DEVELOPMENTAL BIOLOGY
Correction for “Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development,” by Zhe Ji, Ju Youn Lee, Zhenhua Pan, Bingjun Jiang, and Bin Tian, which appeared in issue 17, April 28, 2009, of Proc Natl Acad Sci USA (106:7028–7033; first published April 16, 2009; 10.1073/pnas.0900028106).

The authors note that due to a printer’s error, Fig. 1B appeared incorrectly on page 7028. The corrected figure and its legend appear below.

[Image of corrected figure]

CELL BIOLOGY, COMPUTER SCIENCE

The authors wish to note the following: “We wish to add direct references to a stochastic model of DNA replication previously applied to the *Xenopus laevis* early embryonic divisions. That model was applied to molecular combing experiments on cell free extracts from *Xenopus laevis* embryos.” The additional references appear below.


MICROBIOLOGY

The authors note that in Table 1 on page 11360, the units for “Electrochemical activity,” designated as mA, should have been designated as $\mu$A. This error does not affect the conclusions of the article. The corrected table appears below.

**Table 1. Metal-reduction and electrochemical properties of *S. oneidensis* MR-1, GSPD, and ΔMTRC/OMCA**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Ferricyanide-reduction-extractable Fe(II), $^*$ mM</th>
<th>Electrochemical activity, $\mu$A</th>
</tr>
</thead>
<tbody>
<tr>
<td>MR-1</td>
<td>2.17 ± 0.19</td>
<td>68.0 ± 7.8</td>
</tr>
<tr>
<td>ΔMTRC/OMCA</td>
<td>0.67 ± 0.18</td>
<td>14.3 ± 2.08</td>
</tr>
<tr>
<td>GSPD</td>
<td>0.42 ± 0.02</td>
<td>12.3 ± 0.58</td>
</tr>
</tbody>
</table>

*$^{*}$HFO (20 mM) reduction sampled at 24 h and extracted with 0.5 N HCl overnight. Fe(II) was determined by using the ferrozine assay (30).

www.pnas.org/cgi/doi/10.1073/pnas.0903426106

www.pnas.org/cgi/doi/10.1073/pnas.0905246106
Stochastic hybrid modeling of DNA replication across a complete genome

J. Lygeros*, K. Koutroumpas*, S. Dimopoulos‡, I. Legouras‡, P. Kouretas*, C. Heichinger§, P. Nurse§, and Z. Lygerou†

*Automatic Control Laboratory, ETH 8092 Zurich, Switzerland; ‡Laboratory of General Biology, School of Medicine, University of Patras, 26504 Rio, Patras, Greece; §Laboratory of Yeast Genetics and Cell Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10065; and †Cancer Research United Kingdom, London Laboratories, London WC2A 3PX, England

Communicated by Fotis C. Kafatos, Imperial College London, London, United Kingdom, June 6, 2008 (received for review December 6, 2007)

DNA replication in eukaryotic cells initiates from hundreds of origins along their genomes, leading to complete duplication of genetic information before cell division. The large number of potential origins, coupled with system uncertainty, dictates the need for new analytical tools to capture spatial and temporal patterns of DNA replication genome-wide. We have developed a stochastic hybrid model that reproduces DNA replication throughout a complete genome. The model can capture different modes of DNA replication and is applicable to various organisms. Using genome-wide data on the location and firing efficiencies of origins in the fission yeast, we show how the DNA replication process evolves during S-phase in the presence of stochastic origin firing. Simulations reveal small regions of the genome that extend S-phase to three times its reported duration. The low levels of late replication predicted by the model are below the detection limit of techniques used to measure S-phase length. Parameter sensitivity analysis shows that increased replication fork speeds genome-wide, or additional origins are not sufficient to reduce S-phase to its reported length. We model the redistribution of a limiting initiation factor during S-phase and show that it could shorten S-phase to the reported duration. Alternatively, S-phase may be extended, and what has traditionally been defined as G2 may be occupied by low levels of DNA synthesis with the onset of mitosis delayed by activation of the G2/M checkpoint.

DNA replication, the process of duplication of the cell’s genetic material, must be carried out before every cell division. In bacteria and viruses, replication initiates from a single genomic site in every cell, called the origin of replication (1). Eukaryotic DNA replication is characterized by a higher degree of uncertainty; hundreds to thousands of potential origins exist along the genome that fire with varying efficiencies and at different times during S-phase (2–5). Although the spatial and temporal pattern of DNA replication is not strictly defined in eukaryotic cells, each cell in a population must complete the replication process in an accurate and timely manner. Failure to replicate even a small part of the genome would disrupt proper segregation of the genetic material to the two daughter cells during mitosis, leading to genomic instability.

Initial models of eukaryotic DNA replication, influenced by the replication of bacterial genomes, postulated that defined regions in the genome would act as origins of replication in every cell cycle (6). Indeed, work in the budding yeast permitted the identification of specific sequences that could act as origins with high efficiency (7, 8). However, a simple deterministic model of origin selection could not accommodate the complexity of the replication process observed in Metazoan. Different organisms and cell types exhibit variations on the DNA replication process but a common feature is a degree of uncertainty in origin selection. In Drosophila or Xenopus preblastula embryos, DNA replication initiation occurs at short intervals (8–15 kb) without apparent sequence specificity (9, 10). Similarly, DNA replication can initiate from any DNA fragment introduced on a plasmid in cultured mammalian cells and at random sites within each fragment (11). In the genomic context, specific regions have origin activity (12, 13), but these lack clear sequence characteristics and fire in only a fraction of the cells (2). As a result, initiation events differ between cells in a population. Stochastic selection of active origins from a large pool of putative origins could lead to the random generation of large interorigin gaps, which may take too long to replicate. This complication of stochastic DNA replication initiation is known as the random gap problem (14). Different solutions to this problem have been suggested, including a spacing mechanism for the generation of defined interorigin gaps (15, 16), coordination of fork velocity with interorigin distance (17), and an increase in origin efficiency during S-phase (14, 18).

Schizosaccharomyces pombe (fission yeast) is an attractive model system for studying genome-wide DNA replication (19). In this organism, similar to Metazoan cells, replication originates from specific regions, and no origin defining consensus sequences have been characterized (20). Following publication of the S. pombe genome (21), bioinformatics (22) and high-throughput analyses (23–25) made it possible to map origins of replication along the complete genome. Dai et al. (26) provided evidence that approximately half of the 5,000 intergenic regions present in the S. pombe genome may exhibit origin activity and suggested that the origins active in any given S-phase are recruited stochastically among these intergenic regions. Single-molecule analysis has provided evidence for a stochastic pattern of initiation events in fission yeast; DNA combing experiments showed that interorigin distances have an exponential distribution, indicative of stochastic firing (27). Heichinger et al. (24) used comparative genomic hybridization to map the sites of replication origins genome-wide and to measure the firing efficiency of each origin. They showed that ~400 origins have a firing efficiency of 10–60%, while ~500 more putative origins can be detected, which exhibit efficiencies below 10%. This analysis is in good agreement with microarray analysis based on chromatin immunoprecipitation (25), with single molecule analysis by DNA combing (27), and with 2D gel analysis of replication intermediates (22), and suggests that ~160 origins out of a total of 900 would be active in any S-phase.

Here, we employ new analytical tools developed in the area of stochastic hybrid control (28) to create a versatile model able to capture different modes of DNA replication in different organ-

Author contributions: J.L., P.N., and Z.L. designed research; J.L., K.K., S.D., I.L., and P.K. performed research; K.K. and C.H. analyzed data; and J.L., P.N., and Z.L. wrote the paper.

The authors declare no conflict of interest.

1To whom correspondence may be addressed: ETH Zurich, Automatic Control Laboratory, ETI I 22, Physikstrasse 3, 8092 Zurich, Switzerland. E-mail: lygeros@control.ee.ethz.ch.

2To whom correspondence may be addressed. E-mail: lygerou@med.upatras.gr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805549105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA
isms. Model instantiation using full genome data from fission yeast allows us to simulate DNA replication across the entire genome. Strikingly, in silico analysis reveals that a small part of the genome (~5%) may remain unreplicated for a long time, extending into the cell cycle period traditionally defined as G2-phase. The model’s flexibility allows us to explore possible solutions to the problems brought about by random origin selection.

Results

A Stochastic Hybrid Model of DNA Replication. Stochastic hybrid systems are used in areas such as mathematical finance, transportation, and telecommunications to model systems that involve the interaction of discrete phenomena, continuous phenomena, and uncertainty (28). DNA replication is a stochastic hybrid process, where discrete transitions between origin states are coupled with the continuous movement of replication forks along the genome, while initiation events are characterized by uncertainty in time and space. We have built a stochastic hybrid model that reproduces the DNA replication process genome-wide. The basic features of the model are presented here; a full description including formal definitions is given in supporting information (SI) Text, see also (28) and (29) for the mathematical foundations on which the model is based.

The discrete dynamics of the model capture instantaneous changes in the state of the system, such as origin firing or fork conversion. The state of each origin $i$ is represented by a variable, $S_i$, that takes one of six values (PreR, PassR, RB, RL, RR, PostR) (Fig. 1). All origins start at the Pre-Replicative state (PreR) and end either in the passively replicated (PassR) or the Post-Replicating (PostR) states. The remaining states discriminate origins from which active forks emanate on both directions (RB), only to the left (LR) or only to the right (RL). Transitions between states are governed by guards (G), logical statements that depend on the continuous and stochastic dynamics of the system and determine when the transition will take place. The continuous dynamics of the model capture the evolution of replication forks along the genome and are represented by two differential equations for each active origin, denoting the number of bases replicated by forks moving away from origin $i$ toward the left ($L_i$) or right ($R_i$). Notice that there is a tight coupling between the discrete and continuous states of the process, a defining feature of hybrid systems. Origin firing is characterized by a degree of uncertainty in both the location and time of firing of different origins, captured by the stochastic dynamics of the model. For example, the time $T_i$ at which each origin $i$ will fire in any given cell in a population can be randomly extracted based on its experimentally defined firing probability. Note that each origin is characterized by an intrinsic propensity to fire (referred to hereafter as firing propensity), which denotes the instantaneous probability of firing in the absence of passive replication. The firing propensity differs from the observed firing efficiency (referred to hereafter as observed efficiency), which also depends on the probability of passive replication for this particular origin.

Model Flexibility. The model developed is generally applicable and can, in principle, capture the DNA replication process of any organism for which knowledge of the process of origin selection and activation is available. The model requires as input:

1. The length of the genome in base pairs ($L_i$).
2. The number ($N$) of putative origins of replication and the location $X_i$ of each putative origin $i$ along the genome. The number and location of putative origins can be either deterministic based on experimental data, or extracted randomly based on information such as the average number or average spacing of origins.
3. The firing propensity $\lambda_i$ of each origin $i$. This is related to the probability of the origin firing in a unit of time in the absence of passive replication. The model can accommodate different modes of origin firing in time and space (such as stochasticity in space versus a spacing mechanism, stochasticity in time versus a timing mechanism, or constant firing propensity during S-phase versus increasing or decreasing firing propensity as a function of time). We will provide an example of this below.
4. The speed $v(x)$ with which replication forks progress when they go over position $x$ of the genome. Speed can be determinedistically defined, or depend on system parameters, such as interorigin spacing.

Model Instantiation. We have used experimental data from the fission yeast Schizosaccharomyces pombe for model instantiation. In this organism, there is experimental support for a stochastic mode of origin activation provided at the single cell level (27). High-throughput analyses have permitted the location of the majority of putative origins to be specified along the genome (23–25). Additionally, the firing propensity of each origin along the genome has been determined based on the experimentally defined observed efficiencies in the presence of the drug hydroxyurea, which inhibits fork progression, thereby eliminating passive replication (24). In fission yeast, only a small number of origins genome-wide appear intrinsically late and do not fire during a long hydroxyurea block (24, 27), while for the vast majority of origins, the timing of firing appears correlated to their firing propensities (24), suggesting that stochastic determination of firing times is the prevailing mechanism in the fission yeast (18). Fission yeast offers therefore a good opportunity to capture full genome replication based on experimental data coupled to an understanding of the underlining biology.

The number of origins ($N = 893$), their exact locations ($X_{1, 2, \ldots, N}$) along the three chromosomes of $S. pombe$ (total length $L = 12,039,987$ bases), and the firing propensity of each origin ($FP_i$, for $i=1, 2, \ldots, N$, defined as the fraction of cells in which each origin was observed to fire in hydroxyurea) were used as input (24), see Table S1. A constant speed of $v(x) = 3000$ bases/minute was used genome-wide, based on experimental estimates (24). The firing time, $T_i$ (in minutes), of each origin $i$ was extracted according to an exponential distribution whose rate $\lambda_i$ depends on $FP_i$ (SI Text). Initially, the firing times of different origins in the absence of passive replication were assumed to be statistically independent. This is a valid assumption for fission yeast, based on available experimental data (24). In addition, the firing propensity of each origin as a function of time ($\lambda_i$) was assumed to be constant during S-phase. This means that the total remaining firing propensity of the system decreases as S-phase progresses. An alternative model that relaxes both...
results closely matched the experimentally derived curves (exponential curves generated by \(27\)) fitting their experimental estimation of inter-bubble distances (inter-bubble distance) calculated. Histograms of the fraction of inter-bubble distances falling within a given range (in bins of 10 kb) are shown for: (A) Hydroxyurea, (B) Early S-phase, and (C) Late S-phase. For (C), only replication bubbles <30 kb were included, following \(27\). The curves best-fitting the experimental data of \(27\) have been overlaid for comparison.

Simulation Results. The model was coded in the Matlab environment. Monte-Carlo simulations were used to capture the process of DNA replication at a single cell, extract useful statistics on model predictions at a population level, and validate them against experimental data. Examples of simulation runs are given in Fig. S1 and show the stochastic nature of the process. Diagnostic tests were carried out which showed that observed fluctuations in continuous and discrete system parameters during the evolution of DNA replication behave as predicted and each part of the genome is copied once and only once (see SI Text and Fig. S2).

To validate the model using an independent source of experimental data, the output of simulations was compared to data derived by the analysis of replication intermediates at the single molecule level by DNA combing \(27\). In this method, replication intermediates are labeled \textit{in vivo}, genomic DNA is extracted, DNA fibers stretched and visualized. To reproduce \textit{in silico} the experimental conditions of \(27\), random genomic pieces with a size distribution matching the one of the fibers analyzed in \(27\) were extracted from 2,000 simulations under three circumstances: assuming that the first 160 origins (24) scheduled to fire in each simulation would arrest after 5 kb [mimicking the hydroxyurea block of \(27\)], when 30% of the genome had been replicated in each simulation (early S-phase) and when 70% of the genome had been replicated in each simulation (late S-phase). In Fig. 2, the distribution of distances between replication bubbles (inter-bubble distances) is shown for each case. The exponential curves generated by \(27\) fitting their experimental data have been overlaid for comparison. In all cases, simulated results closely matched the experimentally derived curves (\(R^2\) values of 0.9937, 0.9794, and 0.9749, respectively).

Model output was used to analyze genome-wide replication kinetics \textit{in silico}. In Fig. 3A, the fraction of unreplicated DNA as a function of time is shown, with each curve representing one simulation (single cell). While in all simulations the process of conversion from fully unreplicated \((1)\) to fully replicated \((0)\) follows similar kinetics, differences in individual cells due to stochastic phenomena are evident. Strikingly, while the bulk of DNA replication takes place quickly during the first half of the process, a very long tail is evident in the second half, showing that completion of DNA replication of a remaining small fraction of the genome takes a long time. This is also evident in the rate of new DNA synthesis (Fig. 3B), which shows the expected dumb-bell shape observed experimentally, but continues with a long tail. As a consequence, the mean completion time of the process (Fig. 3C) is unexpectedly long \([67 \text{ min, as opposed to experimental estimates of } \approx 20 \text{ min}\]) (30)). This is not due to less origins firing in each simulation than expected (Fig. 3D, average 159 origins firing in each simulation, very close to the experimental estimate of \(\approx 160\)). In each simulation, at least one region of the genome exhibits a distance between adjacent active origins that is very long (Fig. 3E, mean maximum interorigin distance of 258 kb, replication of which would last 43 min with a fork speed of 3,000 bases/minute, assuming both origins fire at the same time). Different locations along each chromosome delay the process in different simulations (Fig. 3F, showing regions that terminate replication after 99% of the genome has replicated) and even the most problematic regions do not delay the process in >12% of the simulations. This is explained by the random distribution of active origins along the genome (random gap problem). The model predicts that there is a small percentage (<5%) of DNA that remains unreplicated for a long time, that the rate of DNA synthesis at late time points is low and that there is an infrequent appearance within the population of each problematic region. Such a delay in S-phase completion is difficult to detect experimentally and could well have gone unnoticed.
we decreased the parameter small values of 67 to 21 min. As expected, however, the predicted average increase, the predicted average S-phase duration decreases, from the parameters: the intrinsic firing propensity of each origin; the fork completion times revealed by full genome simulations, we proceeded further increases in speed have smaller effects on the overall expected 20 min, speeds of >18 kb/min would be required, which are unrealistically high.

Finally, the number of potential origins, N, was increased by introducing more low efficiency origins at specific locations along the genome (Fig. 4 G–I). The experimental method used for origin identification (24) could not reliably detect origins with <10% efficiency. New origins with efficiencies below this experimental threshold were introduced into biologically relevant chromosomal locations, predicted by bioinformatics analysis to fulfill characteristics of known origin regions (SI Text). Criteria depending on AT content and size were used to gradually introduce additional origins with observed efficiencies of FP = 8%. The addition of new origins causes a decrease in the predicted average S-phase duration, however even after adding all potential origins predicted by our analysis (~50% of all intergenes in the fission yeast genome), S-phase completion times still remained long (46 min).

Firing Propensity Redistribution Model. Our analysis leads to two alternatives: Either S-phase duration in the presence of stochastic firing is longer than experimentally defined, or a further mechanism exists that succeeds in limiting S-phase duration circumventing the random gap problem. One proposed solution to the random gap problem is that the efficiency of unreplicated origins may increase during the course of S-phase (14, 18). Stochasticity could be brought about by a limiting factor that allows firing of a fraction of putative origins. The redistribution of this factor after an origin fires or gets passively replicated would imply a gradual increase in firing propensity of origins still in the prereplicative state as S-phase progresses (18). This biological scenario can be captured in our model by keeping the total system firing propensity (given by the sum of the firing propensities of all origins) constant throughout S-phase. Each origin that fires or becomes passively replicated releases its firing propensity λ, which is redistributed to the remaining preR origins proportionally to their initial firing propensities. We call the resulting variant of the model the “Firing Propensity Redistribution Model” and we describe it in detail in the SI Text.

Fig. 5A shows the evolution of unreplicated DNA over time predicted by the firing propensity redistribution model. While the kinetics of bulk DNA replication are similar to the basic model, the curve has lost its long tail. A similar change is observed in the curve depicting the rate of newly synthesized DNA over time (Fig. 5B). Consequently, the average time of completion of DNA replication has decreased from 67 min in the basic model to 33 min (Fig. 5C), concomitant with a decrease in the mean maximum distance observed between active origins in each simulation (from 258 kb to 135 kb, Fig. 5E). The number of firing origins has increased to a mean of 286 (Fig. 5D). The regions that replicate last in the process are shown in Fig. 5F. Two additional variants of the redistribution model (redistribution upon fork conversion proportional to initial firing propensity and redistribution upon firing proportional to current firing propensity) are discussed in the SI Text. The simulation results for the two variants (data not shown) were similar to the results of Fig. 5, with redistribution upon fork conversion leading to slightly longer S-phase duration (average replication completion time of 37 min).

We conclude that the redistribution during S-phase of a limiting factor acting at the level of initiation of DNA replication is one way by which the problems arising due to stochastic origin activation can be circumvented.

Discussion
Simulating Full Genome DNA Replication. A general and flexible stochastic hybrid model was developed to capture the DNA replication process during the S-phase of the cell cycle. The model in principle applies to any organism; here it was instan-
The fraction of unreplicated DNA is plotted as a function of time for 100 different simulations. Base case average (Fig. 3A) is shown as thick, gray dotted line for comparison. Note the considerably shorter tail. (B) Mean replication rate (bases/min) as a function of time. Base case (Fig. 3B) is shown as a thin line for comparison. Note again the shorter tail for the propensity redistribution model. (C) Histogram of DNA replication completion time. (D) Histogram of the number of origins that fire in each simulation. (E) Fraction of appearance is the fraction of simulations that exhibit each abcissa value. Results in (B–F) based on 2,000 Monte Carlo simulations.

**The Effect of Parameter Values.** The model’s flexibility allowed us to introduce modifications to test possible explanations for this difference. Input parameters (intrinsic firing propensities of all origins, fork speed and number of potential origins) were varied to investigate their effect on S-phase duration. While the parameters tested have a direct effect on S-phase duration, none of these changes alone seems capable of explaining in a biologically realistic way the discrepancy between experimentally measured S-phase duration and the predictions of the model. Local effects, such as an increase in fork speed at problematic regions, could conceivably speed up DNA replication completion. However, different regions create problems in different cells (see Fig. 3F) and local effects at a number of these problematic regions would be required for an effect at the population level. Combining changes in input parameters could potentially reveal a biologically plausible solution, although we have been unable to pin-point a likely combination (data not shown). More complex patterns of origin firing, such as coordination of fork velocity with interorigin distance (17) or defined spacing of active origins (15, 16) could potentially decrease S-phase duration. Such conjectures can be tested by our model and we have made use of the model’s versatility to capture one such possibility.

**Firing Propensity Redistribution.**

The model was modified to accommodate a specific biological hypothesis for the reason behind the probabilistic nature of the firing of origins: the presence of a limiting initiation factor, which is released following origin firing and binds again to origins still at the prereplicative state (18). Since the number of origins in the prereplicative state decreases as S-phase progresses, the probability of the limiting factor binding to a particular prereplicative origin (and therefore the firing probability of this origin) increases along S-phase (18)

The firing propensity redistribution model was developed to code this hypothesis, assuming that the pool of available factor remains roughly constant during S-phase. Using the initial model parameters, the firing propensity redistribution model predicts a considerable reduction in S-phase duration, while also maintaining the average fraction of origins that fire during S-phase in a reasonable range. Strikingly, while bulk replication progresses with dynamics similar to the basic model, the long tail of replication has disappeared. The firing propensity redistribution model reproduces the experimental data of (27) as well as the base-case model does (data not shown), consistent with the bulk of replication progressing in a similar manner in the two models. Firing propensity redistribution therefore offers a possible solution to the random completion problem, without the need to postulate a spacing mechanism for positioning active origins at appropriate locations. This scenario is indeed biologically plausible. The limiting factor should act at the level of origin activation, and not origin licensing that is globally inhibited following S-phase onset. It could be released at firing (Fig. 5) or upon fork conversion (data not shown). Cyclin-dependent kinases or the Dbf4-dependent kinase (DDK), which are required for initiation, are possible candidates for the first option, while Cdc45, previously shown to be limiting for replication in *Xenopus* (32), is a candidate for the second option.

"G2" Replication? An alternative explanation is that DNA replication takes longer than has been generally accepted. The predicted low percentage of unreplicated DNA (<5%, Fig. 3A), the very low rate of replication at late time points (Fig. 3B), and the fact that problematic regions (those that delay the overall process) differ in different cells (Fig. 3F), suggest that such a prolonged phase of DNA replication could have gone unnoticed. The inability of techniques used up to now to detect small
amounts of unreplicated DNA scattered across different regions in different cells means that the length of S-phase may be longer than our current estimates.

An S-phase that extends into what we now define as G2-phase is consistent with known properties of the fission yeast cell cycle and the available experimental data (Fig. 6). G2 makes up the largest part of the fission yeast cell cycle, and it would be inconsequential if small parts of the genome continued replication for an extended time while the cell grows in size. The checkpoint that arrests entry into mitosis in the presence of unreplicated DNA would become essential if the G2/M size control was inoperative. This alternative explanation can explain a number of previous observations in fission yeast. In fission yeast, S + G2 is not shortened beneath a certain minimal period (0.5 of the cell cycle) even in large cells that have reached the critical size for mitosis (33, 34). Furthermore, a wee1 mutant in which the G2/M size control is inoperative and, as a consequence, advances into mitosis at a small size (33, 35) also has a S + G2 length of approximately 0.5 of the cell cycle. Significantly, wee1 is synthetically lethal with a rad3 mutant, defective in the G2/M checkpoint control (36). We postulate that in wee1 mutant cells, the extended S-phase occupies all of the previously designated “G2” period and mitosis is restrained by the Rad3-dependent G2/M checkpoint. When the checkpoint is inacti-

vated, cells cannot restrain entry into mitosis even though S is not complete, and this leads to lethality. The minimal S + G2 (65–120 min depending on growth conditions), is similar to the extended S-phase suggested here (Fig. 3A). Additional support for the extended S-phase hypothesis comes from studies of New-End Take-Off (NETO), a growth transition that is dependent on the completion of DNA replication and takes place in early G2. This occurs at approximately 0.35 of the cycle (37) rather than 0.1 of the cycle, the time when DNA replication is believed to be completed (30). NETO shows heterogeneity among cells in a population (38), consistent with the results presented here.

Any organism with an active checkpoint inhibiting entry into mitosis until DNA replication has been completed might not require a further specific mechanism to inhibit the random gap problem. In this light, part or all of what we now experimentally define as G2-phase might be better described as the period required for the cell to complete replication of the last remaining parts of its genome. Proving or disproving the extended S-phase hypothesis will have to await the development of more sensitive methods, able to detect the low amounts of late replication suggested by the model described here.

ACKNOWLEDGMENTS. We are grateful to Jacqueline Hayles for critical reading of the manuscript and helpful suggestions. This work was supported by the European Commission under the project HYGEIA (FP6-NEST-004995).

Supporting Information

Lygeros et al. 10.1073/pnas.0805549105

SI Text

Model Description

Discrete dynamics. The discrete dynamics of the model capture instantaneous changes in the state of the system. One such change is the transition of an origin from the prereplicative to the postreplicative state. An origin is called prereplicative at the beginning of S-phase, when it is ready to fire and becomes postreplicative either when it fires or when it is passively replicated by a replication fork from an adjacent origin. Although in reality this transition takes some time, it is much faster than the rest of the dynamics of the system (e.g., the movement of the replication forks) and can therefore be accurately abstracted by an instantaneous switch that takes place when either the origin fires, or when a replication fork of an adjacent origin reaches it.

For logistical reasons, the distinction between pre- and postreplicative state needs to be refined further in our model. In particular, we need to distinguish whether the postreplicative state is reached because the origin fired or because it was passively replicated. After an origin has fired, we also need to distinguish cases where both its replication forks are active, from cases where only the right fork is active (because the left fork encountered the right replication fork of another active origin to the left of the origin in question), from cases where the left fork is active (for the symmetric reason), from cases where neither fork is active. The discrete state of each origin, $i$, can be captured by a variable, $S_i$, that takes one of six values,

$$S_i \in \{PreR, RB, RR, LR, PostR, PassR\}.$$ 

Initially all origins are in the prereplicative state, therefore $S_i = PreR$ for all $i = 1, 2, \ldots, N$. The transition from $PreR$ to $RB$ is “spontaneous” and takes place when the origin “decides” that the time to fire has come. In our model, the timing of this transition is probabilistic; the mechanism driving it is described under the paragraph Stochastic Dynamics below. The remaining transitions are “forced,” in the sense that they have to take place when certain conditions on the movement of the replication forks are met. For example, the transition of an origin $i$ from $PreR$ to $PassR$ (that represents passive replication) takes place when either the left replication fork of the first active origin to the right of $i$, or the right replication fork of the first active origin to the left of $i$ reaches the position, $X_i$, of $i$. States $RB$, $LR$ and $RR$ are used to discriminate origins from which active forks emanate to both directions ($RB$), only to the left ($LR$) or only to the right ($RR$) due to encountering a fork progressing from the opposite direction (fork conversion). In the conventional licensing terminology, $PassR$, $PostR$, $RB$, $RR$, and $LR$ are subdivisions of the postreplicative state. Transitions between states are governed by “guards” (listed next to the arrow that represents the corresponding transition in Fig. 1 of the main text). The guards are logical statements involving the variables of the model. When the statement becomes true the corresponding transition is taken.

The guards make use of the right and left neighbors of a given origin, i.e., origins to the left and to the right that are actively replicating at this instant

$$LN(i) = \max\{j < i | S_j \in \{PreR, PostR, PassR\}\}$$

$$RN(i) = \min\{j > i | S_j \in \{PreR, PostR, PassR\}\}$$

The guard for the transition from $PreR$ to $PassR$ can then be succinctly written as

$$G_{PreR \rightarrow PassR} = [X_{LN(i)} + R_{LN(i)} \geq X_i] \lor [X_{RN(i)} - L_{RN(i)} \leq X_i],$$

where $R_{LN(i)}$ and $L_{RN(i)}$ denote the progress of the right and left replication forks (respectively) of these neighboring origins, and $\lor$ denotes “or.” The remaining transition guards are

$$G_{RR \rightarrow RR} = [X_{LN(i)} + R_{LN(i)} \geq X_i - L_i]$$

$$G_{PreR \rightarrow RB} = [i \geq T_i]$$

$$G_{RR \rightarrow LR} = [X_{RN(i)} - L_{RN(i)} \leq X_i + R_i]$$

$$G_{RR \rightarrow PostR} = G_{RB \rightarrow LR}$$

$$G_{LR \rightarrow PassR} = G_{RR \rightarrow PassR}$$

The guard $G_{PreR \rightarrow RB}$ reflects the probabilistic transition that takes place when origin $i$ fires; $T_i$ denotes the current time and $T_i$ the random origin firing time, whose probability distribution is discussed below.

Continuous dynamics. The continuous dynamics of the model capture evolutions that are slow compared to the discrete transitions discussed above and can therefore be considered as taking place continuously over a time interval. In our model the only such evolution is the movement of the replication forks. The progress of the replication forks is measured in terms of the number of bases replicated. Even though this number is finite by nature, it is very large and it is therefore reasonably accurate to capture the progress of the replication fork by a continuous quantity that changes according to a differential equation. The rationale behind this approximation becomes clearer if we consider normalizing the number of replicated bases by the total number of bases, $L$, (which is more than 12 million for S. pombe) and looking at the resulting number, which will be between $\theta$ and $I$ (reflecting the fraction of the genome that has been replicated).

When origin $i$ fires it gives rise to two replication forks moving away from the origin to the left and to the right. We denote by $L_i$ and $R_i$ the number of bases that these forks have replicated respectively. The forks move (i.e., $L_i$ and $R_i$ increase) at a velocity $\nu(x)$ which depends on the position, $x$, of the genome currently being replicated by the fork. The progress of the forks stops when they encounter replication forks moving in the opposite direction (reflecting in the discrete state of the model). The progress of the replication forks can therefore be captured by two differential equations

$$\frac{dR_i(t)}{dt} = \begin{cases} \nu(X_i + R_i(t)) & \text{if } S_i(t) \in \{RB, RR\} \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{dL_i(t)}{dt} = \begin{cases} \nu(X_i - L_i(t)) & \text{if } S_i(t) \in \{RB, LR\} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, the moment origin $i$ fires the number of bases its forks have replicated is zero, therefore we initialize the differential equations with $L_i(0) = R_i(0) = 0$. Notice that there is a tight coupling between the discrete and continuous states of the process: The transitions of the discrete state depend on the value of the continuous state and the evolution of the continuous state depends on the value of the discrete state. This tight coupling of
the continuous and the discrete is the defining feature of hybrid systems.

**Stochastic dynamics.** In our model we assume that, in the absence of passive replication, the firing time, $T_i$, (in minutes), of origin $i$ is governed by an exponential distribution. More specifically, we assume that in the absence of passive replication the probability that origin $i$ has not fired by time $t$ (in minutes) decreases exponentially in $t$:

$$\text{Probability}[T_i \geq t] = e^{-\lambda t}.$$  

Recall that the positive number $\lambda_i$ reflects the firing propensity of origin $i$ in a unit of time in the absence of passive replication. The exponential distribution offers theoretical and practical advantages and is widely used in many applications, among them telecommunication networks, manufacturing, insurance, and finance. For example, the exponential distribution is memoryless. Moreover, the parameter $\lambda_i$ may in general depend on the discrete state of the model (e.g., the number of origins that are prereplicative) and/or the continuous state (e.g., the fraction of unreplicated DNA). This observation is exploited below for the development and simulation of the firing propensity redistribution model.

The parameter $\lambda_i$ for each origin needs to be selected based on experimental data. To do this we use the fraction of cells, $FP_i$, in which origin $i$ was observed to fire in hydroxyurea experiments, where replication forks are halted after moving only for a few thousand bases, thus preventing passive replication. It is easy to see that under the exponential distribution and in the absence of passive replication the probability that origin $i$ fires by a given time $T_f$ (in minutes) is given by

$$\int_0^{T_f} \lambda_i e^{-\lambda t} dt = 1 - e^{-\lambda T_f}.$$

If we equate this number to $FP_i$, we can obtain an estimate of the intrinsic firing propensity, $\lambda_i$, of origin $i$ as a function of the experimentally observed fraction $FP_i$:

$$\lambda_i = -\frac{1}{T_f} \ln(1 - FP_i).$$

Clearly there is some arbitrariness involved in the choice of $T_f$. Initially we set $T_f = 20$ minutes to reflect the generally accepted view that the S-phase for *S pombe* lasts about 20 min. The sensitivity of the results with respect to this choice is discussed in the main body of the paper.

We note two subtle assumptions of our model. The first is that the firing processes of different origins are assumed to be statistically independent. The second is that the value of $\lambda_i$ is assumed to be constant during S-phase. This implicitly means that origin efficiency (weak vs. strong) and origin firing time (early vs. late) both boil down to the same mechanism; strong origins will also tend to fire early and weak origins will tend to fire late. The reason is that the mean of the exponential distribution is $1/\lambda_i$. Therefore, a high value of $\lambda_i$ implies that, in the absence of passive replication, origin $i$ will tend to fire both more often and earlier. In the presence of passive replication the correlation will not be perfect, of course. An alternative firing propensity re-distribution model that relaxes both assumptions is discussed below.

**Firing propensity redistribution model.** The structure of the firing propensity redistribution model is the same as that of the basic model outlined above. The only difference lies in the calculation of the intrinsic firing propensities which change every time an origin fires or gets passively replicated. As before, let $\lambda_i$ denote the initial firing propensity of origin $i$, computed by the method discussed in the previous section and let $\bar{\lambda}_i(t)$ denote the firing propensity of origin $i$ at time $t$. Initially we set $\lambda_i(0) = \lambda_i$. Let also

$$Q(t) = \{i|S_i = \text{PreR}\} \subseteq \{1, 2, \ldots, N\}$$

denote the set of indices of the origins in the prereplicative state at time $t$; initially we set $Q(0) = \{1, 2, \ldots, N\}$. Assume that at some time, $t$, origin $i$ either fires, or becomes passively replicated. At this time, the index $i$ is dropped from the set $Q(t)$, and the firing propensity of origin $i$ is redistributed among the origins remaining in $Q(t)$. We assume that the redistribution is proportional to the initial firing propensity of the remaining prereplicative origins, in other words

$$Q(t) = Q(t^-) - [i]$$

and

$$\bar{\lambda}_i(t) = \bar{\lambda}_i(t^-) + \frac{\lambda_i(t^-)}{\sum_{k \in Q(t)} \lambda_k} \sum_{k \in Q(t)} \lambda_k(t^-)$$

for all $j \in Q(t)$ where $t^-$ denotes the time just before the discrete transition (firing or passive replication) takes place; more formally, $Q(t^-) = \lim_{\tau \rightarrow t^-} Q(\tau)$. It is easy to see that in this case the firing propensity of individual origins increases monotonically along the S-phase, while the total firing propensity remains constant

$$\bar{\lambda}(t) = \sum_{i \in Q(t)} \bar{\lambda}_i(t) = \sum_{i = 1}^N \lambda_i.$$

Two alternative redistribution models were also implemented. The first alternative is the same as above, with the exception that redistribution takes place according to the current firing propensity of the *PreR* origins and not their initial firing propensity. In other words

$$\bar{\lambda}_i(t) = \frac{\bar{\lambda}_i(t^-)}{\sum_{k \in Q(t)} \lambda_k} \sum_{k \in Q(t)} \lambda_k(t^-)$$

for all $j \in Q(t)$.

For this first alternative redistribution model the total firing propensity also remains constant.

The second alternative redistribution model captures the situation where redistribution takes place upon fork conversion and not upon firing. In this case, redistribution takes place either when origin $i$ is passively replicated (transition *PreR* to *PassR*) or when forks meet (transitions *RR* to *RR*, *RB* to *LR*, RR to *PostR*, or LR to *PostR*). In both cases, the computation of $Q(t)$ proceeds as before. In the former case all of the firing propensity $\bar{\lambda}_i(t^-)$ gets redistributed among origins in $Q(t)$ proportionately to their initial firing propensities. In the latter case, one half of the firing propensity $\lambda_i(t^-)/2$ gets redistributed. One can see that in this case the total firing propensity will not be exactly constant, since origins in the *RR*, *LR*, and *LR* states are not in $Q(t)$ but still hold on to (part of) their firing propensity.

In all cases, the parameters used in the results $[L, N, T_f, \nu(x)]$ were the same as for the base case of the original model.

**Fission yeast instantiation: Modeled genome areas and input parameters.**

The locations and intrinsic firing propensities of origins along the fission yeast genome were taken from (Heichinger et al., 2006). The genome sequence release of the fission yeast genome of April 2004 from the Sanger Center was used. Highly repetitive regions of the fission yeast genome were excluded from the origin mapping analysis and were therefore also excluded from simulations. These are telomeric and subtelomeric repeats, centromeric and subcentromeric repeats, and ribosomal RNA repeats. The left and right arms of chromosomes 2 and 3 were modeled as separate pieces to avoid artifacts from artificially
joining them following extraction of centromeric regions. Chromosome 1 was modeled as one piece, since the length of the unmapped centromeric region was ~15 kb, which is no different from the mean inter-origin distance. DNA replication was modeled for a total area of \( L = 12,039,987 \) bases containing \( N = 893 \) potential origins. Specifically the modeled regions were:

- Chromosome 1: base 90,532 to 5,494,088. 5.4 Mb containing 408 origins
- Chromosome 2-Left arm: base 95,383 to 1,592,054. 1.5 Mb containing 105 origins.
- Chromosome 2-Right arm: base 1,649,764 to 4,426,827. 2.8 Mb containing 207 origins.
- Chromosome 3-Left arm: base 33,072 to 1,065,271. 1 Mb containing 83 origins.
- Chromosome 3-Right arm: base 1,144,065 to 2,431,896. 1.4 Mb containing 90 origins.

Location and intrinsic firing propensities of origins used as model input are shown in supporting information (SI) Table S1.

To specify the locations of additional putative origins, a bioinformatics analysis of the properties of all known fission yeast origins was carried out. Consistent with previous analyses, mapped fission yeast origins were shown to localize to intergenic regions which were over 500 bp in length. The maximum AT content of 500 bp windows within each intergenic region was shown to be a good predictor of origin activity. A moving AT content threshold was therefore used to progressively increase the number of intergenic regions that were included as weak origins.

**Hardware and software.** The model was implemented in Matlab 2006b on a dual core (2 × 1.83GHz) computer with 2048MB of RAM. The results are generally based on 2,000 Monte-Carlo runs. To keep the memory requirements manageable, continuous states were added as they became necessary and dropped after they ceased being active. Forced transitions (all except the transition from \( \text{PreR} \) to \( \text{RB} \)) were simulated by the event detection facilities of the ode45 Matlab routine used to integrate the continuous dynamics. The spontaneous, probabilistic transition from \( \text{PreR} \) to \( \text{RB} \) was simulated by extracting a random variable uniformly in the unit interval, taking its logarithm, dividing by \(-\lambda_\ell\), and waiting for the simulation time to reach the resulting value. The probability distribution for the time at which this happens obeys the exponential distribution with rate \( \lambda_\ell \).

**Simulation and diagnostics.** A simple calculation suggests that, even though the model used to capture the behavior of individual origins is relatively simple, the resulting genome-wide model can be very complex. The number of discrete states and the dimension of the continuous state space can be very large, with up to \( 2^N \) continuous states and \( 6^6 \) discrete states being activated, with \( N = 893 \) for the fission yeast genome. While only a fraction of these states will be visited in any one execution of the model, the analysis and even the simulation of such a model can be a formidable task. In the implementation of our model for computer simulation special care had to be taken to ensure that the simulations were faithful to the model and simulation time and memory requirements were kept reasonable.

Fig. S1 shows examples of simulations that represent full genome replication in individual cells for the basic model. In Fig. S1A, the replication process in a specific region of the genome, containing 11 origins, is depicted for two simulations (C1 and C2). The stochastic nature of the process is evident in the different location of active origins in each simulation, the different timing of firing of each origin and the different total time required to complete replication of this genomic region. In Fig. S1B, the replication time of each position \( x \) along a second genomic region is shown for four simulations. The stochasticity of the replication process is evident.

To ensure that the programming code accurately captures the model dynamics, two separate implementations were generated. The first was based on a continuous time simulation of the differential equations governing the evolution of the continuous state, coupled with event detection to simulate the evolution of the discrete state. The second implementation is based on an explicit algebraic solution of the differential equations, followed by a discrete time simulation from one discrete event to the next. The former implementation is more general but also more computationally demanding, while the latter only applies to cases where the fork speed is constant, but can be considerably faster. Both implementations were tested on the fission yeast instantiation discussed here and produced identical results (albeit at a different computational cost). This is a clear indication that both computational implementations indeed capture the dynamics of the mathematical model correctly.

Next, to ensure that the model dynamics correctly capture our current understanding of the DNA replication process, several tests were carried out using the two computational implementations. The simulation tests showed the following properties:

1. The fraction of unreplicated DNA always starts at 1 and decays monotonically to 0.
2. The total amount of DNA in the nucleus starts with 1 genome length and monotonically increases to 2 genome lengths.
3. All parts of the genome (at inter-putative origin granularity) started with one copy and ended with two.
4. All origins started in the \( \text{PreR} \) state and ended up either in the \( \text{PostR} \) state or in the \( \text{PassR} \) state. Further, the number of \( \text{PreR} \) origins was monotone decreasing while the number of \( \text{PostR} \) and \( \text{PassR} \) origins was monotone increasing.
5. The number of \( \text{RR} \) origins was always (roughly) equal to the number of \( \text{LR} \) origins. Small differences are due to origins near the edges of the chromosome pieces completing the replication to the end of the piece.

Fig. S2 demonstrates these properties for two runs of the model, one for the base-case model (without redistribution, left column) and one for the firing propensity redistribution model (right column).

---

**A**

<table>
<thead>
<tr>
<th>i</th>
<th>Time: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ii</th>
<th>Time: 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iii</th>
<th>Time: 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iv</th>
<th>Time: 1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>v</th>
<th>Time: 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vi</th>
<th>Time: 2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vii</th>
<th>Time: 3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>viii</th>
<th>Time: 4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ix</th>
<th>Time: 5.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>x</th>
<th>Time: 7.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>xi</th>
<th>Time: 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>xii</th>
<th>Time: 12.84</th>
</tr>
</thead>
<tbody>
<tr>
<td>c₁</td>
<td></td>
</tr>
<tr>
<td>c₂</td>
<td></td>
</tr>
</tbody>
</table>

**B**

![Graphs showing time of replication](image)

**Fig. S1.** Example simulation runs of DNA replication across the fission yeast genome. (A) Snapshots of a “movie” of two simulations of the replication process in a given piece of the genome containing 11 putative origins. For each snapshot the bar at the top shows the locations of potential origins, the two bars labeled C₁ and C₂ show the two simulations, and the time corresponding to the snapshot is listed in the bottom; origins that have fired in the simulations are marked by a vertical line and the replicated pieces of the DNA by a thick horizontal line. Stochastic phenomena in the location of active origins and timing of firing are evident. As a consequence, the time required to complete replication of this region differs between simulations (10 min for C₁ versus 13 min for C₂). (B) Results of 4 different simulations a 40-kb piece of the genome containing 5 putative origins, showing the time of replication of each genomic region. Locations of potential origins are marked at the top, the horizontal axis represents location in the genome and the vertical axis the time at which each location was replicated in the particular simulation. Differences in the timing profile in different simulations are evident.
Fig. S2. Properties of base-case and firing propensity redistribution model. (A, C, and E) Base case model, (B, D, and F) Firing propensity redistribution model. (A) and (B) show the total system firing propensity (the sum of the firing propensities of origins available to fire, in blue) and the mean firing propensity of origins available to fire (in green) during the course of S-phase. Note that the total system firing propensity decreases during the course of S-phase for the base-case model and remains constant for the firing propensity redistribution model. In contrast, the firing propensity of a given origin remains constant in the base case model and increases in the redistribution model. The mean firing propensity tends to decrease somewhat in the base model, as more efficient origins tend to fire earlier. (C) and (D) show the number of origins that are found in each one of the six discrete states during the course of S-phase. PreR red, PassR purple, RB light blue, RL dark blue, RR black, and PassR yellow. Note that the RL and RR curves effectively overlap and that for both models, the majority of origins are passively replicated. (E) and (F) show the number of inter-origin locations which have one copy (black line) or two copies (red line) during the course of S-phase.
Other Supporting Information Files

Table S1.xls