Minireview

Surveying genome replication
Stephen Kearsey

Address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. E-mail: stephen.kearsey@zoo.ox.ac.uk

Abstract

Two recent studies have added microarrays to the toolkit used to analyze the origins of replication in yeast chromosomes, providing a fuller picture of how genomic DNA replication is organized.

Eukaryotes differ fundamentally from prokaryotes in the chromosomal organization of the sites at which DNA replication is initiated. In *Escherichia coli*, for instance, two replication forks originating from a single origin of replication are responsible for replicating the entire genome, whereas the replication of eukaryotic chromosomes during S phase of the cell cycle starts from many origins. Increasing the number of origins allows DNA polymerases to work at many sites simultaneously, speeding up replication, and this was presumably a prerequisite for the evolution of large genomes. Working out how origins are distributed in eukaryotic chromosomes is therefore addressing a fundamental question about replication. Origins have been most extensively studied in the budding yeast *Saccharomyces cerevisiae*, where they consist of short sequences (less than 100-150 base-pairs). Although budding yeast origins share some consensus sequence similarities, genetic and biochemical assays are required to map them precisely. Over the years, detailed work has systematically mapped the origins on two yeast chromosomes [1-3], but two recent papers [4,5] have used microarray technology to provide a whole-genome view of how replication origins are distributed and when they function only in S phase.

In the first paper, Wyrick *et al.* [4] mapped the chromosomal binding sites of the proteins associated with initiation. Yeast replication origins are bound by the proteins that make up the origin recognition complex (ORC) throughout the cell cycle, and additional proteins associate with ORC before S phase and form a pre-replicative complex (pre-RC) which is competent to initiate replication. Six MCM (mini-chromosome maintenance) proteins are key components of the pre-RC, and these factors are likely to provide the DNA helicase

function required for DNA synthesis [6]. Wyrick *et al.* [4] used the chromatin immunoprecipitation (CHIP) technique in combination with microarrays to determine the chromosomal binding sites of ORC or of MCM proteins in G1 phase of the cell cycle, and thus the location of potential origins (Figure 1a).

After cross-linking DNA to protein and fragmenting the DNA, antibodies against ORC or MCM proteins were used to purify sequences associated with these proteins. DNA was hybridized against microarrays containing probes for over 12,000 loci across the yeast genome, allowing high-resolution mapping of potential origins (generally to regions of a kilobase or less). Comparison of the hybridization data with the locations of known initiation sites made it possible to establish criteria for identifying the location of unknown origins. This approach identified 22 out of 25 known origins on chromosome III, and most (79%) sites that were identified as putative origins actually functioned as initiation sites (on plasmids), indicating that the approach is effective for locating most of the origins in the genome. The 429 putative origins identified by Wyrick *et al.* [4] are not randomly distributed but occur in intergenic zones and also cluster near telomeres. Surprisingly, intergenic zones containing elements deriving from transposable elements and tRNA genes had a higher than expected probability of containing a putative origin.

Another way of locating origins was described by Fangman's group [5], who analyzed the time at which different chromosomal regions replicate (Figure 1b). The method relies on the classical Meselson-Stahl approach, using heavy isotopes to follow the semi-conservative replication of DNA. Cells are
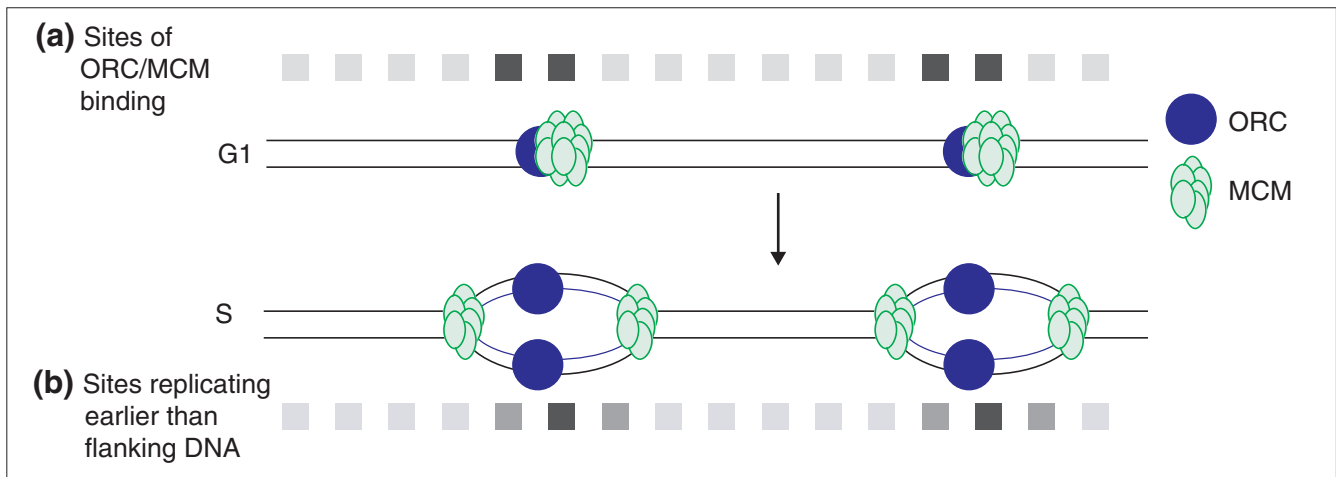
Figure 1
The principle of methods used to map replication origins. **(a)** The location of pre-replicative complexes can be detected by purifying DNA cross-linked to ORC/MCM proteins, followed by hybridization to a microarray (represented by gray squares). Positive hybridization (black squares) indicates the location of potential origins. **(b)** Hybridization of early-replicating DNA (see text for details) to a microarray allows the identification of sequences that replicate before flanking DNA, and these earliest-replicating sequences must include replication origins (black/dark gray squares).

grown for many generations in medium containing heavy isotopes (^{13}C and ^{15}N) so that both strands of the DNA incorporate the density label (so-called heavy-heavy or HH DNA). Labeled cells are arrested before the start of DNA replication and transferred to 'light' medium (containing normal isotopes of carbon and nitrogen), before allowing progress into S phase to resume. Newly synthesized DNA will thus have a heavy parental strand and a light newly synthesized strand (heavy-light or HL DNA). Samples are taken at different times after the start of S phase and, after fragmenting the DNA, HH and HL DNA can be separated by cesium chloride density gradient centrifugation. The replication time of a particular sequence is given by the point at which it is converted from HH (unreplicated) to HL (replicated) DNA.

To give a whole-genome picture of replication, Fangman's group [5] hybridized the HH and HL fractions from different time points to microarrays containing thousands of oligonucleotides, allowing sampling of sequences every 10 kilobases along each chromosome. Initially all sequences in the array hybridize to HH (unreplicated) DNA only, but, as S phase proceeds, individual sequences start to hybridize to HL (replicated) DNA. Figure 2b shows a replication profile for chromosome II, in which the time of replication is plotted against chromosomal position. Replication origins are defined by peaks that correspond to sequences that replicate before flanking DNA, and termination sites are represented by valleys that replicate later than flanking DNA. Information can also be obtained from the slope of the graph: a steep slope indicates a rapid transition from early to late replication and thus implies a slow rate of fork movement, whereas a shallow slope indicates a region of rapid fork movement. Shoulders on peaks may represent

either inefficient origins or zones where the rate of fork movement changes.

This analysis provides a refined view of the chronology of origin firing. It was previously known that some origins fire early in S phase and others later, but this analysis shows that there is really a gradient of replication-activation times, with most origins firing in the middle of S phase. Adjacent origins tend to be activated at around the same time, with different chromosomes showing a range of activation times. Compared to average sequences, origins near centromeres tend to fire early and those near telomeres late, although they are not necessarily the first and last sequences to replicate in the chromosome. Centromeres and telomeres may thus have a position effect on the timing of origin function. Interestingly, the timing of replication of the two telomeres on a single chromosome seems to be correlated, hinting that interaction between the two chromosome ends could be relevant to the timing of replication onset. Chromosomal regions are replicated not only at different times but also at different rates, with most regions replicated at rates in the range 1 to 4 kilobases per minute. This range could reflect local differences in chromatin structure, or possibly differences in the proteins assembled at replication forks.

How do the approaches taken in these studies [4,5] compare? Figure 2 compares the results for chromosome II, where 29 origins were identified by ORC/MCM binding and 25 by replication timing. The techniques give consistent results for the location of most origins. ORC/MCM binding predicts more initiation sites partly because origins that are close together may not be resolved by the replication-timing method. Also, some potential origins may not function in
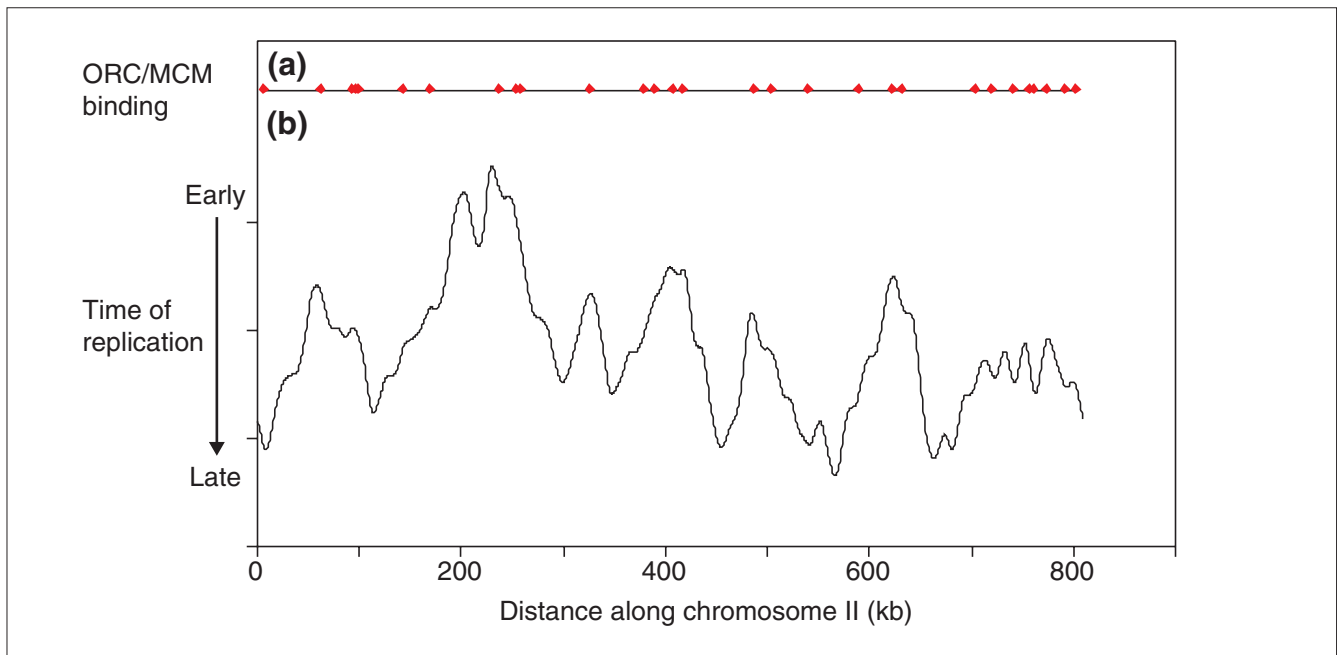
Figure 2
A comparison of origin-mapping techniques applied to yeast chromosome II (see text for further details). **(a)** Origins predicted by ORC/MCM binding [4]. **(b)** A replication profile of chromosome II [5]; peaks represent the locations of origins.

most cell cycles, because of their time of firing. For example, a potential late-replicating origin may be passively replicated by a passing fork from an early origin before it has a chance to initiate itself and thus will not be detected in the replication-timing profile. Earlier work has shown that these inefficient origins will function if flanking origins are deleted [7], and perhaps they have a function under unusual circumstances. For instance, stalling of a replication fork might prevent passive replication of a late origin, and under these circumstances firing of the late origin might contribute to the completion of chromosome replication. Less easy to explain are origins that are clearly detected by replication timing but not by ORC/MCM binding. Further work will be required to determine whether this represents some technical problem in detecting ORC/MCM binding by the CHIP method at some sites, or whether there is in fact something unusual about these origins.

The two papers [4,5] show us how a eukaryotic genome replicates in space and time, but many questions remain as to the significance of origin location and timing. The frequency of origins in chromosomes is clearly important for determining the time required for S phase, but provided that origins are adequately spaced, does it matter whether they fire early or late, or where they are? One problem here is that the molecular mechanism that sets an origin's timing of firing is not understood. Maybe there is something inherent in the mechanism leading to S-phase onset that leads to lack of synchronicity in origin firing. There may be advantages in

staggering origin firing, to reduce the demand for replication proteins and for the nucleotide precursors for DNA synthesis at any one time. Also, a checkpoint mechanism exploits this lack of synchronicity to preserve genome integrity during replication stress. If replication forks encounter problems in synthesizing DNA, for instance as a result of DNA damage, this triggers a checkpoint mechanism that inhibits the firing of late origins, thus slowing down S phase [8-10]. As DNA polymerases may stall or make errors when trying to copy a damaged template, inhibiting late origins preserves genome integrity by providing extra time for DNA repair, and clearly this mechanism would not work if all origins fired synchronously at the start of S phase.

Another factor that has been explored to determine its relevance to origin location and replication timing is transcription. In higher eukaryotes, transcriptionally active parts of the genome are replicated earlier than heterochromatin, but in yeast there are no striking correlations between transcription and replication timing, with the exception of telomeric regions. Transcription is repressed in telomeric regions [11], but it seems unlikely that this is caused directly by late replication, and a more plausible explanation is that some aspect of local chromatin structure inhibits both transcription and early replication. As well as functioning in replication, ORC is known to have a role in transcriptional control at the silent mating-type loci of yeast. In this case, it promotes the association with chromatin of the proteins involved in silencing - in other words, the formation of local heterochromatin

needed for inactivation of mating-type genes - so a non-replication function is relevant to the location of some ORC binding sites in the yeast genome. As far as is known, ORC does not have this function at other chromosomal regions in yeast, but in *Drosophila* ORC may have a general role in establishing heterochromatin [12].

Further insight into the significance of origin location and timing should come from applying microarray techniques to the examination of mutants defective in replication proteins. In addition to the methods described here, which of necessity look at average properties of populations of cells, it will be interesting to look at origin location and timing in single chromosomes using, for instance, DNA combing [13], to see how much S-phase to S-phase variability there can be in genome replication. Finally, application of these methods to other eukaryotes should provide new ways of mapping origins that have previously been difficult to study. In fission yeast and metazoan chromosomes, origins are much larger than in budding yeast and, under some circumstances, such as in early *Xenopus* development, specific DNA sequences are not required for initiation. We can look forward over the coming years to a much more detailed understanding of evolutionary and developmental changes in the genomic organization of replication.

References
1. Friedman KL, Brewer BJ, Fangman WL: **Replication profile of *Saccharomyces cerevisiae* chromosome VI.** *Genes Cells* 1997, **2**:667-678.
2. Poloumienko A, Dershowitz A, De J, Newlon CS: **Completion of replication map of *Saccharomyces cerevisiae* chromosome III.** *Mol Biol Cell* 2001, **12**:3317-3327.
3. Yamashita M, Hori Y, Shinomiya T, Obuse C, Tsurimoto T, Yoshikawa H, Shirahige K: **The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI.** *Genes Cells* 1997, **2**:655-665.
4. Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM: **Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins.** *Science* 2001, **294**:2357-2360.
5. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, Lockhart DJ, Davis RW, Brewer BJ, Fangman WL: **Replication dynamics of the yeast genome.** *Science* 2001, **294**:115-121.
6. Labib K, Diffley JFX: **Is the MCM2-7 complex the eukaryotic DNA replication fork helicase?** *Curr Opin Genet Dev* 2001, **11**:64-70.
7. Santocanale C, Sharma K, Diffley JF: **Activation of dormant origins of DNA replication in budding yeast.** *Genes Dev* 1999, **13**:2360-2364.
8. Santocanale C, Diffley JF: A **Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication.** *Nature* 1998, **395**:615-618.
9. Shirahige K, Hori Y, Shiraishi K, Yamashita M, Takahashi K, Obuse C, Tsurimoto T, Yoshikawa H: **Regulation of DNA-replication origins during cell-cycle progression.** *Nature* 1998, **395**:618-621.
10. Tercero JA, Diffley JF: **Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint.** *Nature* 2001, **412**:553-557.
11. Dubrana K, Perrod S, Gasser SM: **Turning telomeres off and on.** *Curr Opin Cell Biol* 2001, **13**:281-289.
12. Pak DT, Pflumm M, Chesnokov I, Huang DW, Kellum R, Marr J, Romanowski P, Botchan MR: **Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes.** *Cell* 1997, **91**:311-323.
13. Legronne A, Pasero P, Bensimon A, Schwob E: **Monitoring S phase progression globally and locally using BrdU incorporation in TK+ yeast strains.** *Nucleic Acids Res* 2001, **29**:1433-1442.