Supplementary information for Zhang *et al.* "Cell-to-cell variability and robustness in S-phase duration from genome replication kinetics"

Supplementary Text

Fitting replication timing data from experiments using the model

This section describes our fitting procedure based on the model. The fitted parameters were used in simulations of genome replication kinetics can giving the distribution of S-phase duration and of replication time of one chromosome (Fig. S10).

We used flow cytometry (FACS) data to re-normalize replication timing as follows. If the base line value of average DNA copy-number a is remarkably larger than 1, and/or its plateau value b is remarkably smaller than 2, we use the formula $y = a + \frac{(b-a)(t-T_0))^r}{(t-T_0)^r+(t_c-T_0)^r}\theta(t-T_0)$ to fit the FACS data and normalize replication timing data by $\phi_{\text{norm}}(x,t) = 1 + \frac{\phi(x,t)-a}{b-a}$, where ϕ is the replication probability function [1].

We used fixed origin locations from the literature and optimized the fit for the parameters γ, T_0, v and λ_i iteratively. The objective function was defined as the L2 distance (the average of squared differences) of the experimental and theoretical replication probability timing profile (Fig. S10), i.e., as $\sqrt{\sum_i \sum_j (\phi_{model}(x_i,t_j) - \phi_{exp.}(x_i,t_j))^2/(N_x N_t)}$, where N_x and N_t are the numbers of the measured loci and time points respectively.

Initialization of the parameters for the fits was performed as follows. Firing rate exponent γ and fork velocity v were initialized at arbitrary values (typically γ at 0, v at 2 kb/min). The start of S phase T_0 was initially set when genome copy number from the normalized FACs data (from the interval $[a,b]$ to $[1,2]$) is first larger than a fixed threshold (e.g. 1.05) and each origin strength λ_i starts from the value fitted with the time-course data at this origin.

Fitting was performed with following iterative rule. 1) for a parameter x, assume it has a step length Δ_x, and a memorized step length $\Delta_x^{'} = 2\Delta_x$, 2) set $r = \Delta_x/\Delta_x^{'}$ and $\Delta_x^{'} = \Delta_x$, if $x + \Delta_x$ gives a better fit than x, let $x = x + \Delta_x$, otherwise (i) if $|r| = 1$, we update $\Delta_x \to \Delta_x/2$ (ii) if $|r| = 0.5$, set $\Delta_x \to -\Delta_x$; 3) repeat 2) until the termination condition is satisfied. λ_1, λ_2, ..., λ_n for each

chromosome are updated iteratively given γ, v and T_0 and in each iteration, one λ_i is chosen randomly to be updated. T_0 is updated iteratively given γ and v. v is updated iteratively given γ. For γ, we tested some discrete values between 0 and 3. Supplementary Fig. S8a,b indicate the best fit value of γ for *S.cerevisiae* and *L.kluiveri*, and Supplementary Fig. S8c shows one example of the best fit.

Role of chromosome boundaries in replication timing

In some simulations, we used circularized chromosomes for easier comparison with the analytical estimates, but relative to a circular chromosome, a linear chromosome has lower symmetry because of the boundary at both ends. To verify that this assumption does not qualitatively affect the results, we circularized the empirical *S.cerevisiae* chromosomes by linking their ends respectively, and simulated their replication kinetics with the estimated parameters. The results (Fig. S2) show that the circularized chromosomes always replicate faster than the linear chromosomes, but their durations do not differ much (the average deviation is in all cases less than 15%).

Determination of the parameters α, β and t_0 in the formula for the distribution of T_i

Eq. 3 in the main text, describing the replication timing of one inter-origin region contains the parameters α, β and t_0, which need to be related to the biologically measurable parameters (inter-origin distance and origin rates). To estimate such parameters for the distribution of T_i we used two methods. The first is a fit of all the T_i data taken from the simulation of the given chromosome, and the second is to fit the specific T_i data (replication times of the central inter-origin region) extracted from simulation of a linear chromosomal fragment where inter-origin distances and origin strengths are sampled from known distributions (different samples for different runs of the simulation). In this second method, each run of the simulation is carried out considering inter-origin distances and origin strengths with the same averages as the original chromosome. Both methods give the same distribution for T_i, which agrees very well with Eq. 3 of the main text (See Fig. S3).

We mainly used the second method since it does not depend on origin configuration of the original chromosome. The detailed procedure is the following. First, we defined a characteristic distance $d_c = (\frac{\gamma+1}{\langle\lambda\rangle}\log(\frac{1}{1-x}))^{\frac{1}{1+\gamma}}v$, where $x < 1$ (e.g. 0.99) and assume $n_c = \min(\lfloor d_c/\langle d\rangle\rfloor + 1, \lfloor n/2\rfloor)+1$. Then we produced a linear chromosomal fragment with $2n_c$ origins, in which two origins are always located at the ends. Next, we simulated many realizations for the replication of this chromosome. In each simulation run, we sampled inter-origin distance d_i, origin strength λ_j and origin firing time $t_f^{(j)}$ from $\Gamma(\frac{\langle d\rangle^2}{\sigma^2(d)}, \frac{\langle d\rangle}{\sigma^2(d)})$, $\Gamma(\frac{\langle\lambda\rangle^2}{\sigma^2(\lambda)}, \frac{\langle\lambda\rangle}{\sigma^2(\lambda)})$ and $f(t) = \lambda_i t^\gamma \theta(t)\exp(-\lambda_i\frac{t^{\gamma+1}}{\gamma+1})$ respectively, where $i \in \{1, 2, ..., 2n_c - 1\}$ and $j \in \{1, 2, ..., 2n_c\}$. The statistics over different realizations gives the distribution

of the replication time of the central inter-origin region (T_{n_c}), which was fitted with Eq. S1 to obtain α, β and t_0.

Analytical derivation of an approximate distribution of S-phase duration T_S based on extreme value theory.

This section gives further details on the analytical calculation for the extreme-value estimate of the distribution of S-phase duration. We assume that replication timing of one inter-origin region T_i obeys the stretched exponential distribution

$$F(t) = P(T_i < t) = 1 - e^{-\alpha(t-t_0)^\beta} \ , \tag{S1}$$

where $t \geqslant t_0$ and $\alpha > 0$. The parameters α, β and t_0 were obtained as described in the previous section. We define $M_n = \max(T_1, T_2, ..., T_n)$. By taking $a_n = 1/(\alpha^{1/\beta}\beta(\log n)^{1-1/\beta})$ and $b_n = (\log n/\alpha)^{1/\beta} + t_0$, and applying the Fisher-Tippett-Gnedenko theorem, we can prove that

$$\lim_{n\to\infty} P((M_n - b_n)/a_n \leq t) = \exp(-\exp(-t)) \triangleq G(t) \ , \tag{S2}$$

where $G(t)$ is the standard Gumbel distribution.

When n is sufficiently large, we can make the approximation $P((M_n - b_n)/a_n \leq t) \approx G(t)$. If we define $\tilde{t} = a_n t + b_n$, we have $P(M_n \leq \tilde{t}) \approx G((\tilde{t} - b_n)/a_n)$.

Finally, we can represent the distribution of T_S ($=M_n$) approximately as

$$P(T_S \leq t) \approx \exp(-\exp(-\frac{t-b_n}{a_n})) = \exp\left\{-\exp\left[\beta \log n \left(1 - (\alpha/\log n)^{1/\beta}(t-t_0)\right)\right]\right\} \tag{S3}$$

Here n is the origin number, and α, β and t_0 are connected to the model parameters describing replication kinetics, v, γ, inter-origin distances $(d_1, d_2, ..., d_n)$ and origin strengths $(\lambda_1, \lambda_2, ..., \lambda_n)$.

We now discuss how α, β and t_0 can be expressed as functions of simplified parameters by numerically solving some approximate equations. We consider a "characteristic" inter-origin region with the distance $\langle d \rangle$ and origin strength $\langle \lambda \rangle$, and we assume that the replication of the inter-origin region is mainly carried out by the forks originated from the two nearest origins, both of which are typically activated, Thus we have

$$T_i \approx \langle d \rangle /2v + (t_f^l + t_f^r)/2, \tag{S4}$$

where t_f^l and t_f^r are the firing time of the left origin and the right origin respectively. Since t_0 is the minimal replication time of inter-origin region and the firing time has zero as a lower bound, one has

$$t_0 = \min(T_i) = \langle d \rangle /2v. \tag{S5}$$

From equation S4, we can further obtain

$$\langle T_i \rangle \approx \langle d \rangle /2v + \langle t_f \rangle \tag{S6}$$

and

$$\sigma(T_i) \approx \sigma(t_f) \tag{S7}$$

In addition, we have

$$\langle T_i \rangle = \alpha^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta}+1\right) + t_0, \tag{S8}$$

$$\sigma(T_i) = \alpha^{-\frac{2}{\beta}} \left[\Gamma\left(\frac{2}{\beta}+1\right) - \Gamma^2\left(\frac{1}{\beta}+1\right)\right], \tag{S9}$$

$$\langle t_f \rangle = \left(\frac{\gamma+1}{\langle\lambda\rangle}\right)^{\frac{1}{\gamma+1}} \Gamma\left(\frac{\gamma+2}{\gamma+1}\right), \tag{S10}$$

and

$$\sigma(t_f) = \left(\frac{\gamma+1}{\langle\lambda\rangle}\right)^{\frac{1}{\gamma+1}} \sqrt{\Gamma\left(\frac{\gamma+3}{\gamma+1}\right) - \Gamma^2\left(\frac{\gamma+2}{\gamma+1}\right)} \tag{S11}$$

Based on equations S5-S11, α and β can be numerically solved as functions of v, γ, $\langle d\rangle$ and $\langle\lambda\rangle$. Our simulations in the EVD regime, and using empirically realistic values of the parameters are in line with equations S5-S7.

References

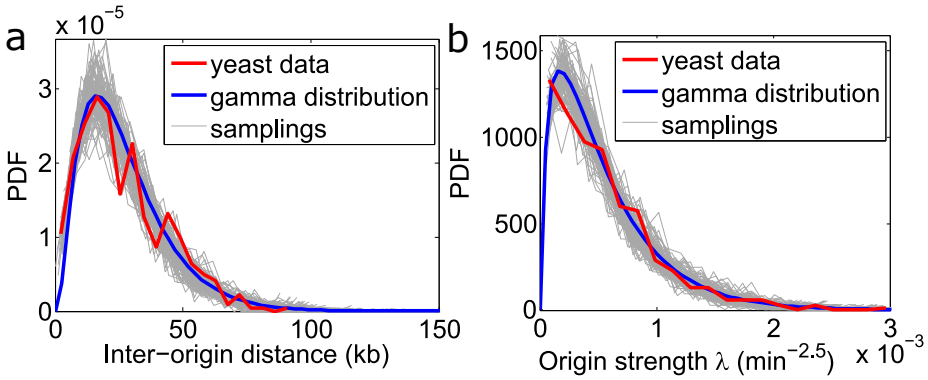
[1] Nicolas Agier, Orso Maria Romano, Fabrice Touzain, Marco Cosentino Lagomarsino, and Gilles Fischer. The spatiotemporal program of replication in the genome of lachancea kluyveri. *Genome Biol. Evol.*, 5(2):370–388, 2013.

[2] Michelle Hawkins, Renata Retkute, Carolin A Müller, Nazan Saner, Tomoyuki U Tanaka, Alessandro PS de Moura, and Conrad A Nieduszynski. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. *Cell Rep.*, 5(4):1132–1141, 2013.

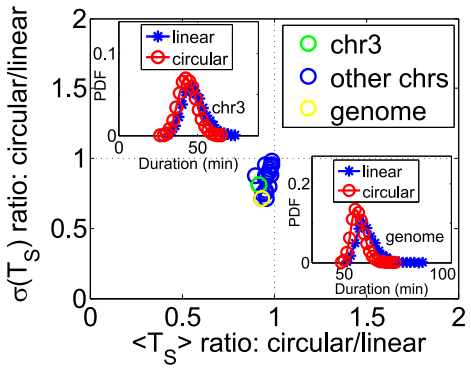
[3] Christian Heichinger, Christopher J Penkett, Jürg Bähler, and Paul Nurse. Genome-wide characterization of fission yeast dna replication origins. *EMBO J.*, 25(21):5171–5179, 2006.

[4] Jack A Vincent, Tracey J Kwong, and Toshio Tsukiyama. Atp-dependent chromatin remodeling shapes the dna replication landscape. *Nat. Struct. Mol. Biol.*, 15(5):477–484, 2008.

Supplementary Figures

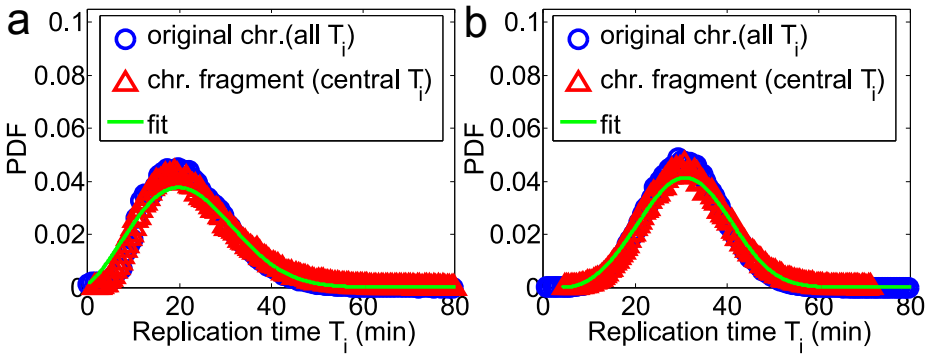


Supplementary Figure S1: **The hypothesis of gamma-distributed inter-origin distances and origin firing rates used to generate randomized chromosomes is in line with empirical data**. The plots compare inter-origin distances (a) and firing rates (b) distributions used for the model (blue continuous line) with *S. cerevisiae* data from ref. [2] (red line), and 100 samplings of the assumed distributions with the same number of instances as the empirical case (thin grey lines). Empirical firing rates were inferred setting $\gamma = 1.5$ (the best-fit value for the data in ref. [2].

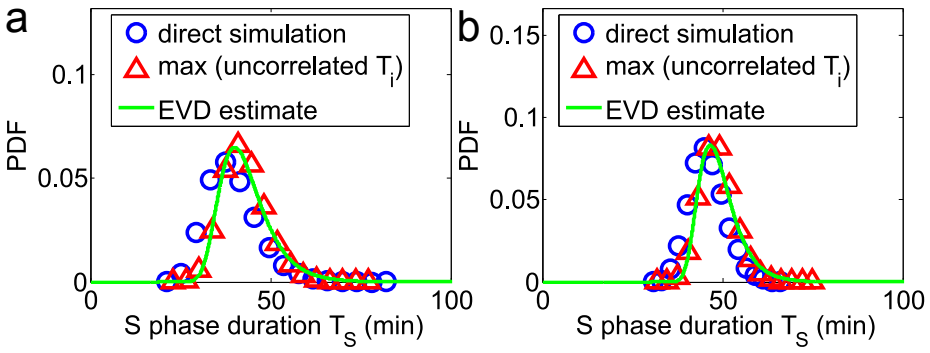


Supplementary Figure S2: **Comparison of S-phase duration of *S.cerevisiae* chromosomes and genome and their circularized versions indicates that the boundary effect on replication timing is small.** Circular chromosomes were obtained by linking two ends of the linear chromosome. The circular genome was gotten by linking all the linear chromosomes via their ends successively. Ratio of T_S average (SD) between *S.cerevisiae* linear chromosomes and the genome and the circularized versions is close to 1. The insets show that the distribution of T_S of chromosome 3 and the genome and their circularized versions are similar. The parameters giving best fit to *S. cerevisiae* data from ref. [2] were used (in particular, $\gamma = 1.5$).

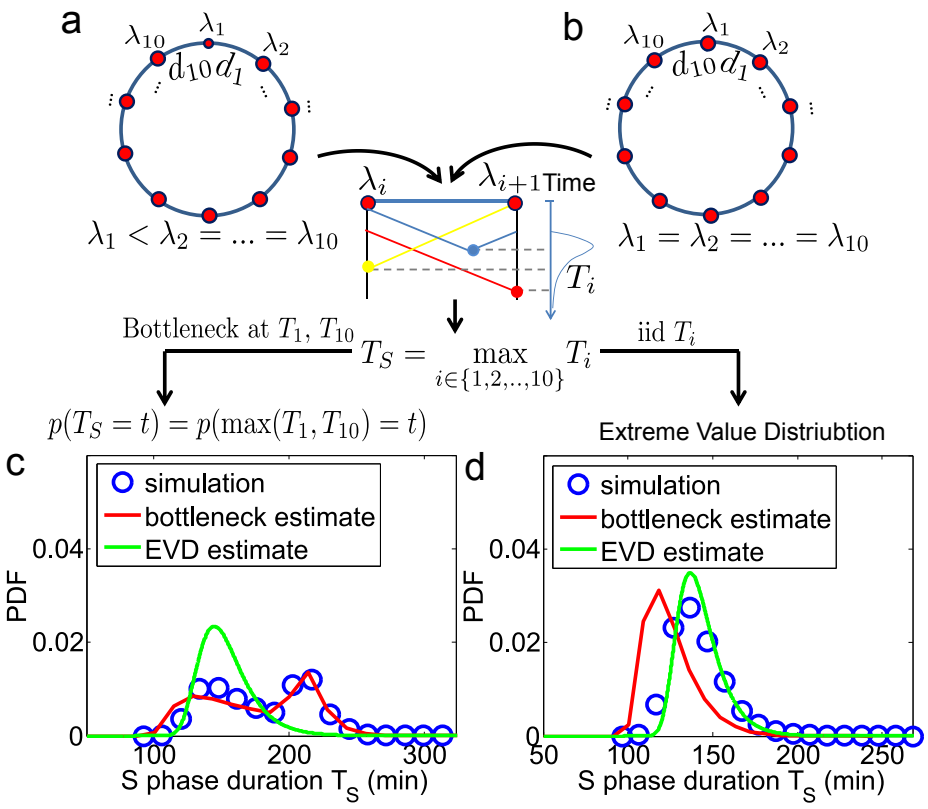
.



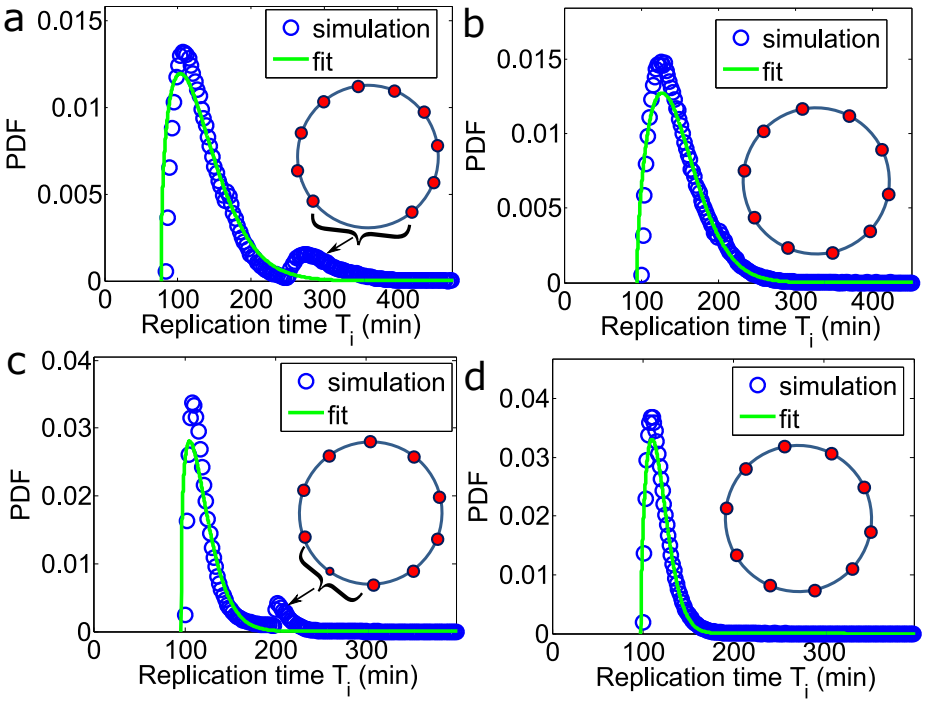
Supplementary Figure S3: **Justification of the assumption for the inter-origin replication timing distribution (Eq. 3 of the main text).** We used two methods of obtaining the data for the distribution of replication time (T_i) of inter-origin regions, both of which are in good agreement with the theoretical formula. Blue circles: distribution obtained by the simulation of an circular chromosome (original chromosome) where origin strengths and inter-origin distances are sampled with Eq.2; red triangles: distribution of replication time of the central inter-origin region in a linear chromosomal fragment where origin strengths and inter-origin distances are sampled with Eq.2 in each run of the simulation; the continuous line is a fit with Eq. 3. For (a), chromosome parameters: $\gamma = 0$, $n = 20$ (original) or 16 (linear fragment), $v = 1.88\ kb/\min$, $\langle d \rangle = 28.13\ kb$, $\sigma(d) = 13.46\ kb$, $\langle \lambda \rangle = 0.045\ \min^{-1}$, $\sigma(\lambda) = 0.036\ \min^{-1}$, and the fitted parameters: $\alpha = 3.72 \times 10^{-4}\ \min^{-\beta}$, $\beta = 2.42$, $t_0 = -1.07\ \min$. For (b), chromosome parameters: $\gamma = 1.5$ (best fit), $n = 20$ (original) or 10 (linear fragment), $v = 1.81\ kb/\min$, , $\langle d \rangle = 28.13\ kb$, $\sigma(d) = 13.46\ kb$, $\langle \lambda \rangle = 6.17 \times 10^{-4}\ \min^{-2.5}$, $\sigma(\lambda) = 5.53 \times 10^{-4}\ \min^{-2.5}$, and the fitted parameters: $\alpha = 1.79 \times 10^{-5}\ \min^{-\beta}$, $\beta = 3.21$, $t_0 = 4.16\ \min$.



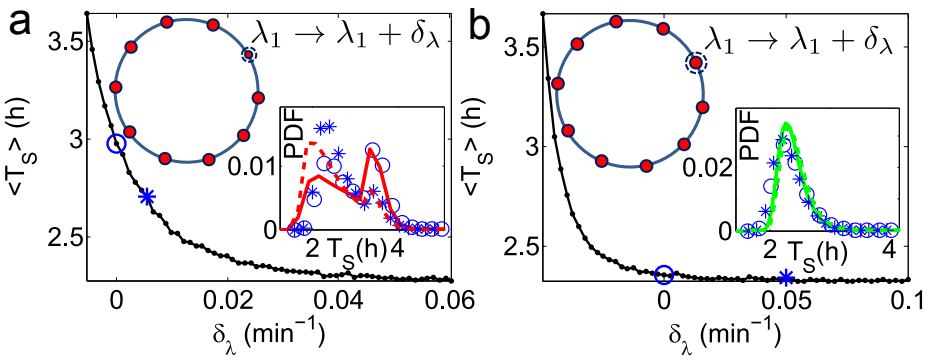
Supplementary Figure S4: **The correlation between replication times of adjacent inter-origin regions has little effect on the distribution of S-phase duration T_S.** The plot shows the distribution of S-phase duration from direct simulation (blue circles; correlated T_i) compared to sampling of $T_i (i \in \{1, 2, ..., n\})$ from $F(t) = 1 - e^{-\alpha(t-t_0)^\beta}$ independently and taking their maximum (red triangles; uncorrelated T_i). Both methods agree well with the EVD estimate based on Eq. 4 of the main text (green continuous line). The plots refers to a circular chromosome with two different parameter sets, compatible with yeast data: (a) $\gamma = 0$, $n = 20$, $v = 1.88\ kb/\min$, $\langle d \rangle = 28.13\ kb$, $\sigma(d) = 13.46\ kb$, $\langle \lambda \rangle = 0.045\ \min^{-1}$, $\sigma(\lambda) = 0.036\ \min^{-1}$, (b) $\gamma = 1.5$, $n = 20$, $v = 1.81\ kb/\min$, $\langle d \rangle = 28.13\ kb$, $\sigma(d) = 13.46\ kb$, $\langle \lambda \rangle = 6.17 \times 10^{-4}\ \min^{-2.5}$, $\sigma(\lambda) = 5.53 \times 10^{-4}\ \min^{-2.5}$. Origin strengths and inter-origin distances are sampled with Eq. 2 of the main text.



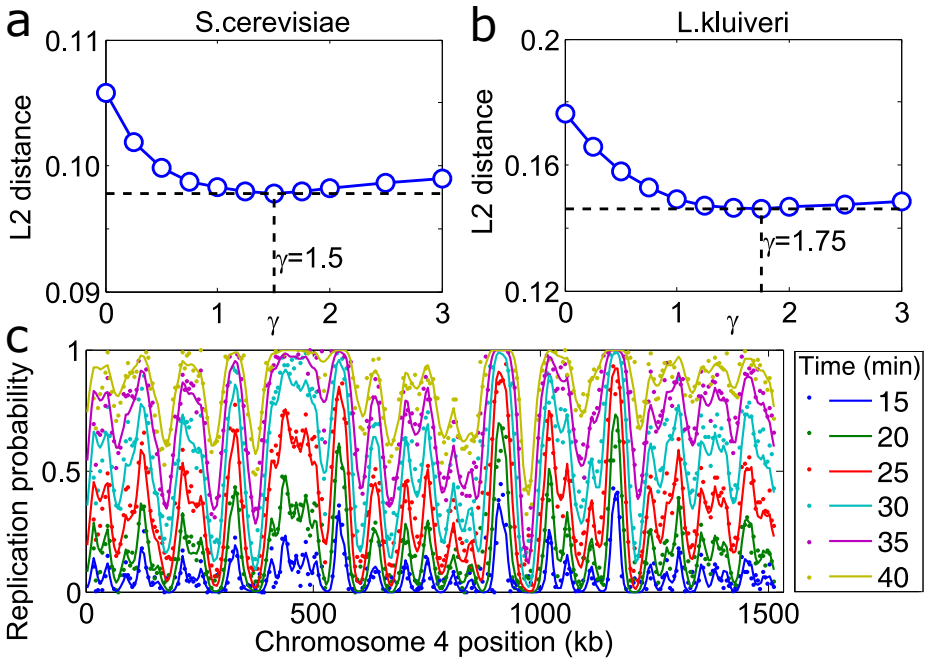
Supplementary Figure S5: **Replication regimes determined by firing rates.** (a) Due to a single slow-firing origin, the two neighboring bottleneck inter-origin regions (labelled by the index 10 and 1 in panels a and b) typically complete replication much later than the rest. Hence, T_S will be typically equal to $\max(T_1, T_{10})$ (origin strengths in the example are $\lambda_i = 0.055$ min^{-1} for all origins except $\lambda_1 = 0.0055$ min^{-1}). (b) If the replication times of all inter-origin regions are comparable, and they are considered independent and identically-distributed (iid) random variables, the distribution of T_S can be obtained by extreme-value-distribution (EVD) theory (origin strengths are $\lambda_i = 0.05$ min^{-1}). Simulations of the model (blue circles), when one inter-origin distance is much larger than the others (c), and when all inter-origin distances and strengths are comparable (d), agree with the corresponding analytical calculations (red and green curves). (Origin number $n = 10$ origins, fork velocity $v = 1$ $kb/$ min, origin strength $d_i = 200$ kb.)



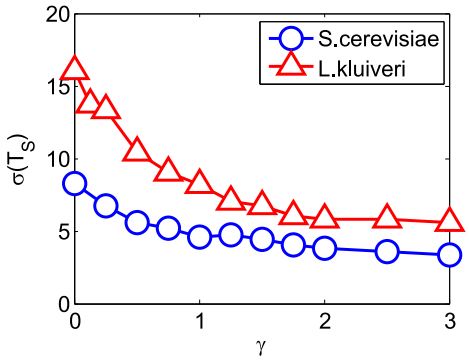
Supplementary Figure S6: **In the bottleneck regime, the slowest region in replication causes the appearance of small peaks in the right tail of the distribution of T_i, leading to the failure of the EVD estimate.** The plots come from simulations with parameter sets shown in Fig. S5 and in Fig. 2 of the main text. For the bottleneck cases shown in Fig. 2 of the main text (a) and Figure S5 (c), a small peak emerges in the right tail of T_i distribution due to the slowest replication of the bottleneck regions. Conversely, in the EVD regime, the right peak does not exist (b,d).



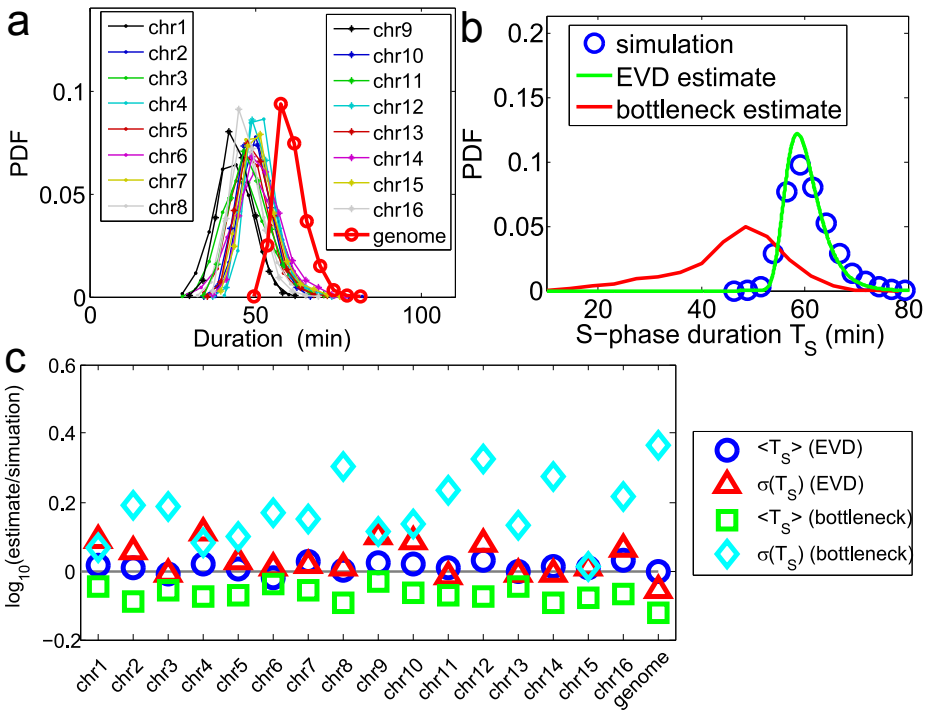
Supplementary Figure S7: **Effects of perturbations of a pair of inter-origin regions on S-phase duration.** (a) The bottleneck regions of the chromosome shown in Fig. S5a are perturbed by increasing the strength of origin 1 by δ_λ (i.e., $\lambda_1 \rightarrow \lambda_1 + \delta_\lambda$). The inset shows that the perturbation changes the distribution of T_S (circles are simulations for the unperturbed chromosome, and stars correspond to $\delta_\lambda = \lambda_1$; the two curves are the analytical estimates in the bottleneck regime). (b) The same perturbation as in (a) is performed on the strength of one origin of the chromosome shown in Fig. S5b, which lies in the EVD regime. Symbols are as in (a). The distribution of T_S is robust to this perturbation.



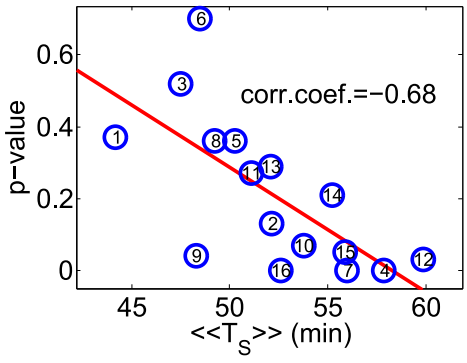
Supplementary Figure S8: **The goodness of the fit of the model with the empirical data depends on exponent factor γ and with the best γ, the model can be efficiently fit to the empirical replication data.** The empirical data of *L.kluveri* and *S.cerevisiae* are from ref. [2, 1]. (a,b) The L2 distance between theoretical and empirical replication probability profiles is minimized at $\gamma = 1.5$ (for *S.cerevisiae*) or $\gamma = 1.75$ (for *L.kluveri*) (c) The model gives a good fit to the empirical replication probability $\phi(x,t)$ from *S. cerevisiae* chromosome 4. Dots and continuous lines indicate experimental and theoretical data respectively, which are both averaged with bins of 5kb. Different colors indicate different measurement times.



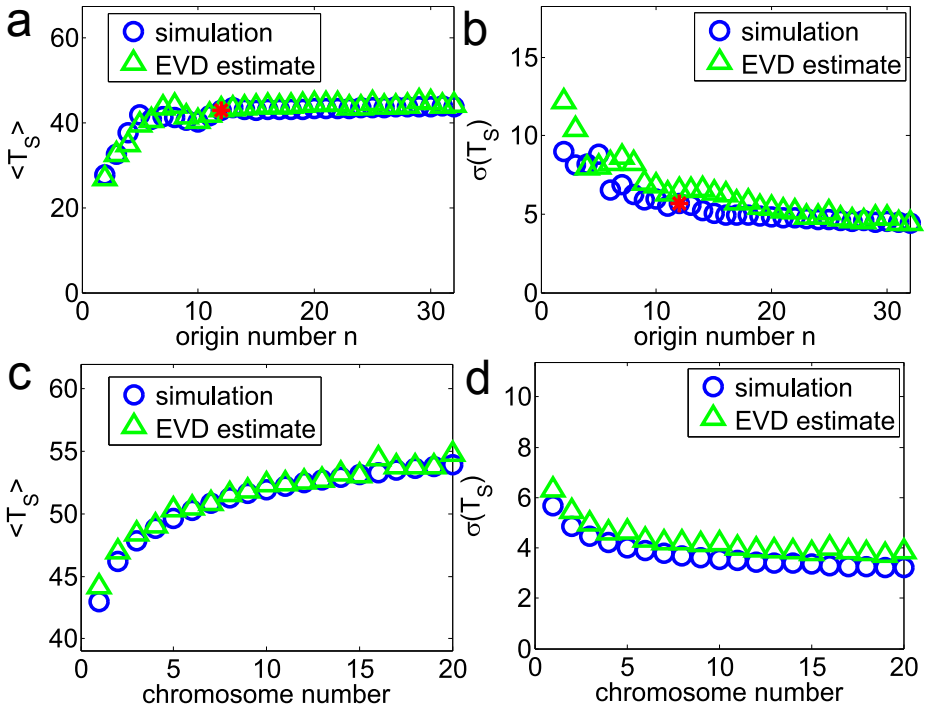
Supplementary Figure S9: **The standard deviation of S-phase duration T_S of *S. cerevisiae* and *L.kluveri* decreases with the parameter γ.** The plot is obtained from simulations with the best-fitting parameters of empirical data, using data from ref. [1, 2] (See Fig. S10 and S8)



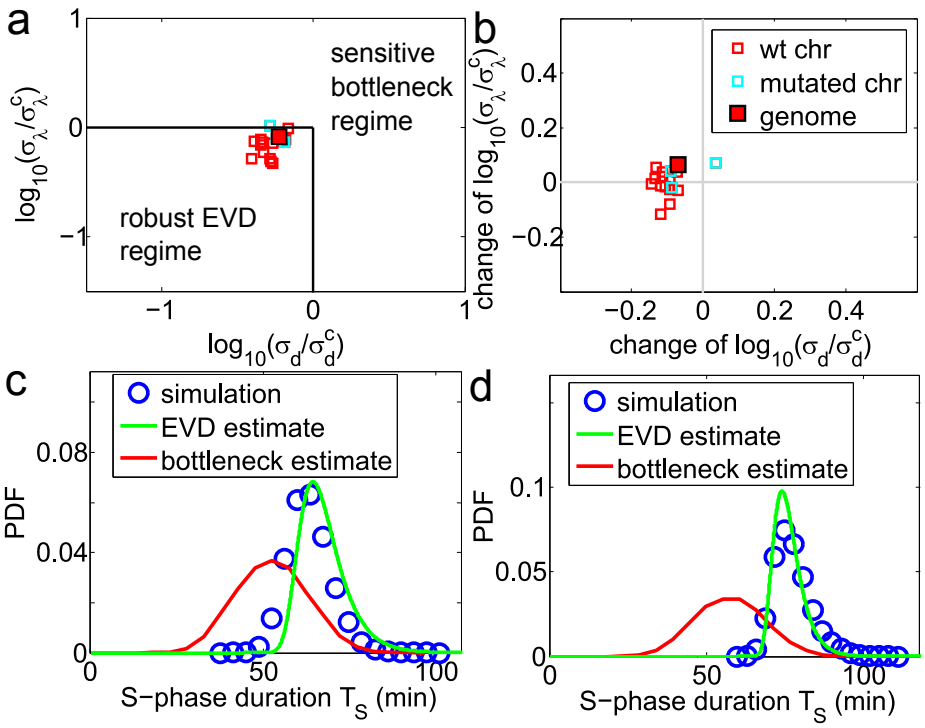
Supplementary Figure S10: **Simulated and estimated prediction for the cell-to-cell variability of S-phase duration, using the best-fitting parameters for *S. cerevisiae*.** (a) The plot shows the probability density function (PDF) of the predicted duration of the replication of chromosomes and genome of *S. cerevisiae* from simulations. The average duration of S phase compares well with measurements from flow cytometry [2]. (b) The simulated distribution (PDF) of the replication timing of the genome (circle), is well predicted by EVD estimate (green line) rather than the bottleneck estimate (red line). (c) Comparison of the average and standard deviation of the duration of the replication from analytical estimates and the simulation. The EVD estimate predicts the replication timing of the genome and all the chromosomes better than the bottleneck estimate. Data from ref. [2]. chr: chromosome.



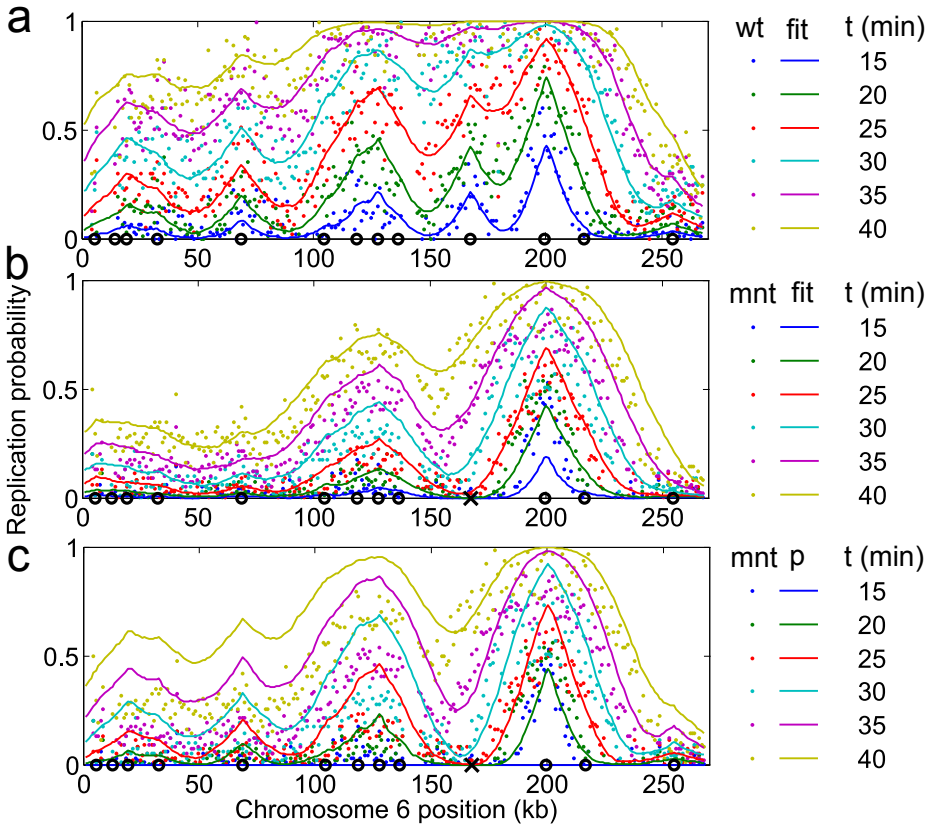
Supplementary Figure S11: **Stronger bias towards smaller replication times for slower chromosomes.** The plot shows that the p-value of the mean T_S for each *S. cerevisiae* chromosome (circles, numbered 1-16) against randomized chromosomes is negatively correlated with the typical replication timing T_S of the randomized chromosomes. See Fig. 5b of the main text. Randomized chromosomes have the same averaged inter-origin distance and averaged origin strength. The typical time in the x-axis is defined as a double mean over realizations of the parameters and over cells, i.e., realizations of the process at fixed parameters. The P-value is defined as the fraction of the mean T_S from randomized chromosomes smaller than the mean empirical T_S over the number of randomised samples.



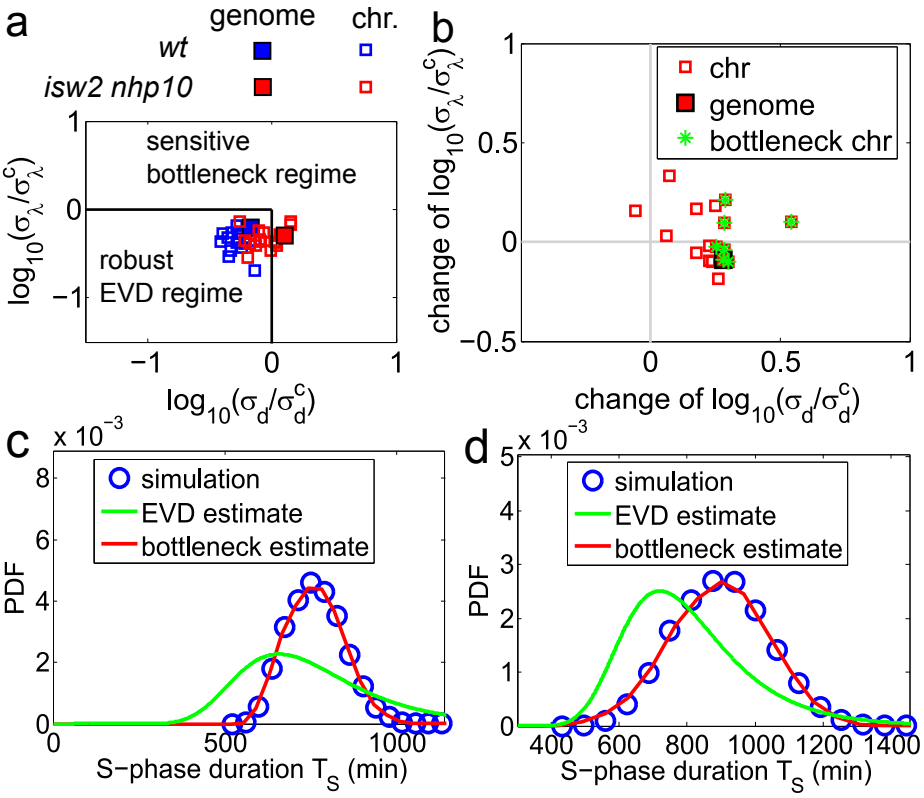
Supplementary Figure S12: **Change of the overall replication timing and its cell-to-cell variability with number of inter-origin regions and with number of chromosomes.** (a) Average of replication duration of *S. cerevisiae* chromosome I (parameters from the fit of data from ref. [2]) increases with origin number. The value saturates around n=10. (b) The standard deviation decreases with n. Red stars indicate the empirical value of n. (c,d) The average of the completion time for replication of *S.cerevisiae* chromosome 1 increases with the number of copies, whereas the standard deviation decreases.



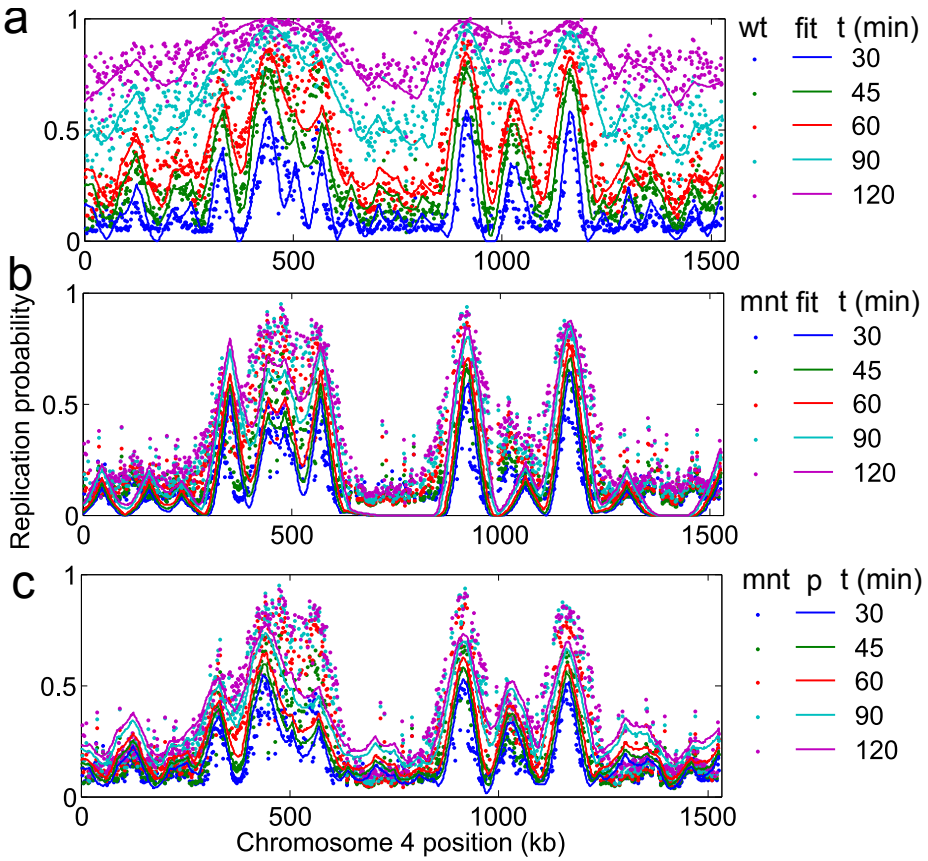
Supplementary Figure S13: ***S.cerevisiae*** **remains in the extreme-value regime under inactivation of three origins in chromosomes 6, 7 and 10.** (a) The phase diagram indicates that all the chromosomes (except for chromosome 7) and the genome remain in the extreme-value regime when origins are removed. (b) The overall relative variability of inter-origin distances for the mutant strain does not change much compared to the wt strain. The xy-axes indicate the change of the overall relative variability of inter-origin distances ($\log_{10}(\sigma_d/\sigma_d^c)$) and origin strengths ($\log_{10}(\sigma_\lambda/\sigma_\lambda^c)$) of the origin mutant strain compared to the wt strain. (c,d) The extreme-value estimate predicts well the replication duration of chromosomes (e.g. chromosome 7 shown in panel c) and the genome (panel d). The plots refer to fits of data of *S.cerevisiae* origin-impaired mutant and wt strain from ref. [2].



Supplementary Figure S14: The model gives a satisfactory prediction of replication timing profiles of *S.cerevisiae* origin mutant (mnt) strains. The plots refer to chromosome 6 as an example, and assess the performance of the model with parameters based on the wild-type fit on the mutant data, when simulations are run without the inactivated origins. Dots correspond to experimental data from ref. [2], and lines indicate a model fit or a model prediction (p). Different dot colors correspond to different times. The black circles indicate origin locations. The black cross mark shows the location of the inactivated origin. (a) Model fit of replication timing profiles of the wt strain. (b) Model fit of replication timing profiles of the origin mutant strain. (c) Model *prediction* of mutant replication timing profiles based on the best-fit parameters from the wt data.



Supplementary Figure S15: ***S.cerevisiae isw2/nhp10* mutant treated with DNA alkylating agent MMS (affecting replication forks) drives S-phase to the bottleneck regime.** (a) The phase diagram (see Fig. 4 and 5 in the main text) indicates that all the chromosomes and the genome of the wt strain are in the EVD regime while some chromosomes (4, 6, 12, 13, 14 and 15) and the genome of the mutant are in the bottleneck regime (b) The relative variability of the inter-origin distances for the chromosomes and the genome of the mutant strain is higher that of the wt strain (except for chromosome 1). The green stars indicate that the chromosomes/genome of the mutant strain is inside the bottleneck regime. The xy-axes indicate the change of the overall relative variability of inter-origin distances ($\log_{10}(\sigma_d/\sigma_d^c)$) and origin strengths ($\log_{10}(\sigma_\lambda/\sigma_\lambda^c)$) of the *isw2/nhp10* mutant strain compared to the wt strain. (c,d) The replication duration of some of *S.cerevisiae* chromosomes, e.g. chr. 13 (shown in panel c) and 15 (panel d), in the mutant strain is well predicted by the bottleneck estimate rather than EVD estimate. Data of MMS (DNA alkylating agent methyl methanesulfonate) treated wild-type and *isw2 nhp10* mutant strains of *S.cerevisiae* from ref. [4]. Origin locations are obtained from the literature [2]. Origins with zero firing rate from the fit were deleted in the statistics on inter-origin distances and origin strengths.



Supplementary Figure S16: The model predicts well replication timing profiles of *S.cerevisiae isw2 nhp10* mutant (mnt) strains. The plots refer to chromosome 4 as an example. Dots correspond to experimental data of MMS-treated wild-type and *isw2 nhp10* mutant strains of *S.cerevisiae* from ref. [4], and lines indicates a model fit or prediction (p). Different dot colors correspond to different times. (a) Model fit of replication timing profiles of chromosome 4 of the wt strain. Origin locations from ref. [2] were used in this fit. (b) Model fit of replication timing profiles of chromosome 4 of the *isw2 nhp10* mutant strain. (c) Model prediction of mutant replication timing profiles based on the best-fit parameters from the wt data. The model parameters correspond to best-fit values of γ and origin strengths from the wt data. For the prediction, all origin rates from the wt best fit were multiplied by an adjusted global constant factor (about 1/8), and fork speed and replication initial time were taken from the fit of mutant data.

Supplementary Table S1: The parameters for genomes and chromosomes of *S.cerevisiae*, *L.kluyveri* and *S.pombe* from the best fit of genome-wide time-course replication data with the model.

Parameters for *S.cerevisiae (SC)*, *L.kluyveri* and *S.pombe* genomes

Species*	$\gamma^\dagger$	$v^\dagger$ (kb/min)	$T_0^\dagger$ (min)	$\langle d\rangle^\ddagger$ (kb)	$\sigma_d^\ddagger$ (kb)	$\langle\lambda\rangle^\ddagger$ ($\mathrm{min}^{-\gamma-1}$)	$\sigma_\lambda^\ddagger$ ($\mathrm{min}^{-\gamma-1}$)
SC wt1	1.5	1.8	1.3	26.1	16.9	5.3×10^{-4}	4.5×10^{-4}
SC mut1	1.5	2.0	5.0	26.2	17.2	1.9×10^{-4}	2.1×10^{-4}
SC wt2	0.25	0.84	-13	37.3	22.9	3.1×10^{-3}	2.0×10^{-3}
SC mut2	0.75	0.27	-161	85.4	64.3	8.4×10^{-5}	5.6×10^{-5}
L.kluyveri	1.75	2.5	72.5	47.0	24.6	9.2×10^{-5}	6.2×10^{-5}
S.pombe	2.0	2.55	20.1	45.0	27.9	5.9×10^{-6}	3.7×10^{-6}

Parameters for *S.cerevisiae* wt1 chromosomes 1-8 (c1-c8)

Parameter	c1	c2	c3	c4	c5	c6	c7	c8
n§	12	34	15	51	20	13	41	26
$\langle d\rangle$ (kb)	19.3	23.6	20.0	30.0	28.1	20.8	26.2	21.3
σ_d (kb)	12.8	13.6	13.2	20.0	13.5	12.3	18.0	14.8
$\langle\lambda\rangle$ ($\times10^{-4}\,\mathrm{min}^{-2.5}$)	4.5	4.3	5.0	6.0	6.2	4.7	5.0	4.8
σ_λ ($\times10^{-4}\,\mathrm{min}^{-2.5}$)	4.1	3.1	5.9	4.5	5.5	6.5	4.5	4.1

Parameters for *S.cerevisiae* wt1 chromosomes 9-16 (c9-c16)

Parameter	c9	c10	c11	c12	c13	c14	c15	c16
n	20	26	24	40	35	25	40	37
$\langle d\rangle$ (kb)	22.2	29.1	27.7	26.9	26.1	31.3	27.0	25.5
σ_d (kb)	16.4	17.7	17.3	20.5	15.6	16.0	19.2	15.0
$\langle\lambda\rangle$ ($\times10^{-4}\,\mathrm{min}^{-2.5}$)	4.9	5.4	5.7	5.5	5.6	5.3	5.5	5.2
σ_λ ($\times10^{-4}\,\mathrm{min}^{-2.5}$)	3.4	4.5	4.3	6.0	4.3	4.4	4.4	3.8

Parameters for *S.cerevisiae* mut1 chromosomes 1-8 (c1-c8)

Parameter	c1	c2	c3	c4	c5	c6	c7	c8
n	12	34	15	51	20	12	40	26
$\langle d\rangle$ (kb)	19.4	23.6	20.0	29.7	28.1	22.4	26.8	21.1
σ_d (kb)	12.7	13.7	13.3	20.0	13.5	17.1	19.2	14.8
$\langle\lambda\rangle$ ($\times10^{-4}\,\mathrm{min}^{-2.5}$)	1.7	1.5	2.8	2.0	2.7	2.2	1.8	1.7
σ_λ ($\times10^{-4}\,\mathrm{min}^{-2.5}$)	1.8	1.7	3.4	2.1	2.9	4.3	1.7	1.5

Continued on next page.

Supplementary Table S1: *Continued from previous page*

Parameters for *S.cerevisiae* mut1 chromosomes 9-16 (c9-c16)

Parameter	c9	c10	c11	c12	c13	c14	c15	c16
n	20	25	24	40	35	25	40	37
$\langle d \rangle$ (kb)	22.2	30.1	27.7	26.8	26.0	31.2	26.9	25.5
σ_d (kb)	16.4	18.8	17.3	20.2	15.7	16.0	19.1	15.0
$\langle \lambda \rangle$ ($\times 10^{-4}\,\mathrm{min}^{-2.5}$)	2.2	2.2	1.7	1.9	1.9	1.7	2.1	1.5
σ_λ ($\times 10^{-4}\,\mathrm{min}^{-2.5}$)	1.7	2.5	1.2	2.5	1.9	1.8	2.1	1.3

Parameters for *S.cerevisiae* wt2 chromosomes 1-8 (c1-c8) treated with MMS

Parameter	c1	c2	c3	c4	c5	c6	c7	c8
n	6	20	10	36	17	7	32	15
$\langle d \rangle$ (kb)	41.0	40.6	33.4	41.7	32.8	41.3	33.5	35.6
σ_d (kb)	40.4	26.1	19.2	25.4	16.2	35.3	18.3	20.2
$\langle \lambda \rangle$ ($\times 10^{-3}\,\mathrm{min}^{-1.25}$)	3.4	3.0	3.6	2.7	3.3	4.4	3.0	2.8
σ_λ ($\times 10^{-3}\,\mathrm{min}^{-1.25}$)	1.7	1.5	2.0	1.6	2.1	2.9	2.2	1.3

Parameters for *S.cerevisiae* wt2 chromosomes 9-16 (c9-c16) treated with MMS

Parameter	c9	c10	c11	c12	c13	c14	c15	c16
n	12	20	20	29	22	18	31	26
$\langle d \rangle$ (kb)	36.6	37.4	33.4	36.5	42.7	42.6	34.7	35.4
σ_d (kb)	23.6	21.8	19.4	25.8	26.5	19.1	21.2	22.8
$\langle \lambda \rangle$ ($\times 10^{-3}\,\mathrm{min}^{-1.25}$)	3.7	3.5	2.6	3.2	3.6	3.0	2.9	2.8
σ_λ ($\times 10^{-3}\,\mathrm{min}^{-1.25}$)	3.0	2.9	2.0	1.7	1.8	2.1	1.9	1.7

Parameters for *S.cerevisiae* mut2 chromosomes 1-8 (c1-c8) treated with MMS

Parameter	c1	c2	c3	c4	c5	c6	c7	c8
n	4	10	6	20	7	3	13	6
$\langle d \rangle$ (kb)	63.8	80.6	51.3	75.5	80.5	119.0	83.1	93.3
σ_d (kb)	51.5	54.4	28.7	70.1	42.6	114.0	50.4	61.0
$\langle \lambda \rangle$ ($\times 10^{-5}\,\mathrm{min}^{-1.75}$)	5.2	9.2	9.4	5.2	10.0	12.2	9.1	7.9
σ_λ ($\times 10^{-5}\,\mathrm{min}^{-1.75}$)	3.9	6.1	9.6	4.7	4.5	8.7	4.8	5.0

Continued on next page.

Supplementary Table S1: *Continued from previous page*

Parameters for *S.cerevisiae* mut2 chromosomes 9-16 (c9-c16) treated with MMS

Parameter	c9	c10	c11	c12	c13	c14	c15	c16
n	4	9	6	14	9	5	15	13
$\langle d \rangle$ (kb)	114.0	87.9	108.0	78.5	99.3	166.0	75.9	71.2
σ_d (kb)	63.4	50.2	58.7	67.9	79.4	117.0	67.5	36.3
$\langle \lambda \rangle$ ($\times 10^{-5}\,\mathrm{min}^{-1.75}$)	16.4	10.5	8.9	8.1	10.1	10.2	6.9	7.2
σ_λ ($\times 10^{-5}\,\mathrm{min}^{-1.75}$)	5.1	6.3	4.2	4.3	5.7	7.4	4.6	4.9

Parameters for *L.kluyveri* chromosomes 1-8 (c1-c8)

Parameter	c1	c2	c3	c4	c5	c6	c7	c8
n	24	25	27	27	30	31	43	39
$\langle d \rangle$ (kb)	42.6	44.9	46.9	49.1	43.9	44.9	41.3	59.8
σ_d (kb)	29.0	22.6	18.0	24.4	14.1	21.1	22.0	34.4
$\langle \lambda \rangle$ ($\times 10^{-5}\,\mathrm{min}^{-2.75}$)	7.4	9.8	11.6	9.7	7.0	8.9	7.9	11.4
σ_λ ($\times 10^{-5}\,\mathrm{min}^{-2.75}$)	4.6	7.1	5.7	7.1	5.1	5.3	5.9	6.8

Parameters for *S.pombe* chromosomes 1-3 (c1-c3)

Parameter	c1	c2	c3
n	125	107	52
$\langle d \rangle$ (kb)	45.1	43.5	47.4
σ_d (kb)	26.9	31.0	23.3
$\langle \lambda \rangle$ ($\times 10^{-6}\,\mathrm{min}^{-3}$)	5.5	5.0	8.5
σ_λ ($\times 10^{-6}\,\mathrm{min}^{-3}$)	3.2	3.1	4.7

* *SC* wt1 and *SC* mut1 are the wide-type and origin mutant strains of *S.cerevisiae* respectively from Hawkins et al. [2]. *SC* wt2 and *SC* mut2 are the wide-type and *isw2nhp10* mutant strains of *S.cerevisiae* respectively from vincent et al. [4].

† global parameters

‡ statistics of local parameters (inter-origin distances and origin strengths).

§ origin numbers of *L.kluyveri*, *S.cerevisiae* and *S.pombe* are from Agier et al. [1], Hawkins et al. [2] and Heichinger et al. [3] respectively. As for *S.cerevisiae* origin mutant, three inactivated origins were deleted from the origin list. For *S.cerevisiae isw2nhp10* mutant, origins with zero strengths were removed.