ORIGINAL PAPER

A model for the spatiotemporal organization of DNA replication in *Saccharomyces cerevisiae*

T. W. Spiesser · E. Klipp · Matteo Barberis

Abstract DNA replication in eukaryotes is considered to proceed according to a precise program in which each chromosomal region is duplicated in a defined temporal order. However, recent studies reveal an intrinsic temporal disorder in the replication of yeast chromosome VI. Here we provide a model of the chromosomal duplication to study the temporal sequence of origin activation in budding yeast. The model comprises four parameters that influence the DNA replication system: the lengths of the chromosomes, the explicit chromosomal positions for all replication origins as well as their distinct initiation times and the replication fork migration rate. The designed model is able to reproduce the available experimental data in form of replication profiles. The dynamics of DNA replication was monitored during simulations of wild type and randomly perturbed replication conditions. Severe loss of origin function showed only little influence on the replication dynamics, so systematic deletions of origins (or loss of efficiency) were simulated to provide predictions to be tested experimentally. The simulations provide new insights into the complex system of DNA replication, showing that the system is robust to perturbation, and giving hints about the influence of a possible disordered firing.

T. W. Spiesser · E. Klipp · M. Barberis (✉)
Computational Systems Biology,
Max Planck Institute for Molecular Genetics,
Ihnestraße 73, 14195 Berlin, Germany
e-mail: barberis@molgen.mpg.de

Introduction

The budding yeast *Saccharomyces cerevisiae* is a model organism used to study fundamental processes relevant to all life forms (Menacho-Marquez and Murguia 2007). Some of these processes are affected more frequently than others by genetic and epigenetic alterations in cancer. One of them is the essential action of copying all the information of an organism in the form of its deoxyribonucleic acid (DNA) to ensure the maintenance of genomic integrity. The genomic duplication requires a complex coordination of successive events to initiate DNA replication and to distribute fully replicated chromosomes into the daughter cells (Bell and Dutta 2002; Diffley and Labib 2002). The initiation of DNA replication temporally stretches from the Mitosis phase (M phase) over the Gap1 phase (G1 phase) into the early Synthesis phase (S phase). However, the chromosomal duplication is confined to the S phase of the cell cycle. Successful replication requires that the entire genome of an organism is duplicated without errors in a timely fashion only once per cell cycle. Therefore, DNA replication has evolved into a tightly regulated process, involving the coordinated action of numerous factors.

In prokaryotes, replication starts from a single well-defined site and proceeds with a speed of up to 500 nucleotides per minute until it terminates at the end of the genome. This mechanism leads to a homogeneous replication pattern that is identical in every cell cycle. The genome of *S. cerevisiae* consists of 16 chromosomes, spanning a total length of about 13.5 million base pairs (bp) and if the replication machinery were to use the same

single site strategy, DNA replication would take several days to complete. On account of this, replication of eukaryotic genomes initiates from multiple discrete sites distributed over the chromosomes, so called origins of replication. During the G1 phase of the cell cycle, replication origins are prepared to fire, a process that is referred to as origin licensing (Weinreich et al. 2004), and the density of active replication origins in the chromosomes of eukaryotic cells determines S phase dynamics and chromosome stability during mitosis (Bielinsky 2003). In *S. cerevisiae*, a direct correlation between the length of S phase and the number of the replication origins has been demonstrated (van Brabant et al. 2001). Not all replication origins are initiated with an equivalent efficiency and eventually only a specific selection of them is destined to fire (Shirahige et al. 1993). Furthermore, it has been demonstrated recently that there is a hierarchy of preferential initiation of origins that correlates with local transcription patterns (Donato et al. 2006).

Experimental and computational studies have identified and mapped over 700 potential origin function target sites on the genome of *S. cerevisiae* (Feng et al. 2006; Nieduszynski et al. 2006; Raghuraman et al. 2001; Wyrick et al. 2001; Xu et al. 2006; Yabuki et al. 2007). A number of studies have suggested that yeast chromosomes contain early and late replicating domains and exhibit replication timing profiles that are consistent with a highly regulated chronological program (Nieduszynski et al. 2006; Yabuki et al. 2007; McCune et al. 2008), which is reproducible even under altered conditions (Alvino et al. 2007). These nearly homogeneous replication kinetics favour the argument that, in budding yeast, the origins of replication fire according to a temporal program, as it has been reported for bacterial replication (Jacob and Brenner 1963). However, recent studies have revealed an intrinsic temporal disorder in the replication of yeast chromosome VI (Czajkowsky et al. 2008), suggesting that there is no obligate order of origin firing and that the observed temporal pattern of replication could be explained largely by variable properties of origin firing without the need to invoke temporal staggering of initiations at different origins. This stochastic component is indeed contained in the replication process for its distant cousin fission yeast (Patel et al. 2006). This observation would place budding yeast yet closer to the other eukaryotes, where it has been considered to be rather the exception in the general organization of eukaryotic replication (Rhind 2006). Therefore, even though intensively studied, the spatiotemporal organization of the selective origin activation in *S. cerevisiae* remains unclear.

Every origin within the yeast genome can be characterized by specific properties: location in the chromosome, initiation time of firing, emanating fork rate (replication speed), efficiency of firing. Chromosomal positions and firing times for a certain number of origins have been reported (Nieduszynski et al. 2007), and fork rate values are available (Rivin and Fangman 1980; Raghuraman et al. 2001; Yabuki et al. 2007). However, only few data are available about individual origin efficiencies (Yamashita et al. 1997), which refer to the frequency at which an origin initiates DNA replication (fires) within a population of cells.

In this work we provide a deterministic model for the DNA replication dynamics, based upon four replication parameters, to study the temporal sequence of origin activation in *S. cerevisiae*. The parameters are the length of the chromosomes, the positions of the origins, the initiation firing times of the origins and the replication fork migration rates. Single origin efficiencies, the fifth major parameter influencing the replication process, is not included in the model as an adjustable parameter, but is implicitly incorporated. The model of the DNA replication is validated via its ability to reproduce experimental data in the form of replication profiles. We continuously monitor the dynamics of the chromosomal duplication during simulations of wild type and perturbed replication conditions. Furthermore, we perform simulations of systematic origin deletion in order to provide predictions, which could be tested experimentally.

This work aims at amplifying the knowledge and further understanding of the mathematically poorly elucidated DNA replication process in budding yeast. Understanding DNA replication in *S. cerevisiae* is not a trivial goal. Due to the high degree of conservation of the replication machinery, the study of replication in this model organism accounts for nearly all life forms and must not be seen as an isolated process, but rather as one step towards the understanding of a crucial event, whose deregulation is often fatal and can lead to severe genetic disease in humans, like cystic fibrosis or cancer.

Materials and methods

Model characteristics and available data

1. *DNA units*. In the model, a DNA unit (u) is defined as a 500 bp block of DNA. Hence, in the simulation each chromosome is composed of a series of DNA units, corresponding to its original size (L_{org}) divided by 500 to yield the internal resolution size L_{res}. To acknowledge the correct size of the chromosomes, L_{res} is always rounded up. The size of the DNA units (500 bp) defines the resolution of the simulation. The size of the chromosomes was obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG)

(Kanehisa and Goto 2000; Kanehisa et al. 2006; Kanehisa et al. 2008).

2. *Origin location.* The location of the replication origins on the chromosomes is sequentially pre-determined (Newlon and Theis 1993). An 11 bp region, the autonomous replicating sequence (ARS) consensus sequence (ACS), can be found within every 200 bp sequence that exhibits origin activity in the budding yeast (Theis and Newlon 1997). The chromosomal locations of the replication origins can be found in the *S. cerevisiae* OriDB database, version 1.1.1 (Nieduszynski et al. 2007).

3. *Origin initiation.* Initiation times have been assessed for replication origins (Raghuraman et al. 2001; Yabuki et al. 2007). They are assembled in the *S. cerevisiae* OriDB, version 1.1.1 database (Nieduszynski et al. 2007) as well. In this work we consider the initiation times provided by a heavy:light (HL) timing study (Raghuraman et al. 2001).

4. *Fork migration rate.* The replication bubble grows bi-directionally and both replication forks migrate at a certain rate (v). According to the data reported in Raghuraman et al. 2001, fork rates range from 0.5 to 11 kb/min, with a mean of 2.9 kb/min and a median of 2.3 kb/min. Similar mean values were obtained in different studies: 2.8 ± 0.1 kb/min (Yabuki et al. 2007) and 3.7 kb/min (Rivin and Fangman 1980). In this model we assume that the forks migrate constantly throughout S phase at an approximate rate of 3 kb/min.

The *S. cerevisiae* OriDB, version 1.1.1 database (Nieduszynski et al. 2007) contains 732 replication origins target sites, approximatively 60% (454) of which are considered in this work. The selection is based on the availability of both chromosomal location and firing time (derived from the HL analysis) for every replication origin. A complete list of the replication origins, the location on the chromosomes and the firing times used in this work are reported in the electronic supplementary material, Table S1.

The spatiotemporal model

Figure 1 illustrates the model and its parametrization. As described above, the DNA is divided into units of equal length (500 bp). A two-dimensional array element (A) of size L_{res} is assigned to every chromosome. Additionally, two DNA units are added to A, introducing artificial boundaries, accounting for the left (A_0) and right (A_{Lres+1}) end of the chromosomes. The array element A contains all discrete DNA unit positions ($A_{(0:Lres+1)}$) and the status of the replication for the position. This is represented by a Boolean Variable, which is set 'FALSE' by default indicating that the DNA has not been replicated at this position
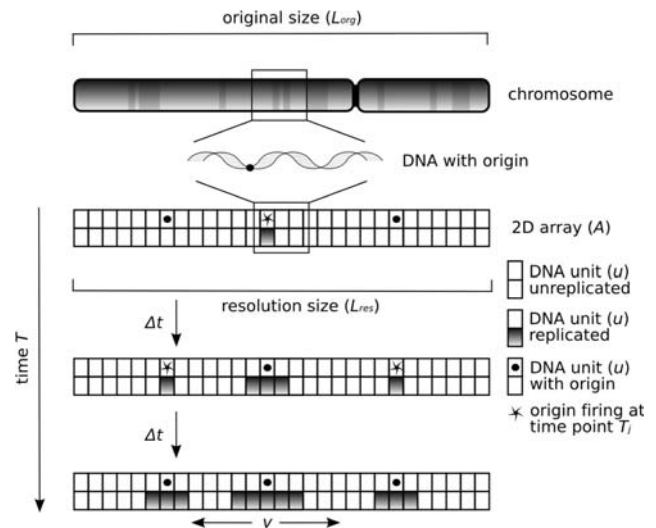

Fig. 1 Scheme of the chromosomal duplication model and its parametrization. The features and the algorithm are explained in the main text

yet, and set 'TRUE' only at the end positions of the chromosomes. Another two-dimensional array element (O) stores origin information: origin name, origin position on the virtual chromosome A, origin activation time in seconds and the origin activation status, a Boolean Variable, set 'FALSE' by default, indicating that the origin has not been activated yet. A variable T represents the replication time.

T is the sum of all discrete time steps t_i, with ($i = 1{:}n$)

$$T = \sum_{n=1}^{n} t_i \tag{1}$$

where n is the number of discrete time steps needed to complete DNA replication. One time step equals the time (Δt), that the replication fork needs to go through one DNA unit (Δu), hence

$$\Delta t = \frac{\Delta u}{\Delta v} \tag{2}$$

where $\Delta u = 500$ bp and $\Delta v = 3{,}000$ bp/min and therefore

$$\Delta t = \frac{500 \, \text{bp}}{3{,}000 \, \text{bp/ min}} = \frac{1}{6} \text{min} = 10 \text{s}. \tag{3}$$

The variable T_j, with $j \in (1, n)$ specifies the replication time at every discrete time point during the simulation. An algorithm for the DNA replication has been implemented as follows. At every time point T_j the program reviews the array O to find the origins that initiate at that time. If found, the Boolean Variables for these origins in O are set to 'TRUE', indicating that they have fired and cannot do so again. Furthermore, the Boolean Variables in A at the origins positions (e.g. A_{ori1} and A_{ori2}) are set 'TRUE' as well, indicating that these regions now have been replicated. For simplicity, the activation of origins is
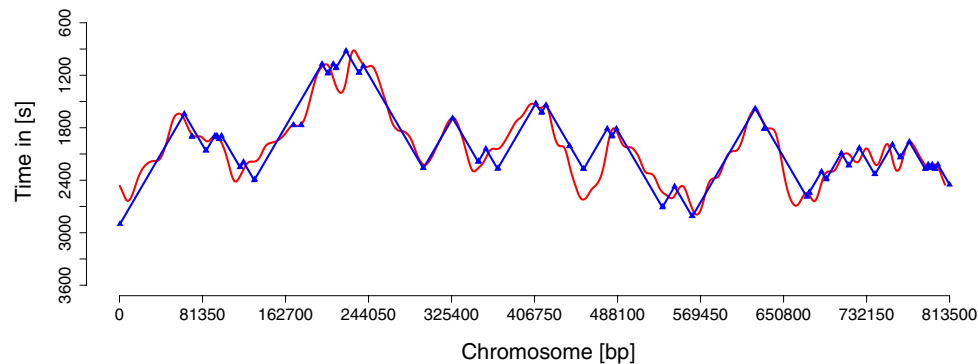
Fig. 2 Replication profiles of chromosome II. The smooth curve is recalculated according to the microarray-based heavy:light data from Raghuraman et al. 2001, whereas the straight curve represents the simulated profile obtained with the spatiotemporal model. The replication time in seconds is plotted as a function of chromosome coordinate in base pairs (bp)

assumed to occur at the beginning of the time steps, for which reason a unit is either replicated completely or not at all. The discretization error introduced by this approximation decreases with the DNA unit size. Every origin issues two replication forks upon activation, each traveling in opposite directions in the course of the chromosomal duplication. Therefore, at time point T_{j+1} the program checks if the positions left and right of a replicated region (e.g. A_{ori1-1}, A_{ori1+1} and A_{ori2-1}, A_{ori2+1}) have not been replicated (set 'FALSE') yet, and if so, sets the Boolean Variable to 'TRUE'. In this manner the replication forks migrate in both directions, until they meet either the end of the chromosome, or a region that has already been replicated. Every position of every replication fork is stored at every time point of the simulation. The way of every replication fork through the genome during the simulation can be retraced and their final positions and times can be observed. The simulation stops once the whole chromosome is replicated.

Replication profile data

Experimental replication profiles, which can be found in the literature (Raghuraman et al. 2001) are used to assess the model performance. The profiles are derived from a microarray based HL timing study. After growth in an isotopically dense culture medium, cells are released into S phase (after α-factor-induced G1 phase arrest), and replicated (HL) DNAs and unreplicated [heavy:heavy (HH)] DNAs are isolated from samples collected at 10, 14, 19, 25, 33, 44 and 60 min (Raghuraman et al. 2001). Replication profiles for all chromosomes and the corresponding data in tabular form can be found in the electronic supplementary material, where the original data were used to recalculate the replication profiles. Figure 2 shows the replication profile of chromosome II as a showcase. For all recalculated replication profiles see electronic supplementary material, Fig. S1. Furthermore, the data were used to

calculate the total replication time for all chromosomes. Subtraction of the highest peak from the lowest valley yields the total replication time. It should be noted at this point that the authors (Raghuraman et al. 2001) deleted regions of low probe density from their replication profiles. However, these regions are still consistent in their corresponding data. Therefore, these regions appear in the recalculated profiles as large artifacts, as well as they extend the calculated total replication time [see electronic supplementary material, Fig. S1 (l)].

Software

The spatiotemporal model has been implemented using the programming language Python (van Rossum 1995).

Results

Generation of the replication profiles

The spatiotemporal organization of the DNA replication process can be visualized by means of replication profiles. A replication profile is the plot of the replication time as a function of the position in the chromosome. In a replication profile peaks correspond to origins of replication, and valleys correspond to termination zones. The earlier an origin fires, the taller is its respective peak within the profile. Shoulders along the lines connecting peaks and valleys can either result from timely collisions of a firing origin and an oncoming replication fork, or they could also be the result of change in the fork migration rate, or inefficient origins. The slope of the line connecting a peak and a valley gives the direction and rate of the fork migration.

The simulation of the chromosomal duplication has been performed, as described in "Materials and methods" with a fork rate value equal to 3 kb/min. Sixteen replication

profiles were generated, one for each chromosome, in order to highlight the spatiotemporal organization of the simulated DNA replication. Figure 2 shows the replication profiles for chromosome II. The smooth curve is recalculated from the data provided by Raghuraman et al. 2001, as described in "Materials and methods", and the straight curve shows the simulated profile. All essential features of the experimental profile were captured in the simulation.

However, we observed a deviation in the slope of the lines, representing the speed of the fork migration. The lines of the simulated curve are straight, for a constant migration rate is implemented, whereas the experimental curve is smooth with a varying slope, indicating different fork rates. Most simulated regions reflect experimental data with high accuracy and only few regions with lower accuracy. We found similar results for all 16 chromosomes (see electronic supplementary material, Fig. S1). As reported in the work of Raghuraman et al. 2001, the fork rates range from 0.5 to 11 kb/min with a mean of 2.9 kb/min. Changes (increase or decrease) in the value of the fork rate could lead to different results in the computed simulations, implying more precise results in some regions and less accuracy in other regions. In addition, it is likely that, for some inefficient origins, the direction of fork migration during DNA synthesis may change from one cell division to the next. Moreover, it has been shown in mammalian cells that the replication speed controls the choice of the initiation firing sites on the chromosome (Courbet et al. 2008). However, we aim at a simplifying parametrization for this still not well-defined process to create an accurate, yet comprehensive representation.

We model the chromosome duplication deterministically using the published data for locations and firing times of 454 origins of replication. Since only few data are available about origin firing efficiency (Yamashita et al. 1997), which is nonetheless known to be a key property of the origin activation, we included origin efficiencies in an implicit way. We regarded the efficiencies of a subset of all origins (454 out of 732 reported in the OriDB) as to be 100%, which is a strong assumption. However, an approximation of the replication with 454 origins that fire with an efficiency equal to 100% represents a single replication event in a cell with 732 origins that fire at about 60% average efficiency. Since the number of actively engaged origins per cell cycle has been reported to be roughly around 400 (Wyrick et al. 2001; Takeda and Dutta 2005), this approximation seems reasonable. Employing this approach, the model does not represent a single cell behavior per se (no intrinsic noise in efficiencies and firing times) but reflects the average of a cell population. In other words, the model stands for a likely replication event in the average single cell, because it has been parametrized with population averaged data.

Chromosome duplication in the *clb5Δ* mutant

The activation of the replication machinery has still to be highlighted in many of its regulatory events, but a relevant step is the phosphorylation of different substrates by the Cdk1–Clb5,6 kinase complex that induces the firing of the DNA replication origins (Bell and Dutta 2002; Takeda and Dutta 2005). In a recent work, we described the steps which lead to the firing of DNA replication origins with a simple probabilistic model that considers the availability of the Cdk1–Clb5,6 nuclear concentration as the main input (Barberis and Klipp 2007). This model provides an explanation for the replication status of specific mutants which influence the entry into S phase, pointing out the direct correlation between the Cdk1–Clb5 activity and the temporal activation of the replication origins (Barberis and Klipp 2007). In support of this, *clb5Δ* cells suffer a significant decrease in the firing efficiency of some origins, in particular for those classified as late-S phase origins (Donaldson et al. 1998). Clb6 activates instead the early replication origins (Donaldson et al. 1998).

In the work of McCune et al. 2008, the activation of the replication origins has been investigated, comparing the temporal program versus the disordered firing, analysing cells lacking the initiator factor of DNA replication Clb5. Therefore, we tested the model in the *clb5Δ* mutant. Operatively, we stopped origin firing at 1,645 s. The replication profile computed for the chromosome II in a *clb5Δ* mutant is reported in Fig. 3. We found that multiple zones suffer significant delays in replication, whilst others are unaffected. Interestingly, the delayed regions correspond to the so-called *CLB5*-dependent regions (CDRs) experimentally observed in the work of McCune et al. 2008. These regions match sequences of the genome which on average replicate late in S phase (Alvino et al. 2007; Raghuraman et al. 2001), and each of the late replication origins reported in the work of Donaldson et al. 1998 resides in CDR regions. The simulations of the *clb5Δ* mutant are reported in the electronic supplementary material, Fig. S2 (compare CDR regions with the experimental profiles in Fig. 4; McCune et al. 2008). In detail we found a perfect match for nine chromosomes (from I to VIII, and XI), a good fit in the majority of the sequence length for chromosomes IX, X and XIV, and a small or no match for chromosomes XII, XIII, XV and XVI.

This analysis is in agreement with the fact that the *clb5Δ* mutant only affects late origins, whereas the early origins fire normally. Therefore, the precise time at which origins stop to fire in absence of *CLB5* is important. We use 1,645 s as the time point, after which there is no more origin activation, because it represents the mean value of the distribution of the experimentally determined origin activation times (see electronic supplementary material,

Fig. 3 Replication profiles of chromosome II in a *clb5Δ* background. The *dotted line* represents the simulated profile for wild type cells, whereas the *straight one* represents the computed profile for the *clb5Δ* mutant
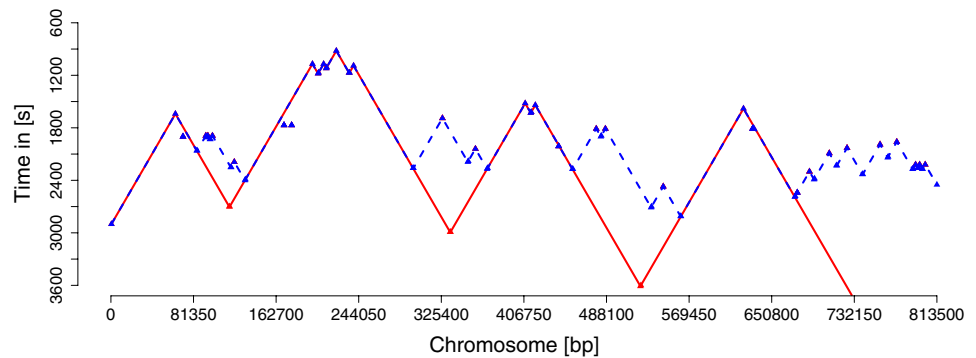
Fig. 4 Simulated replication kinetics of chromosome II. The simulations are performed for wild type (**a**) and perturbed conditions (**b**). In the case of perturbed conditions, the simulation has been performed considering 30 reduced sets of replication origins derived from the random deletion of 50% of the original origins
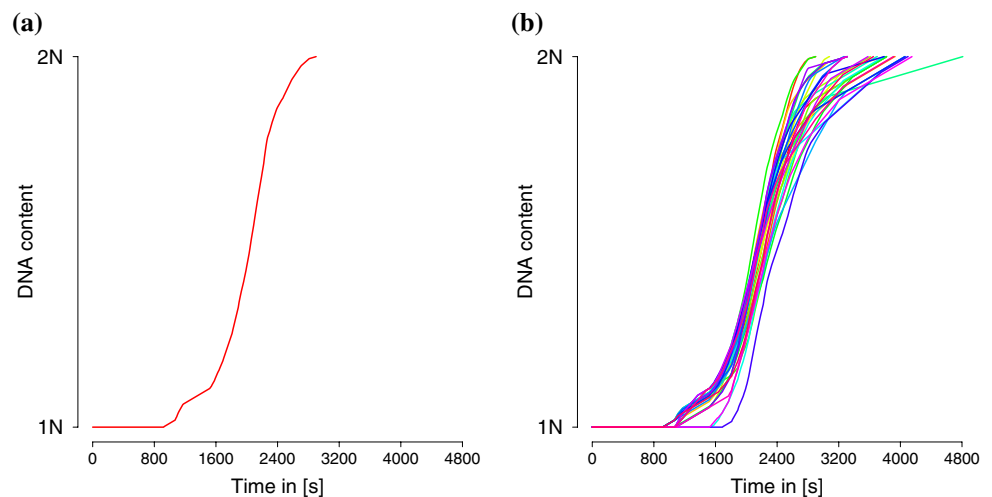
Fig. S3). Thus, the origins are divided in an early half (Clb5-unaffected) and in a late half (Clb5-affected). However, it is likely that Clb5 activates every origin not at the same time at every cell cycle, but with a certain variation. Intrinsic noise will affect the time of the activation of the Clb5-dependent origins that will become more like a time span (of some seconds or minutes). Therefore, the considered value of 1,645 is an approximation, which for some chromosomes might be quite accurate, but for others it might not be. This affects the results we observed in the following way: the chromosomes containing more early origins will be less sensitive to *CLB5* deletion, whereas the chromosomes with more late origins will be more sensitive.

The general agreement of the replication kinetics between wild type and *clb5Δ* in the computed and experimental profiles supports the temporal program of the origin activation in budding yeast, as predicted (McCune et al. 2008).

Impact of origin deletion on DNA replication

Saccharomyces cerevisiae has well-defined, site-specific origins, many of which are efficient and fire as many as 90% of S phases (Fangman and Brewer 1991; Newlon et al. 1991). These characteristics lead to nearly homogeneous

replication kinetics (Raghuraman et al. 2001). Despite the fact that DNA replication in budding yeast seems to follow a temporal program of origin activation, it has been reported that there is a stochastic component which can influence the process (Czajkowsky et al. 2008; McCune et al. 2008). In fact, the activation of some origins in the CDR regions, more closely fits a disordered, stochastic firing. They show no peak time of firing or are activated over a broad distribution of activation times in different cells in the population (McCune et al. 2008). In addition, it has been reported that variants of a stochastic firing model are compatible with a temporal staggered initiation of the replication origins in fission yeast (Lygeros et al. 2008; Rhind 2006).

In order to investigate the impact of change in the origin activation pattern on the replication dynamics, replication kinetics for all chromosomes have been computed repeatedly (30 times) with reduced sets of considered origins. The subsets are composed by random deletion of 50% of the original origins. This accounts for the change in environmental conditions (i.e. stress condition, checkpoint activation) or inefficient firing, which could reduce the global origin firing efficiency from 60 to 30%. Comparison of the replication kinetics for chromosome II exhibited under wild type (Fig. 4, left) and perturbed (Fig. 4, right)

conditions shows that a 50% deletion of replication origins
yields a prolonged chromosomal replication time. How-
ever, we do not observe fundamental alterations in the
general shape of the replication kinetics, which indicates
that conditional change leading to a 50% efficiency
reduction of origin firing does not change the replication
dynamics of the chromosomal duplication.

Moreover, we found that for most chromosomes the
replication kinetics seem to show a remarkable resistance
to origin reduction (see electronic supplementary material,
Figs. S4, S5). The chromosomal duplication initiates within
a short timeframe, which is consistent throughout the
replication process, and only disperses towards replication
termination. Concerning retardation, we found that 50% of
origin deletion leads on average to a circa 12 min delay in
duplication completion for chromosome II. The remaining
chromosome kinetics show similar results (see electronic
supplementary material, Figs. S4, S5). The outcome of the
random perturbation of the system shows that the replica-
tion process is robust against firing failure or efficiency
variation, and suggests that the replication kinetics dis-
played by a cell can be widely independent from the
temporal program of the origin activation.

Simulating a stepwise loss of origin function

Despite the contribution that multiple origins per chromo-
some may make to efficient genome duplication in
S. cerevisiae, it is widely accepted that there are many
more replication origins than needed for the timely repli-
cation during the S phase (Bielinsky 2003). In fact, several
origins on chromosome III can be deleted without sub-
stantially affecting the ability to faithfully inherit this
chromosome during cell division (Dershowitz and Newlon
1993; Dershowitz et al. 2007).

To further understand the relationship between origin
activation and replication time, we simulated the chromo-
somal replication with a decreasing number of active origins
and monitored the change of the replication time. In the
previous simulations we have observed that during pertur-
bation of the system, the replication kinetics for the
chromosomes are very similar, even though they are repli-
cated with different sets of origins. Therefore, we ignored
which specific selections of origins were used in the simu-
lations and thus studied the relationship between the number
of activated origins and the replication time directly. To this
purpose, we used the same chromosomal location for ori-
gins and the same firing times, only the activated origins
change randomly. The model predicts how the replication
time of the average replication event would change, if a
certain percentage of the origins were to be defective,
deleted or inefficient. It is difficult to investigate the direct
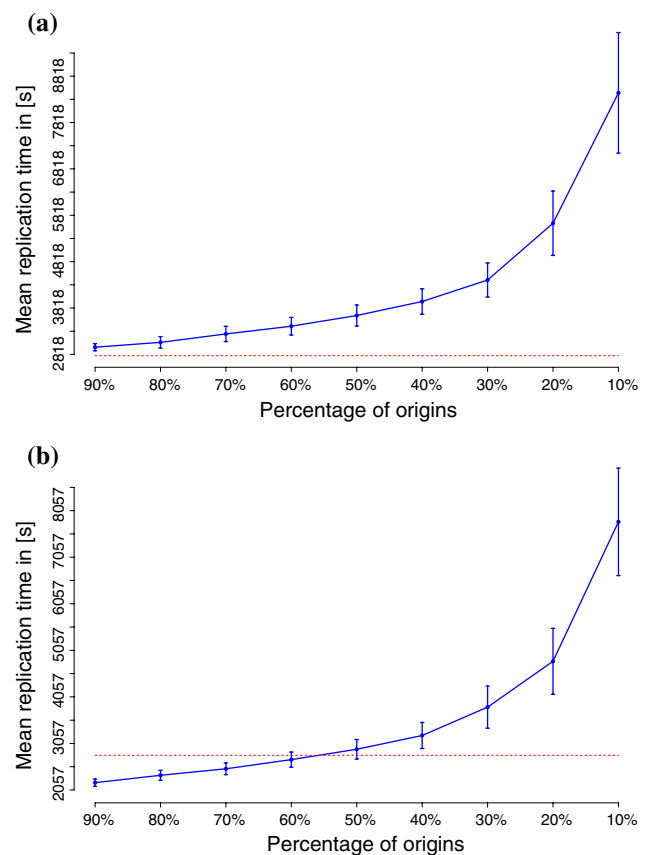effect of activated origins and replication time in living

Fig. 5 Mean replication time (in seconds) for chromosomes II (**a**)
and XVI (**b**). *Solid line* represents the curve for descending
percentage of the considered replication origins (from 90 to 10%).
Error bars show the standard deviation of 10,000 simulations.
Dashed line indicates the experimental replication time for each
chromosome, according to Raghuraman et al. 2001

systems, because the deletion of the origins often leads to
the activation of adjacent usually inefficient/dormant ori-
gins. This mechanism ensures to the cell the successful
chromosomal replication. Therefore, a systematic compu-
tational study is useful to highlight the relationship between
a controlled quantity of active origins and the replication
time.

Mean replication times for descending percentages of
active origins (from 90 to 10%) have been computed for all
chromosomes. The origin sets have been reduced stepwise
(10%) and randomly selected. The simulations for every
fraction of remaining origins were repeated 10,000 times.
Mean and standard deviation for every fraction of
remaining origins are displayed for every chromosome
(Fig. 5; electronic supplementary material, Fig. S6). The
average delay for 50% remaining origins is summarized in
Table 1. The calculations for the chromosome II show that,
with a decreasing percentage of remaining origins, the
mean replication time increases, as well as the standard
deviation (Fig. 5a). This is the case for all chromosomes,

Table 1 Average delay in chromosomal duplication time, under 50% origin deletion condition, calculated after 10,000 simulations of DNA replication

Chromosome	Average delay (50% origin deletion)	Active origins in % (crossing experiments/ simulations)
I	7 min 00 s	30–40
II	12 min 36 s	100
III	2 min 29 s	50–70
IV	18 min 54 s	70–90
V	15 min 29 s	90–100
VI	3 min 52 s	60–90
VII	12 min 37 s	100
VIII	9 min 53 s	40–50
IX	5 min 59 s	30–40
X	11 min 22 s	40–50
XI	13 min 30 s	100
XII	14 min 34 s	40–50
XIII	16 min 17 s	50–70
XIV	20 min 48 s	100
XV	17 min 08 s	100
XVI	14 min 01 s	50–60

The percentage of origins is indicated, which is required to simulate the chromosomal duplication in the experimentally measured time

although the intensity of the increase differs amongst the chromosomes. Interestingly, the experimentally assessed duplication times can be obtained using only a certain subset of activated origins, and the subsets are different for every chromosome and composed randomly. An example is reported for chromosome XVI (Fig. 5b). The experimental replication time, derived from Raghuraman et al. 2001, is indicated as a dashed line. The simulation shows that chromosome XVI duplication could be achieved, in the experimentally measured time, with subsets of only 50–60% randomly selected origins (Fig. 5b; Table 1), as indicated by the intersection of dashed line and solid curve. This percentage differs for every chromosome, and for some chromosomes the replication can only be simulated in the appropriate time with 100% of the origins, e.g. for chromosome II (Fig. 5a; Table 1). Furthermore, it is important to consider that inaccuracies within the experimental replication times (see "Materials and methods" for details) affect the estimates of origin subsets in a way that, where the experimental times should be smaller, the estimated subsets should be larger.

The simulations nicely mirror the robustness of the replication process against perturbations in origin firing, as a result of loss of the origin function or change in the total efficiency. Using a systems study, we highlight the relationship between origin activation and replication time in the average cell population in budding yeast. The reduction

in origin firing up to, e.g. 50% in chromosome II can be compensated within the system resulting in a delay of about 12 min in replication completion (Figs. 4, 5). This is the case obviously only if no other late/dormant origins fire. A similar effect can be observed for the remaining chromosomes (Table 1). The average delay in chromosomal duplication increases with the size of the chromosomes (Fig. 6a), and decreases with an increasing origin density on the chromosomes (Fig. 6b). The origin density is the ratio between the number of origins on a chromosome and the chromosome size.

Discussion

The goal of this work is to provide a model for the DNA replication dynamics, based on four replication system parameters, to study the temporal sequence of origin activation in *S. cerevisiae*. The system parameters are: (1) location of the origins on the chromosome, (2) firing time of the origins, (3) speed of the moving replication fork and (4) length of the chromosomes. The parameters used in the analysis were obtained from experimental data (see "Materials and methods" for details). In the spatiotemporal model of DNA replication, two limiting factors impinge the biological validity of the model: the approximation of the fork migration rate with the mean of the experimentally determined value of 2.9 kb/min (Raghuraman et al. 2001), and the implicit consideration of the origin efficiencies.

The model has been used to generate replication profiles, which plot replication time as function of the chromosome coordinate. They have been compared to the replication profiles reported in the literature (Raghuraman et al. 2001). The comparison has shown that the model is generally able to reproduce the experimental replication profiles (Fig. 2; electronic supplementary material, Fig. S1). Some disagreements between simulations and experiments can be observed, essentially due to two different reasons. First, using an approximated value for the fork migration rate, the rates of motion are constant and do not take into account changes in the speed. This results in small scale inaccuracies in the replication profiles. However, this does not explain large, but locally restricted aberrations in the profiles. Described artifacts in the experimentally produced replication profiles, which were deleted by the authors (Raghuraman et al. 2001) due to low probe density in the microarray can explain this phenomenon. We, therefore, conclude that the modeling performance is even more accurate than it appears at first sight, for no significant differences can be found once the described artifacts are ignored. Nevertheless, the accuracy of the model could perhaps be increased by consideration of a dynamic fork rate function. Different fork rates at different chromosome

Fig. 6 Average delay in chromosomal duplication time (in minutes) plotted over the length (**a**) and the origin density of the chromosomes (**b**). The average delay is calculated after 10,000 simulations of DNA replication under perturbed conditions (50% origin deletion). The origin density is the ratio of the number of origins on a chromosome and the chromosome size
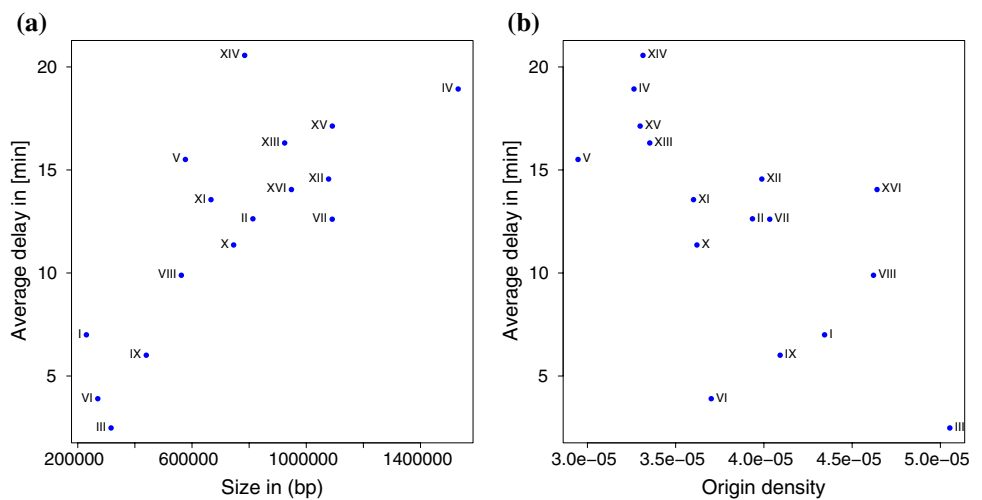
regions could have either regulatory functions or could be caused by higher order structures of the chromosome (protein binding, 3-D effects, etc.). Therefore, a rate function that is adapted to those different, biological characteristics influencing the migration rate, could enhance the performance.

We do not include single origin efficiencies as an adjustable parameter, which means leaving out a key property of the origins and, with it, its stochastic influence on the replication process. However, we based our modeling on the assumption that in one cell cycle there are about 400 origins that fire with the efficiency of 100%, when indeed there are much more origins (732) that could be potentially used. Thus, we approximated the overall efficiency of initiation in a cell with 732 origins at roughly 60%. Previous studies indicate that the excess of origins can help the cell to ensure the duplication under stressed conditions (Dershowitz and Newlon 1993; Dershowitz et al. 2007). This means that our modeling reflects DNA replication of a particular cell cycle and—due to the parametrization of the model with population averaged data—it represents the average DNA replication event in a budding yeast cell. These assumptions could be relaxed when more experimental data will become available.

S. cerevisiae has a 13.5 Mb genome distributed over 16 chromosomes, and therefore each single yeast chromosome is considerably smaller than the 4.6-Mb *E. coli* genome. Yet, yeast replication origins occur on average every 20-40 kb, a hundred times more densely distributed that one would predict by comparison to the *E. coli* genome. The difference in fork migration rates may explain in part the need for multiple replication origins per eukaryotic chromosome. DNA replication forks migrate at rates about 30 times slower in yeast compared to *E. coli*—fork migration rates of about 3 kb/min compared to about 100 kb/min (Raghuraman et al. 2001; Rivin and Fangman

1980). The use of multiple initiation events per chromosome probably compensates for slower fork migration rates in maintaining an efficient rate of genome duplication and S phase progression in eukaryotic cells. Based on the values discussed above, *S. cerevisiae* would need about 100 replication origins to duplicate its genome at a rate sufficient to accommodate its S phase, about four times less than the current estimates for origin numbers in this organism (Raghuraman et al. 2001; Wyrick et al. 2001). Therefore, for the purpose of genome duplication, yeast replication origins are redundant, and it is interesting to investigate the relation between the number of active origins and the replication time. We used the model to systematically study this relationship. To assess the impact of particular sets of origins on the replication time, we computed replication kinetics under wild type and perturbed conditions. The replication kinetics mirror the dynamic of the replication system and are therefore a useful tool to investigate the influence of conditional changes on the system. Perturbing the replication process by severe loss of the replication origin function due to their random deletion showed only little influence on the replication dynamics (Fig. 4). Therefore, we could neglect the effect of specific origin sets on the time of DNA replication and systematically deactivate an increasing number of origins. As expected, the analysis showed that the more origins that were deactivated, the more time was needed to complete the chromosomal duplication, but interestingly highlights that the experimentally assessed duplication times can be obtained using only a certain subset of activated origins (Fig. 5).

In the model, we implemented directed movement for the DNA polymerase. Therefore, we do not allow backward movement during our simulations and, thus, we argue that the anticipated relationship between distance and time is close to linear. However, this linear relationship is not

directly visible in our results since we monitor the mean replication time with respect to the removal of origins, which one could also interpret as a system with an increasing failure rate over time. The replication time is dependent on the longest distance that a replication fork covers, which is the maximum value of the inter-origin spacing (extreme value of the distance between the origins). Successive removal of origins from the chromosome results in longer distances between the remaining origins. If we interpret this system as one with an increasing failure rate over time, we could describe this system with an extreme value distribution (EVD), being in our case the distance between the origins. However, we can only describe our results to a certain extend by such an EVD, because naturally the firing times influence the system as well. Normally distributed firing times (electronic supplementary material, Fig. S3) lead to exponentially distributed waiting times, and this effect smoothens the curve that we obtain.

The analysis showed that the replication system is robust against perturbations. This suggests that a purely deterministic program of the origin activation in budding yeast might be enough only at the first glance on the system, but possibly not to describe all of its properties. If a temporal program is influenced by stochastic patterns, we would expect the replication system to cope more easily with perturbations, and therefore to successfully complete DNA replication with hardly any substantial changes in the dynamics of the replication. Where in the purely deterministic system the defects in origin firing due to a perturbation would be more severe (i.e. stress condition, origin deletion, inactivation of some specific initiation factor which stimulate origins activation), a stochastic component would always provoke some random activation of origins. Hence, a stochastic influence, increasing its robustness can be advantageous for the system.

Moreover, we found that the length of a chromosome and its origin density have an impact on the robustness. In fact, the replication delay under perturbed conditions is increased for larger chromosomes, whereas the average delay is decreased for the chromosomes that have a higher origin density (Fig. 6). Consequentially, the increase in the delay could be interpreted as a decrease of robustness and the decrease in the delay could be seen as an increase in the robustness. Altogether, this suggests that smaller chromosomes with higher origin density are more robust towards perturbation. It is tempting to speculate that this could be an explanation for why organisms have evolved to rather have a number of smaller chromosomes, instead of only a large one. In any case, it seems favourable for an organism to possess a high number of origins, a selection of which is finally activated to duplicate the DNA within the required timeframe.

In conclusion, we have successfully constructed a simple, yet accurate deterministic spatiotemporal model for DNA replication in budding yeast, which reproduces the trends exhibited during chromosomal duplication. The results of our analysis suggest that the replication system is robust against perturbations, and that there might be a stochastic component in the temporal activation of the replication origins, especially under perturbed conditions. The observed robustness could be tested experimentally by deleting origins progressively and evaluating the replication time for each chromosome. Our future goal would be to investigate the influence of stochasticity on the temporal program of origin activation in budding yeast more closely. Noteworthy, a partially deterministic and partially stochastic order of DNA replication was already addressed in a model for DNA replication in mammalian cells (Takahashi 1987). In the light of this evidence, our model could well be suitable for further and more accurate investigation of the temporal origin activation in budding yeast, in particular as soon as experimental data concerning origin efficiencies will become available. Moreover, the computational analysis could be extended to eventually link DNA replication to the classical cell cycle machinery and its relevant checkpoints.

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 27:6396–6406

Barberis M, Klipp E (2007) Insights into the network controlling the G(1)/S transition in budding yeast. Genome Inform 18:85–99

Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. Annu Rev Biochem 71:333–374

Bielinsky AK (2003) Replication origins: why do we need so many? Cell Cycle 2:307–309

Courbet S, Gay S, Arnoult N, Wronka G, Anglana G, Brison O, Debatisse M (2008) Replication fork movement sets chromatin loop size and origin choice in mammalian cells. Nature 455:557–560

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

Dershowitz A, Newlon CS (1993) The effect on chromosome stability of deleting replication origins. Mol Cell Biol 13:391–398

Dershowitz A, Snyder M, Sbia M, Skurnick JH, Ong LY, Newlon CS (2007) Linear derivatives of *Saccharomyces cerevisiae* chromosome III can be maintained in the absence of autonomously replicating sequence elements. Mol Cell Biol 27:4652–4663

Diffley JFX, Labib K (2002) The chromosome replication cycle. J Cell Sci 115:869–872

Donaldson AD, Raghuraman MK, Friedman KL, Cross FR, Brewer BJ, Fangman WL (1998) CLB5-dependent activation of late replication origins in *S. cerevisiae*. Mol Cell 2:173–182

Donato JJ, Chung SC, Tye BK (2006) Genome-wide hierarchy of replication origin usage in *Saccharomyces cerevisiae*. PLoS Genet 2(9):e141. doi:10.137/journal.pgen.0020141

Fangman WL, Brewer BJ (1991) Activation of replication origins within yeast chromosomes. Annu Rev Cell Biol 7:375–402

Feng W, Collingwood D, Boeck ME, 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

Jacob F, Brenner S (1963) On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon. C R Hebd Seances Acad Sci 256:298–300

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34:D354–D357

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res 36:D480–D484

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: genome-wide replication in clb5Δ *Saccharomyces cerevisiae*. Genetics. doi:10.1534/genetics.108.094359

Menacho-Marquez M, Murguia JR (2007) Yeast on drugs: *Saccharomyces cerevisiae* as a tool for anticancer drug research. Clin Transl Oncol 9:221–228

Newlon CS, Lipchitz LR, Collins I, Deshpande A, Devenish RJ, Green RP, Klein HL, Palzkill TG, Ren RB, Synn S, Woody ST (1991) Analysis of a circular derivative of *Saccharomyces cerevisiae* chromosome III: a physical map and identification and location of ARS elements. Genetics 129:343–357

Newlon CS, Theis JF (1993) The structure and function of yeast ARS elements. Curr Opin Genet Dev 3:752–758

Nieduszynski CA, Knox Y, Donaldson AD (2006) Genome-wide identification of replication origins in yeast by comparative genomics. Genes Dev 20:1874–1879

Nieduszynski CA, Hiraga S, Ak P, Benham CJ, Donaldson AD (2007) OriDB: a DNA replication origin database. Nucleic Acids Res 35:D40–D46

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

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 (2006) DNA replication timing: random thoughts about origin firing. Nat Cell Biol 8:1313–1316

Rivin CJ, Fangman WL (1980) Replication fork rate and origin activation during the S phase of *Saccharomyces cerevisiae*. J Cell Biol 85:108–115

Shirahige K, Iwasaki T, Rashid MB, Ogasawara N, Yoshikawa H (1993) Location and characterization of autonomously replicating sequences from chromosome VI of *Saccharomyces cerevisiae*. Mol Cell Biol 13:5043–5056

Takahashi M (1987) A model for the spatio-temporal organization of DNA replication in mammalian cells. J Theor Biol 129:91–115

Takeda DY, Dutta A (2005) DNA replication and progression through S phase. Oncogene 24:2827–2843

Theis JF, Newlon CS (1997) The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. Proc Natl Acad Sci USA 94:10786–10791

van Brabant AJ, Buchanan CD, Charboneau E, Fangman WL, Brewer BJ (2001) An origin-deficient yeast artificial chromosome triggers a cell cycle checkpoint. Mol Cell 7:705–713

van Rossum G (1995) Python reference manual, CWI Reports. Centre for Mathematics and Computer Science, The Netherlands, CS-R9525

Weinreich M, DeBeer MAP, Fox CA (2004) The activities of eukaryotic replication origins in chromatin. Biochim Biophys Acta 1677:142–157

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 294:2357–2360

Xu W, Aparicio JG, Aparicio OM, Tavare 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 (2007) Mapping of early firing origins on a replication profile of budding yeast. Genes Cells 7:781–789

Yamashita M, Hori Y, Shinomiya T, Obuse C, Tsurimoto T, Yoshikawa H, Shirahige K (1997) The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. Genes Cells 2:655–665