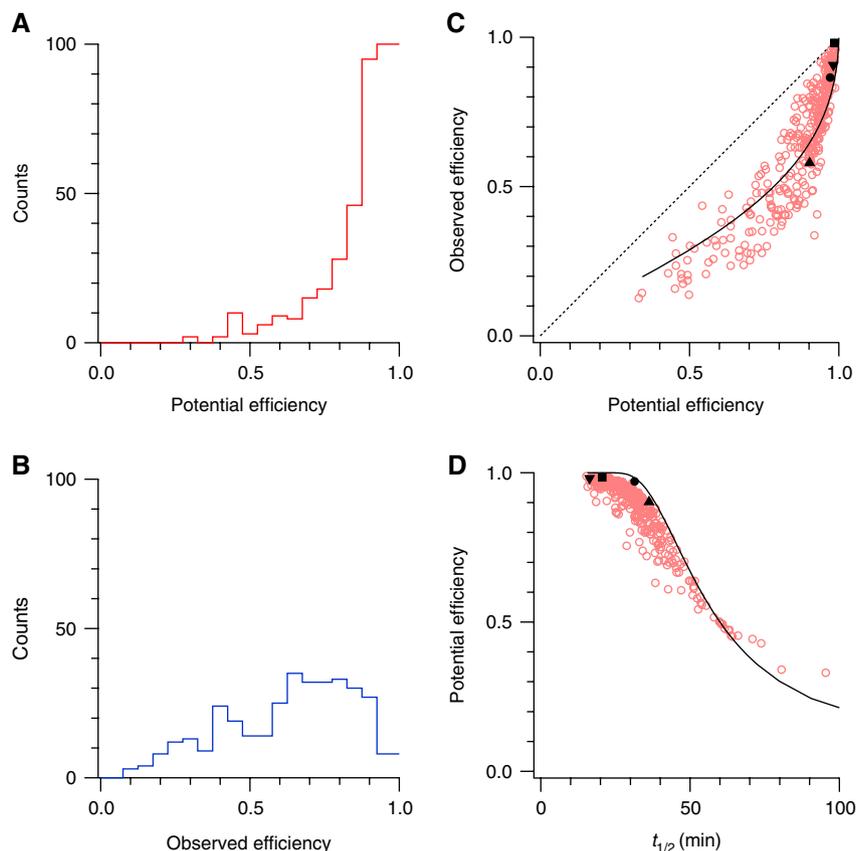


Figure 4 Origin efficiencies. **(A)** Histogram of potential efficiencies with bin width=0.05. More than half of the origins have high potential efficiency (>0.9). Median value=0.91. **(B)** Histogram of observed efficiencies with bin width=0.05. Median value=0.68. **(C)** Observed efficiency versus potential efficiency. Markers would lie on the dotted line if there were no passive replication. Solid markers correspond to the same origins in Figure 3D. The solid curve is a mean-field calculation (see Supplementary Material Section D). **(D)** Potential efficiency versus $t_{1/2}$. There is roughly a one-to-one relationship between the potential efficiency and $t_{1/2}$. Solid markers are as in Figure 3D. The solid curve is calculated using the parameters obtained from the genome-wide MIM fit.

is regulated deterministically in time. We compared the kinetic curves of ARS501 obtained from the earlier slot-blot experiment (Ferguson *et al*, 1991) with that from the present microarray data (McCune *et al*, 2008) and found the two curves differ significantly. The microarray kinetic curve suggests that ARS501 is a typical origin with $t_{rep} \approx 33$ min and $t_{width} \approx 26$ min, whereas the slot-blot data suggest $t_{rep} \approx 33$ and $t_{width} \approx 11$ min (Supplementary Figure 2). For comparing origin properties, we argue that the microarray data is more reliable than the slot-blot data because the former has information about the relative behavior of origins genome wide, whereas the latter contains curves for only a few sites. Simply stated, microarray data contains more relative information about sites than the slot-blot data. According to our analysis, ARS501 is neither unusually efficient (observed efficiency ≈ 0.58) nor unusually late ($t_{1/2} \approx 36$ min).

Temporally alike origins do not cluster

The microarray replication profiles show that different parts of the genome replicate at different times. McCune *et al* studied a mutant yeast strain where origin firings are largely limited to the first half of the cell cycle and found that the typical replication fraction of some regions is unaltered whereas it

decreased in other regions. They thus hypothesized that there are relatively long stretches of chromosomes that replicate early and late and that temporally alike origins are clustered together to form these early- or late-replicating regions (McCune *et al*, 2008).

To investigate this hypothesis, we recall that the distribution of $t_{1/2}$ has a single peak (Figure 5A), suggesting a continuum of typical firing times and arguing against distinct categories of temporally alike origins (e.g. early and late origins). To test the spatial aspect, we calculated the autocorrelation function of the origins' $t_{1/2}$ values (Figure 5B) (Press *et al*, 2007). In this test, a positive value means that, typically, an origin and its n th neighbor are likely to have $t_{1/2}$ that are both larger or both smaller than the average firing time, as would be expected if temporally alike origins were to cluster. Conversely, negative values would mean that an origin and its n th neighbor were likely to have $t_{1/2}$ values that were anti-correlated. We found that the autocorrelation function fluctuates around zero for $t_{1/2}$ but is clearly non-zero at the nearest neighbor for t_{rep} (Figure 5B). The former result indicates that the intrinsic initiation time among origins is independent. The latter result indicates that the time at which neighboring origin sites are replicated is correlated. The correlation arises because a fork from one origin sometimes passively replicates the site of a

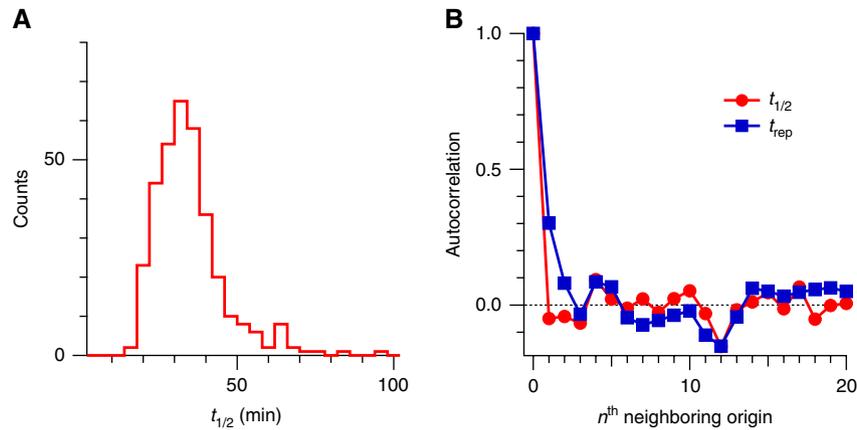


Figure 5 (A) Histogram of $t_{1/2}$ with bin width=4 min. There is no sharp distinction between early and late origins. Median value=31.0 min, s.d.=11.2 min. (B) Autocorrelation of $t_{1/2}$ and t_{rep} . The graph shows the $t_{1/2}$ and t_{rep} correlation between an origin and its n th neighbor. The range of the x axis corresponds to roughly 700 kb of genome distance, as the average interorigin distance extracted is 35 kb.

neighboring origin. Thus, the observation that neighboring origin sites tend to replicate at similar times is consistent with the inference that neighboring origins initiate independently.

The multiple-initiator model can explain the correlation between $t_{1/2}$ and t_w

The SM allows for any combination of $t_{1/2}$ and t_w . The tightness of the trend in Figure 3D is thus remarkable: it suggests that although local variations in origins lead to different $t_{1/2}$ and t_w , replication is organized by a global controlling mechanism that connects the two quantities. In other words, the trend is that earlier-firing origins have tighter control of their initiation timing, whereas later-firing origins show progressively less control. What kind of global and local mechanism can account for such a trend? Here, we propose one plausible scenario.

Replication initiates at origins because there are initiator proteins bound to them. Suppose that every initiator fires according to a global firing-time distribution $\phi_o(t)$. Then, origins that have more initiators loaded will be more efficient. This situation arises because, when multiple initiators bind near an origin site, it is always the earliest firing that counts. Other initiators cannot fire to re-replicate the same site (Sclafani and Holzen, 2007). One can assign an effective firing distribution $\phi_{eff}(t, n)$ to the initiator cluster at an origin, with n being the number of initiators in the cluster. This distribution of the earliest firing times shows the same trend as the curves extracted from the microarray data: earlier-firing origins have narrower distributions (Figure 6A and B).

For moderately large n ($n \geq 10$), the selection of the first initiation among many causes $\phi_{eff}(t, n)$ to tend to a universal distribution, the Weibull distribution, regardless of the ‘details’ of the $\phi_o(t)$ used (Kotz and Nadarajah, 2000). As an example, the shape of $\phi_{eff}(t, n)$ would differ between using an increasing and a decreasing $\phi_o(t)$ but would not alter significantly between a linearly and a quadratically increasing $\phi_o(t)$. This robustness is an advantage of the model because it obviates the need for an accurate form of $\phi_o(t)$.

As the firing-time distribution is shaped by the number of initiators, we call this the ‘multiple-initiator model’ (MIM). Results of the MIM fits are shown in Figure 2 and Supplementary Figure 1. The MIM fits are similar to the SM fits (see Supplementary Material Section B), although the number of parameters used decreased by nearly 1/3. In the SM, each origin has three parameters— x , $t_{1/2}$, and t_w . In the MIM, each origin needs only two— x and n . Furthermore, the trend for t_w versus $t_{1/2}$ are similar between the SM and MIM (Figure 3D). The relationship between potential efficiency and $t_{1/2}$ is also captured (Figure 4D). We emphasize that these similarities are biologically significant because the SM and MIM correspond to two distinct views of replication. The flexibility of allowing each origin its own $t_{1/2}$ and t_w in the SM corresponds to the view that the firing of origins is controlled by an explicit time-counting mechanism, a ‘clock.’ In the MIM, there is no such mechanism; the order in timing is built from stochasticity. Thus, the MIM suggests a new, plausible mechanism for replication timing in *S. cerevisiae*.

Lastly, following what is done for SM, the MIM gave 337 origins (origin parameters tabulated in Supplementary Table II). Out of the 337, 234 colocalize with the 275 origins identified by Alvino *et al* (2007). Out of the remaining 103 origins, 70% colocalize with known origins from the OriDB database (Nieduszynski *et al*, 2007) within 5 kb. Together, the SM and MIM gave 357 distinct origins. Among these, 116 were not identified by Alvino *et al*, and 71% of them colocalize with known origins from the OriDB database within 5 kb.

Discussion

We have developed a general model of replication kinetics and applied it to recently published data from a budding-yeast genome-wide replication time course experiment. We found that origin firing in budding yeast is best described as a stochastic process, in which the average firing time of an origin correlates strongly with its timing precision. The regulation of the timing of origin firing can be explained by the number of initiators loaded at each origin. Our model provides a specific,

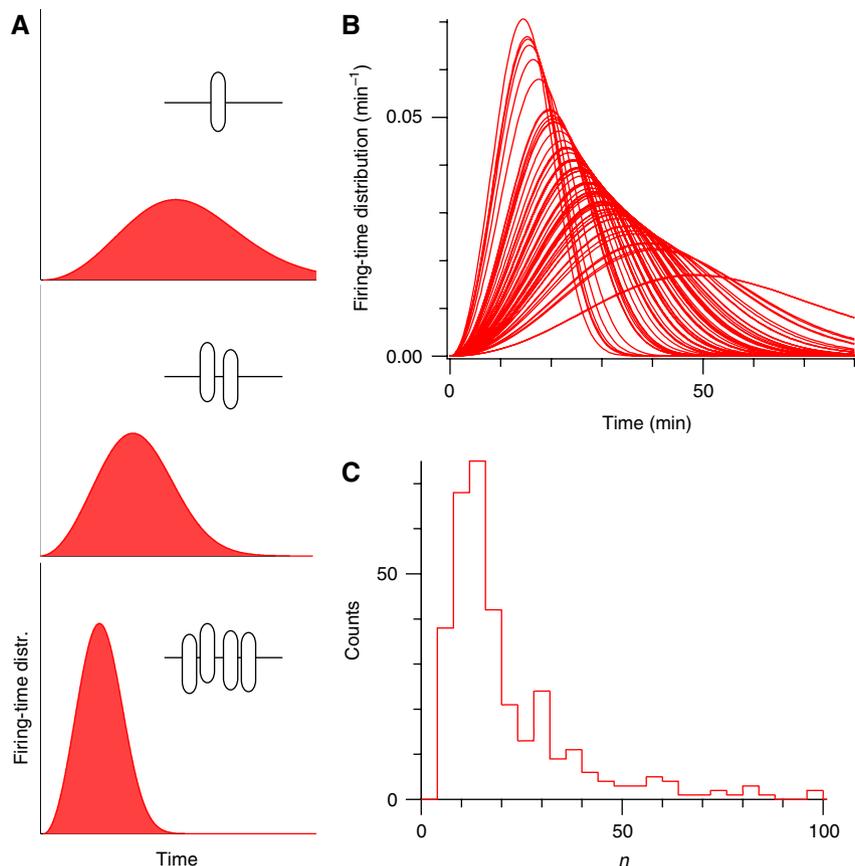


Figure 6 (A) Illustration of the multiple-initiator model. Initiators are loaded onto an origin site. The effective firing-time distribution of the origin narrows and shifts to earlier times with increasing number of initiators. (B) Firing distributions in the multiple-initiator model. The parameter n ranges from 2.2 to 96.4. These distributions resemble those shown in Figure 3A. (C) Histogram of n with bin width=4. Values of n range from 2.2 to 96.4. Median value=13.3, s.d.=16.7.

plausible, testable mechanism for the regulation of replication timing in yeast. As the elements of the model are found generally in eukaryotic organisms, one expects that it may apply to a broad range of eukaryotes.

We initially fit the data with individual sigmoids at each origin, in effect giving each origin an independent firing program. We found that we could replace these empirical firing programs with a universal firing program that describes the behavior of each origin by a single parameter n , which suggests that a fundamental underlying mechanism regulates firing at all origins. Moreover, n can intuitively be interpreted as the number of initiators loaded at an origin in this MIM. Below, we discuss the MIM and propose a plausible initiator candidate, the MCM complex. We also discuss our models in relation to the robustness of replication programs. Lastly, we discuss the limitations of the microarray data analyzed.

A candidate for initiators in the MIM: MCM complexes

A natural candidate for the initiator is the MCM complex (Sclafani and Holzen, 2007). MCM complexes are associated with the unwinding of DNA, one of the initial steps in origin activation, and are loaded in excess onto the DNA prior to S

phase (Edwards *et al*, 2002; Hyrien *et al*, 2003; Bowers *et al*, 2004). A previous experiment using ChIP-PCR showed that origin efficiency is strongly correlated with the number of MCM bind at origins (Wyrick *et al*, 2001). That data show that, on average, there are six times more MCM on efficient origins than on inefficient origins (Wyrick *et al*, 2001; their Supplementary Figure 1). A similar experiment done using ChIP-chip on tiling arrays (Xu *et al*, 2006) does not recapitulate the ChIP-PCR data (Supplementary Figure 3), possibly because of lack of sufficient dynamic range in the ChIP-chip data.

The parameter n for the number of initiators does not have to be an integer, as it represents the average number of initiators bound to an origin. Also, the absolute value of n is coupled to $\phi_o(t)$. Thus, relative variations in n between origins are more significant than absolute values. For the data of McCune *et al* (2008), we found a 43-fold range between the smallest and largest n ($n=2.2$ and 96.4, respectively), which is larger than one might expect for variation in MCM loading at a single origin. However, the majority of origins falls within a 20-fold range, between 2 and 40 (Figure 6C), which is consistent with experimentally observed levels of MCM loading (Edwards *et al*, 2002; Bowers *et al*, 2004). The higher values of n form a fairly sparse tail and may result from multiple origins located < 5 kb from each other. Such origins would not

be resolved in the microarray data and would appear as a single origin with a higher n value.

In addition, other factors such as chromatin structure and origin recognition complex occupancy can influence the loading of initiators. Previous work has hypothesized that the time required for an activator to find an origin can limit the origin-initiation rate (Gauthier and Bechhoefer, 2009). The search times in heterochromatin regions could be longer than in euchromatin regions if initiators in heterochromatin are less accessible to activators (Kerem *et al*, 1984). This in turn implies that initiation rates should be 'scaled down' in heterochromatin regions (see Equation 10 for the exact relationship between initiation rates and firing-time distribution). For the same accessibility reasons, the loading of initiators may also be reduced in the heterochromatin regions. Put together, variations in chromatin structure affect $\phi_o(t)$ and the loading of initiators simultaneously to increase the range of inferred n values. Thus, the fit parameter n represents not simply the number of initiators but rather the combined effects of origin accessibility and initiator multiplicity. As an illustration, in the simple case where the structure of chromatin identically affects initiation loading and origin activation, the number of initiators is \sqrt{n} , and the range of number of initiators would cover a 6.5-fold difference.

In the MIM, the parameter n has direct implications for the timing of the replication program. Figure 6A shows the histogram of all the $t_{1/2}$ values we extracted. The mode of the histogram suggests a typical firing time. The MIM suggests that the typical firing time is related to the average number of initiators loaded onto the origin sites. The histogram also shows that there is no $t_{1/2}$ that is earlier than 15 min. In the context of the MIM, this arises from an upper limit to the number of initiators that can bind and remain on an origin site, which corresponds to the largest n -value found (≈ 43 after normalizing the smallest n to 1). To evaluate whether such a value is reasonable, we make a crude estimate of the largest biologically plausible value of n . The highest value can be associated with a close packing of the double MCM hexamers. Each individual hexamer is 30 nm wide (0.1 kb) (Gómez-Llorente *et al*, 2005), and we imagine them spreading out from an ORC loading site while still closely packed around it. If the hexamers were to spread more than ± 2.5 kb to either side of the ORC loading site, we would, in our analysis, assign more than one origin to that region. The largest value of n for a single origin is then $\approx 5 \text{ kb}/0.1 \text{ kb}=50$, which is greater than the largest value found.

Recently, Wu and Nurse (2009) reported a correlation between the timing of initiation and the timing of origin recognition complex (ORC) binding in fission yeast that also suggests a mechanism for origin-timing control. As ORC recruits multiple MCM (Hyrien *et al*, 2003; Bowers *et al*, 2004), we speculate that early ORC binding provides more opportunities for MCM to be loaded, implying formation of early-firing origins. On the other hand, the finding of Wu and Nurse suggests that MCM loading is just one hypothesis among several, as protein complexes other than MCM may also affect the timing of origin initiation. More detailed models and quantitation of progressive-complex assembly are needed to make further progress.

Robustness of the DNA replication program

Our results show how the replication program in yeast is robust. First, we showed in the SM that although the observed efficiencies varies over a wide range, most of the origins have high potential efficiencies. This result indicates that there are more potential origins than needed to replicate the genome and that many potentially efficient origins appear to be dormant. Thus, if there are fork stalls, the normally dormant (yet highly efficient) origins would initiate to replicate the DNA, safeguarding against replication stress (Blow and Ge, 2009). In this case, the SM reveals that the replication program has built-in redundancy that makes it robust.

Second, Figure 3C shows that the initiation rate extracted using the SM increases throughout most of S phase. A similar result holds in the MIM: the global firing-time distribution $\phi_o(t)$ has to increase through most of S phase to fit the experimental data. Biologically, this design allows any large, unreplicated regions that remain toward the end of S phase to be increasingly likely to be replicated, ensuring timely completion of replication (Yang and Bechhoefer, 2008). An increase in the origin-initiation rate as S phase progresses has been observed in every data set that has been studied, and several plausible mechanisms for such an increase have been suggested (Goldar *et al*, 2009; Gauthier and Bechhoefer, 2009; Rhind *et al*, 2010). In summary, our modeling reveals two types of robustness: one safeguards against stress, whereas the other ensures replication completion.

Limitations of the microarray data

Although the microarray experiments analyzed here provide high-quality data, artifacts and limitations should be addressed. First, the microarray data analyzed has a temporal resolution of 5 min and a spatial resolution of ≈ 2 kb. Because of the limits of spatial resolution, we group origins within 5 kb into a single, more efficient effective origin (see Supplementary Material Section A). Second, the data does not cover the entire range of replication fraction (0–100%), perhaps because of contamination and imperfect signal normalization (see Supplementary Material Section A). We deal with these artifacts by introducing a scaling factor for each time point and a background as fitting parameters (see Materials and methods). Third, from the flow-cytometry data, we estimated S phase to be 60 ± 10 min. Uncertainty in the length of S phase affects the values of the efficiencies extracted but not the relationships shown in Figure 4C and D. Fourth, there is starting-time asynchrony in the cell population probed. We extended the formulation to account for such asynchrony and found it to be consistent with a normal distribution of s.d. 2 min (see Supplementary Material Section E). Refitting the SM to part of the microarray data using the estimated asynchrony, we found that ν shifted by ≈ -0.3 kb/min, $t_{1/2}$ by ≈ -0.5 min, and t_w by ≈ -1 min (Supplementary Figure 8). These changes are small and do not affect our findings presented above. Fifth, all the fit parameters have a small systematic uncertainty that originates from an incomplete knowledge of the likelihood function (see Supplementary Material Section B; Supplementary Figure 4). We found that using an alternative form of the likelihood function leads to shifts in ν by ≈ -0.2 kb/min,

origin positions by $\approx \pm 1$ kb, $t_{1/2}$ by ≈ -3 min, t_w by ≈ -2 min, and n by ≈ -1 (Supplementary Figure 5). Again, these small uncertainties do not affect our overall conclusions.

In summary, artifacts in the microarray data result in small uncertainties in the absolute values of the extracted parameters that do not alter significantly our findings. In particular, the relationship that $t_{1/2} \approx t_w$ remains valid. The $t_{1/2} - t_w$ trend clearly reveals that initiation times and the precision of timing are correlated. This relationship has also been observed by a recent analysis (de Moura *et al*, 2010) and is contrary to the view of replication in which origin initiation is timed in a nearly deterministic manner. We hypothesize that the correlation emerges from the multiplicity of stochastic initiators, and we account for it quantitatively via the MIM.

Materials and methods

Methods summary

Our analysis is described in four parts below. First, we fit sigmoids to the replication fraction of origins from the microarray data and infer that the timing and the variation in timing of origin initiation are correlated. Second, we present the SM, which is based on two elements: origin initiation and fork progression. The model takes into account the effect of passive replication because of all other origins. This framework allows one to calculate the potential efficiency of origins, which is difficult to measure because of passive replication. Third, we detail the parameters used in the SM fits. These parameters include the origin positions, the $t_{1/2}$ and t_w of the firing-time distributions, fork velocity, and factors to deal with experimental artifacts. Last, we describe the parameters used in the MIM and discuss its relationship with the firing-time distribution. Analysis codes are available upon request.

Fitting sigmoids to array data

Let the replication fraction be $f(x, t)$, where x is the genome position and t is the time elapsed since the start of S phase. For microarray data, $f(x, t)$ represents the fraction of the population that has replicated locus x by time t . Looking at the spatial part of the replication fraction, one expects that an early-firing origin that is rarely passively replicated would show a peak at the origin position, as the site is almost always replicated before its surrounding. Thus, peaks in $f(x, t)$ correspond to origin sites. At a fixed x where an origin resides, the peak height then scales with the number of initiations of the particular origin in the population. If an origin fires at a distinct time, the corresponding $f(t)$ would resemble a step function. By contrast, if an origin fires stochastically, $f(t)$ would be smooth.

Using 275 previously identified origin sites (Alvino *et al*, 2007), we extracted the corresponding 275 $f(t)$ curves from the microarray data. These $f(t)$ curves are well described by sigmoids with parameters t_{rep} and t_{width} . We used the Hill equation

$$f(t) = \frac{1}{1 + \left(\frac{t_{rep}}{t}\right)^r}, \quad (1)$$

where r is related to t_{width} via

$$t_{width} = \left(3^{1/r} - 3^{-1/r}\right)t_{rep}, \quad (2)$$

to fit the 275 $f(t)$ and extracted 275 pairs of t_{rep} and t_{width} . The result is plotted in Figure 1C.

There is a distinction between the replication fraction of a locus and the cumulative firing-time distribution of the origin at the locus. The complication is that the locus is replicated by all nearby origins and not just the origin on the site. To untangle these composite effects, we present a more general theory that accounts for passive replication.

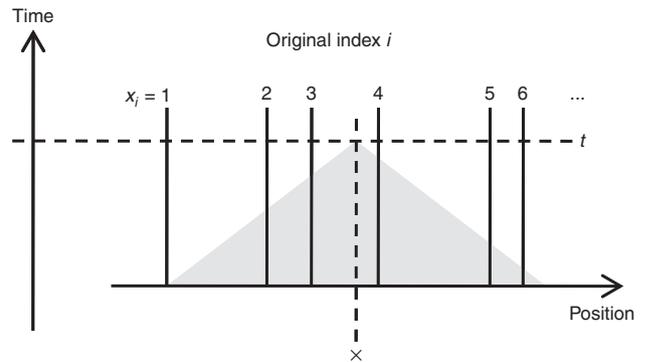


Figure 7 Illustration of Kolmogorov's argument. The replication fraction at (x, t) equals to '1-probability that no origins initiated within the shaded triangle Δ .' Vertical lines indicates the positions of discrete origins; initiation can only occur along these lines.

A general formalism describing DNA replication

The replication program is determined by the properties of origins, which are specified through their positions and ability to initiate throughout S phase. Following previous work (Jun *et al*, 2005), we describe the replication program by an initiation rate $I(x, t)$, defined as the number of times an origin at x would initiate per time per unreplicated length (meaning that the origin has not been passively replicated). Origin positions in budding yeast are well defined and are associated with autonomous replication sequences (Sclafani and Holzen, 2007); we denote these positions by x_i . An origin is then described by the initiation rate $I_i(x, t) = \delta(x - x_i)I_0(x, t)$, where $\delta(x)$ is the Dirac δ function (Boas, 2006) that sets the rate to zero for all $x \neq x_i$. The program $I(x, t)$ is the sum of all $I_i(x, t)$.

Given $I(x, t)$, we can calculate the replication fraction $f(x, t)$ of site x on the genome at a time t after the start of S phase (Kolmogorov, 1937). The result is

$$f(x, t) = 1 - \prod_{\Delta} [1 - I(x', t')\Delta x' \Delta t'], \quad (3)$$

where the product is over all area elements of $\Delta x \Delta t$ in the shaded triangle Δ depicted in Figure 7 (Kolmogorov, 1937). In words, the fraction of replication at a specific position and time is one minus the probability that the position has not been replicated before that time. In the limit $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$, one finds

$$f(x, t) = 1 - e^{-\int_{\Delta} I(x', t') dx' dt'}. \quad (4)$$

To make the double integral explicit, we define a local measure of origin firing,

$$g(\Delta x_p, t) = \int_{x_p}^{x_{p+1}} \delta(x - x_i) dx \int_0^t I_0(x, t') dt' \quad (5)$$

for the interval (x_p, x_{p+1}) , where the index p comes from discretizing a genome of length L :

$$\Delta x = \frac{L}{N}, \quad x_p = p(\Delta x), \quad p = 0, 1, \dots, N-1. \quad (6)$$

The function $g(\Delta x_p, t) = 0$ unless there is an origin contained in the interval (x_p, x_{p+1}) . Replacing the double integral in Equation (4) by a sum using Equation (5), we obtain

$$f(x, t) = 1 - \exp\left[-\sum_{p=0}^{N-1} g\left(\Delta x_p, t - \frac{|x - x_p|}{v}\right)\right], \quad (7)$$

where v is the fork velocity. The Δx_p in the first argument of $g(x, t)$ is an interval, whereas the x_p in the second argument is a point. The second argument, $t - |x - x_p|/v$, is the time along the edge of the triangle. This

equation is used to fit the microarray time course data. Biologically relevant $g(x, t)$ should satisfy the following constraints:

1. $g(\Delta x_p, t < 0) = 0$. This means that the initiation rate is zero [$I(x, t < 0) = 0$] before the start of replication. Applying this constraint to Equation (7), we see that the sum is limited to the domain $(x - t/v) \leq x \leq (x + t/v)$.
2. $\frac{d}{dt}g(\Delta x_p, t) \geq 0$. As $g(\Delta x_p, t) = \int_x^{x+\Delta x} \int_0^t I(x', t') dt' dx'$, this is equivalent to $I(x, t) \geq 0$, meaning that the initiation rate cannot be negative.
3. $g(\Delta x_p, t) \geq 0$. This is a direct consequence of constraints 1 and 2.

We derive a useful quantity, the cumulative initiation probability $\Phi(x_p, t)$, from $g(\Delta x_p, t)$ following a Poisson-process calculation (Rieke et al, 1999):

$$\Phi(x_p, t) = 1 - e^{-g(\Delta x_p, t)}. \quad (8)$$

As $\Phi(x_p, t)$ is only non-zero for $x_p \leq x_i \leq x_{p+1}$, we introduce

$$\phi_i(t) \equiv \frac{d}{dt}\Phi(x_i \in \Delta x_p, t) \equiv \frac{d}{dt}\Phi_i(t) \quad (9)$$

to denote the initiation probability density of each origin i . The x_p in Equation (8) is related to the discretization, whereas the subscript i in Equation (9) relates to the origin position. Rearranging Equation (8) for $g(\Delta x_p, t)$ and differentiating, we obtain the relationship between the initiation rate and the firing-time distribution:

$$I_i(t) \equiv I(x_i \in \Delta x_p, t) = \frac{\phi_i(t)}{1 - \Phi_i(t)}. \quad (10)$$

We note that Equation (10) is also the firing probability density of an origin given that it has not initiated before time t .

Using the probability distributions in Equation (9), we define potential efficiency as the probability that an origin would fire by the end of S phase if there is no passive replication, that is, $\Phi_i(t=t_{\text{end}})$ with t_{end} being the time that S phase ends. By contrast, the efficiency of any origin decreases because of passive replication by neighboring origins. The observed efficiency of the j th origin thus depends not only on $\phi_j(t)$ but also with the $\phi_k(t)$ of neighboring origins. Mathematically, the observed efficiency E_j of the j th origin is then

$$E_j \equiv \int_0^{t_{\text{end}}} dt \phi_j(t) \prod_{k \neq j} \left[1 - \Phi_k \left(t - \frac{|x_k - x_j|}{v} \right) \right], \quad (11)$$

where the integrand is an effective initiation probability density that represents the probability that the j th origin initiates between t and $t + dt$ times the probability that all the other origins have not initiated to passively replicate the j th origin before time t .

A group of origins is effectively a more-efficient origin. In light of this, consider a tightly packed cluster of initiators in the MIM. As it is always the earliest initiation in the cluster that counts, the effective initiation probability of the cluster is the extreme (smallest) value distribution of all the $\phi_j(t)$ s in the cluster. For n initiators near x , the effective cumulative initiation probability is

$$\Phi_{\text{eff}}(x, t, n) = 1 - \prod_{j=1}^n \left[1 - \Phi_j \left(t - \frac{|x - x_j|}{v} \right) \right]. \quad (12)$$

For large n , Equation (12) tends toward the Weibull distribution for a large class of functions Φ_j , a feature of extreme-value statistics (Kotz and Nadarajah, 2000). This is a desirable feature for the MIM (see Results).

Fitting parameters for the SM

Using the model presented above, we performed least-squares fits to the time course microarray data in McCune et al (2008). The fit is done using the Global Fit package in Igor Pro 6.1 (Wavemetrics Inc.; <http://www.wavemetrics.com>). We fit to the unsmoothed microarray data (Supplementary Material Section F; Supplementary Figure 9). To speed up the code, the fit program calls external C-language code for key function evaluations. On a personal computer with an Intel Core2 Duo CPU @ 3.16 GHz, the fit for an average chromosome (≈ 900 points and

≈ 150 parameters) takes about 5 min. The time to fit the entire genome of *S. cerevisiae* is about 10 h. All local parameters of the SM are tabulated in Supplementary Table I. Global ones are in Supplementary Table III. There are three types of genome-wide parameters:

Fork velocity v

Even though the model allows for variable-fork velocity, we found that to fit the microarray data, a constant v across the genome and throughout S phase suffices (see Supplementary Material Section C).

Background b_g

The reason for introducing this parameter comes from the observation that the 10-min time point data do not start close to $f=0$ (Supplementary Material Section A). Although introducing a variable background is possible, we found that a simple constant background is sufficient for the global fit. This parameter is also global in space and time.

Normalization factors α_t

These parameters correct for various artifacts that cause the microarray data to not cover the entire range of possible fractions, 0–100% (see Supplementary Material Section A). We propose a normalization factor α_t for each time point that is genome wide. Combining the normalization factors with the background parameter, we generate a modified replication fraction

$$f_{\text{mod}}(x, t) = \alpha_t [f(x, t) + b_g] \quad (13)$$

to fit the data. Again, in Equation (13), α_t is global in space but not time, whereas b_g is global both in time and space.

There are two types of local parameters pertaining to origins. These parameters are local in space but global in time:

Origin position

Each origin was associated with a unique location x_i along the genome. Initial guesses of the x_i include all 732 origins recorded in the OriDB database (Nieduszynski et al, 2007). We counted origins that were < 5 kb apart as a single origin throughout the fitting process, as the data (resolution ≈ 2 kb) cannot distinguish between a single-efficient origin and a group of less-efficient origins (see Supplementary Material Section A). Before the genome-wide fit, we first fit each chromosome separately to eliminate false origins and origins that do not contribute enough to the replication fraction. This allowed the genome-wide fit to run more smoothly. The criteria for elimination were (1) mode of firing-time distribution < 10 min and efficiency < 0.5 and (2) cumulative firing-time probability < 0.3 at end of S phase (estimated to be 60 min). The first criterion aims to identify peaks in microarray data that originate from contamination instead of origin activity (small fragments of unreplicated DNA along with A-T rich sequence can be counted as replicated DNA, as mentioned in the Supplementary Material Section A). Contamination produces microarray peaks in the early but not the mid and late time points. In the model, this is effectively the same as an origin that fires very early (before the first time point at 10 min) but is inefficient. The second criterion simply eliminates origins that do not contribute much throughout S phase.

Firing-time-distribution parameters

We used the Hill equation (same form as Equation 1) to describe the cumulative firing-time distribution $\Phi(t)$. The two parameters are the median time $t_{1/2}$ at which $\Phi(t=t_{1/2})=0.5$ and the width t_w , defined as the difference $t_{0.75}-t_{0.25}$. This function is a valid cumulative initiation probability function, satisfying the constraints $0 \leq \Phi(t) \leq 1$, $\Phi(t < 0) = 0$, and $d\Phi(t)/dt \geq 0$.

Fitting parameters for the MIM

The MIM, like SM, is also parameterized with the fork velocity, background, normalization factors, and origin positions. The two

firing-time-distribution parameters for each origin in SM are replaced by a single parameter n in MIM. As discussed in the Discussion, n represents the number of initiator molecules that is modified by effects such as chromatin structure and ORC loading. There are additional parameters ($t_{1/2}^*$ and r^*) to describe the global firing-time distribution of all initiators (see Equation 15 below). These are absent from the SM, as the SM does not assume that the firing distributions are related. All local parameters of the MIM are tabulated in Supplementary Table II and all global parameters in Supplementary Table III.

We still use the Hill equation to describe the global cumulative distribution $\Phi_o(t)$. Using Equation 12, the $\Phi_{\text{eff}}(t, n)$ for a cluster of n initiators is

$$\Phi_{\text{eff}}(t, n) = 1 - [1 - \Phi_o(t)]^n, \quad (14)$$

where

$$\Phi_o(t) = \frac{1}{1 + \left(\frac{t_{1/2}^*}{t}\right)^{r^*}}. \quad (15)$$

The quantity $t_{1/2}^*$ is the median time of the firing-time distribution for a single initiator, and r^* is its rate of increase.

Because of the different form of firing-time distribution used in the MIM, the criteria for origin elimination are slightly modified from the SM ones. The MIM predicts strictly that earlier-firing origins are more efficient; thus, no contamination effects are modeled by the MIM. We then eliminated origins via a single criterion of $\Phi_i(t=60 \text{ min}) < 0.4$. The change is consistent with the observation that the firing-time distribution in Equation (14) tends to result in more efficient origins than the Hill function does. Similar to the SM fits, individual-chromosome fits were done first before performing the genome-wide fit.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (<http://www.nature.com/msb>).

Acknowledgements

We thank MK Raghuraman and Gina M Alvino for making available the raw data from McCune et al (2008). We thank Olivier Hyrien for comments and suggestions concerning the upper bound for allowed initiator numbers. We thank Ollie Rando and Melissa Moore for useful comments on the manuscript. We thank Michel Gauthier for discussions and helping with coding for fits. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), the American Cancer Society (ACS), and the Human Frontier Science Program (HFSP).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alvino GM, Collingwood D, Murphy JM, Delrow J, Brewer BJ, Raghuraman MK (2007) Replication in hydroxyurea: it's a matter of time. *Mol Cell Biol* **27**: 6396–6404

Blow JJ, Ge XQ (2009) A model for DNA replication showing how dormant origins safeguard against replication fork failure. *EMBO Rep* **10**: 406–412

Boas ML (2006) *Mathematical Methods in the Physical Sciences*, 3rd edn, New York, NY, USA: John Wiley & Sons

Bowers J, Randell J, Chen S, Bell S (2004) ATP hydrolysis by ORC catalyzes reiterative Mcm2–7 assembly at a defined origin of replication. *Mol Cell* **22**: 967–978

Czajkowsky DM, Liu J, Hamlin JL, Shao Z (2008) DNA combing reveals intrinsic temporal disorder in the replication of yeast chromosome VI. *Mol Biol Cell* **375**: 12–19

Donaldson AD (2005) Shaping time: chromatin structure and the DNA replication programme. *Trends Genet* **21**: 444–449

Edwards MC, Tutter AV, Cvetic C, Gilbert CH, Prokhorova TA, Walter JC (2002) MCM2–7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts. *J Biol Chem* **277**: 33049–33057

Eshaghi M, Karuturi RKM, Li J, Chu Z, Liu ET, Liu J (2007) Global profiling of DNA replication timing and efficiency reveals that efficient replication/firing occurs late during S-phase in *S. pombe*. *PLoS ONE* **2**: e722

Feng W, Max E, Boeck DC, Fox LA, Alvino GM, Fangman WL, Raghuraman MK, Brewer BJ (2006) Genomic mapping of single-stranded DNA in hydroxyurea-challenged yeasts identifies origins of replication. *Nat Cell Biol* **8**: 148–155

Ferguson BM, Brewer BJ, Reynolds AE, Fangman WL (1991) A yeast origin of replication is activated late in S phase. *Cell* **65**: 507–515

Gómez-Llorente Y, Fletcher RJ, Chen XS, Carazo JM, Martín CS (2005) Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from *Methanobacterium thermoautotrophicum*. *J Biol Chem* **280**: 40909–40915

Gauthier MG, Bechhoefer J (2009) Control of DNA replication by anomalous reaction-diffusion kinetics. *Phys Rev Lett* **102**: 158104

Goldar A, Marsolier-Kergoat MC, Hyrien O (2009) Universal temporal profile of replication origin activation in Eukaryotes. *PLoS ONE* **4**: e5899

Goren A, Cedar H (2003) Replicating by the clock. *Nat Rev Mol Cell Biol* **4**: 25–32

Heichinger C, Penkett CJ, Bähler J, Nurse P (2006) Genome-wide characterization of fission yeast DNA replication origins. *EMBO J* **25**: 5171–5179

Herrick J, Bensimon A (1999) Single molecule analysis of DNA replication. *Biochimie* **81**: 859–871

Herrick J, Stanislawski P, Hyrien O, Bensimon A (2000) Replication fork density increases during DNA synthesis in *X. laevis* egg extracts. *J Mol Biol* **300**: 1133–1142

Hyrien O, Goldar A (2010) Mathematical modeling of eukaryotic DNA replication. *Chromosome Res* **18**: 147–161

Hyrien O, Marheineke K, Goldar A (2003) Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays* **25**: 116–125

Jun S, Zhang H, Bechhoefer J (2005) Nucleation and growth in one dimension. I. The generalized Kolmogorov-Johnson-Mehl-Avrami model. *Phys Rev E* **71**: 011908

Kerem BS, Goitein R, Diamond G, Cedar H, Marcus M (1984) Mapping of DNAase I sensitive regions on mitotic chromosomes. *Cell* **38**: 493–499

Kolmogorov AN (1937) On the static theory of crystallization in metals. *Bull Acad Sci USSR, Phys Ser* **1**: 335

Kotz S, Nadarajah S (2000) *Extreme Value Distributions: Theory and Applications*. River Edge, NJ, USA: Imperial College Press

Lygeros J, Koutroumpas K, Dimopoulos S, Legouras I, Kouretas P, Heichinger C, Nurse P, Lygerou Z (2008) Stochastic hybrid modeling of DNA replication across a complete genome. *Proc Natl Acad Sci USA* **105**: 12295–12300

McCune HJ, Danielson LS, Alvino GM, Collingwood D, Delrow JJ, Fangman WL, Brewer BJ, Raghuraman MK (2008) The temporal program of chromosome replication: genomewide replication in *clb5Δ Saccharomyces cerevisiae*. *Genetics* **180**: 1833–1847

de Moura APS, Retkute R, Hawkins M, Nieduszynski CA (2010) Mathematical modelling of whole chromosome replication. *Nucleic Acids Res* (advance online publication 10 May 2010; doi:10.1093/nar/gkq343)

Nieduszynski CA, Hiraga S, Ak P, Benham CJ, Donaldson AD (2007) OriDB: a DNA replication origin database. *Nucleic Acids Res* **35**: D40–D46 (<http://www.oridb.org>)

Patel PK, Arcangioli B, Baker SP, Bensimon A, Rhind N (2006) DNA replication origins fire stochastically in fission yeast. *Mol Biol Cell* **17**: 308–316

- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. Cambridge Cambridge, NY, USA: University Press
- Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL (2001) Replication dynamics of the yeast genome. *Science* **294**: 115–121
- Rhind N, Yang SCH, Bechhoefer J (2010) Reconciling stochastic origin firing with defined replication timing. *Chromosome Res* **18**: 35–43
- Rieke F, Warland D, Steveninck R, Bialek W (1999) *Spikes: Exploring the Neural Code*. London, UK: The MIT Press
- Sclafani RA, Holzen TM (2007) Cell cycle regulation of DNA replication. *Annu Rev Genet* **41**: 237–280
- Sekedat MD, Fenyő D, Rogers RS, Tackett AJ, Aitchison JD, Chait BT (2010) GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Mol Sys Biol* **6**: 353
- Wu PYJ, Nurse P (2009) Establishing the program of origin firing during S phase in fission yeast. *Cell* **136**: 852–864
- Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* **294**: 2357–23650
- Xu W, Aparicio JG, Aparicio OM, Tavaré S (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics* **7**: 276
- Yabuki N, Terashima H, Kitada K (2002) Mapping of early firing origins on a replication profile of budding yeast. *Genes Cells* **7**: 781–789
- Yang SCH, Bechhoefer J (2008) How *Xenopus laevis* embryos replicate reliably: investigating the random-completion problem. *Phys Rev E* **78**: 041917
- Yang SCH, Gauthier MG, Bechhoefer J (2009) Computational methods to study kinetics of DNA replication. In *DNA Replication: Methods and Protocols*, Vengrova S, Dalgaard JZ (eds), pp 555–574. Humana Press: New York c/o Springer Science + Business Media
- Zhang H, Bechhoefer J (2006) Reconstructing DNA replication kinetics from small DNA fragments. *Phys Rev E* **73**: 051903



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.