Chromosomal Mcm2-7 distribution is a primary driver of genome replication timing in budding yeast, fission yeast and mouse

Eric J. Foss, Smitha Sripathy, Tonbelle Gatbonton-Schwager, Hyunchang Kwak, Adam H. Thiesen, Uyen Lao, Antonio Bedalov

Abstract

The spatio-temporal program of genome replication across eukaryotes is thought to be driven both by the uneven loading of pre-replication complexes (pre-RCs) across the genome at the onset of S-phase, and by differences in the timing of activation of these complexes during S-phase. To determine the degree to which distribution of pre-RC loading alone could account for chromosomal replication patterns, we identified the binding sites of the Mcm2-7 helicase complex, a key component of the pre-RC that is required for initiation of DNA replication, in budding yeast, fission yeast and mouse. In budding yeast, we detected Mcm2 binding in sharply focused peaks, generally with a single double hexamer bound at known origins of replication. In fission yeast, Mcm2 binding, while still concentrated at known origins, was more diffuse, often with 6 to 8 helicase complexes distributed along 0.5-1.5 kb sized origins, and with significantly more binding between origins. Finally, in mouse, we found even more diffuse Mcm2-7 distribution, with the density of Mcm2-7 binding in G1 recapitulating to a remarkable degree the replication program implemented in S-phase. Computer simulations that assign each licensed origin an equal probability of firing show that the observed Mcm2-7 density distribution in G1 across all three species largely recapitulated the DNA replication program. We conclude that the pattern of origin licensing from yeast to mammals is sufficient to explain most differences in replication timing without invoking an overarching temporal program of origin firing.

Results

Initiation of DNA replication requires activation of the MCM2-7 helicase complex (Mcm2-7) at sites at which this complex was loaded during the preceding G2/M and G1 phases [1-4]. However, the order in which different regions of the genome complete replication is widely assumed to reflect not only the locations of Mcm2-7 loading but also a program of control above the level of loading, in which certain complexes are activated before others [5, 6]. The distribution of Mcm2-7 is therefore expected to constrain but not fully determine the program of genome replication. Mcm2-7 encircles approximately 60 base pairs (bp) of DNA as a double hexamer whose monomer components are juxtaposed head-to-head at their N-termini with C-

termini facing away from each other [7]. To exploit this arrangement of Mcm2-7 hexamers to identify their exact loading sites *in vivo*, we tagged its Mcm2, Mcm4 and Mcm6 subunits at their C-termini in *Saccharomyces cerevisiae* with micrococcal nuclease (MNase) [8], permeabilized G1-arrested cells, activated the MNase by addition of calcium, and prepared libraries for paired end sequencing using total extracted DNA without any size fractionation [9]. The strains with tagged MCM proteins exhibited growth rates comparable to those in wild type, indicating that the presence of the tag did not interfere with the proteins' function (Fig. 1A). Because the PCR involved in library preparation preferentially amplifies short fragments, which in turn are mostly generated by MNase activity, we expected the resulting libraries to reflect predominantly the sites where double Mcm2-7 hexamers have been loaded; furthermore, these sites should be congruent for the three different libraries. Consistent with this expectation, our sequencing results for the libraries prepared from the three different tagged subunits were focused in sharp peaks that coincided with each other (Fig 1B). Furthermore, the size distribution of fragments for all three libraries peaked in the 50-62 bp size range, which is consistent with the 62 base pairs that have been shown to be protected by Mcm2-7 double-hexamers using cryo-electron microscopy (cryo-EM; Fig 1C) [7]. [The peak of fragments in the 150-200 bp size range reflects cleavage by the MNase-tagged protein between flanking nucleosomes. We have seen this phenomenon with multiple MNase-tagged DNA binding proteins (data not shown).] Individual Mcm2-7 footprints can be visualized as heat maps, with fragment size plotted according to genomic location, and read depths represented by color intensity [9]. A typical example of this is shown in Figure 1D: The Mcm2 footprint at ARS1103, which was among the 10 most abundant footprints in the genome, was composed of reads in the 50-70 bp size range located immediately adjacent to the ARS consensus sequence (ACS) at this origin. Nearly identical footprints were generated by MNase-tagged Mcm4 and Mcm6, confirming that the footprints indeed reflect Mcm2-7 binding (Fig 1D). The prominent MCM2 signal at ARS1103 (Fig 2, left middle panel), which is a highly efficient origin that fires in a large fraction of cells, suggests that at least one Mcm2-7 double hexamer is loaded at this origin in a large fraction of cells; the virtual absence of Mcm2-7 signal in the surrounding 1.5 kb of DNA, rules out the presence of multiple Mcm2-7 complexes at this origin. We conclude that the

Mcm2-7 abundance at individual origins is largely a reflection of the fraction of cells in which a single complex had been loaded.

We next used the same approach to identify Mcm2-7 localization in the fission yeast *Schizosaccharomyces pombe*. Like in budding yeast, the distribution of fragment length was bimodal, with the major peak consistent with the size protected by MCM complexes (Supplemental Figure 1A). Also similar to budding yeast, the Mcm2 peaks largely coincided with previously identified origins of replication (central panel). However, two features of Mcm2-7 distribution were notably different: First, unlike *S. cerevisiae*, where we found predominantly a single Mcm2-7 double-hexamer at most of the origins, we found 4-12 adjacent Mcm2-7 footprints distributed over 500-1500 bp at most of the known origins in *S. pomb*e (Fig 2, top middle panel, and Supplemental Figure 2). Such a distribution of Mcm2-7 footprints is consistent with the previously described initiation zones in *S. pombe* [10]. Second, we found significantly more Mcm2-7 between the origins (Fig 2, central panel). These results demonstrate the utility of this technique in elucidating organism-specific differences in the distribution of MCM complexes.

We next determined the distribution of Mcm2-7 complexes by MNase-seq in mouse Patski cells. We expressed mouse Mcm2 tagged with MNase using a lentiviral vector, choosing an expression level that was not above the physiologic levels (Supplemental Fig 3). As we had seen in yeast, the distribution of sequenced fragment sizes was bimodal, with the major peak composed of fragments from 50 to 100 bp (Supplemental Fig 2B). We therefore concluded that the architecture of individual MCM binding sites in mouse is similar to that in budding and fission yeasts, though the distribution of MCM binding sites across the genome was much more uniform.

Although MCM binding in mouse was relatively ubiquitous, the density of binding sites per megabase varied approximately 2.5-fold across the genome. In order to explore the effect of this variation on replication, we compared the Mcm2-7 distribution to replication profiles that

we generated by deep sequencing of G1 and S phase fractions of log phase cells that had been flow-sorted on the basis of DNA content [11]; this technique is based on the principle that early replicating regions in the S phase fraction will be more abundant than late replicating regions. The resulting replication profiles strikingly recapitulated the MCM binding profiles generated from G1-arrested cells, with early replicating domains corresponding to regions of higher Mcm2-7 density (Fig 2, right middle and right bottom panels). We conclude that Mcm2-7 density in G1 cells foretells the replication program in the ensuing S-phase.

In order to determine whether differences in the density of MCM binding in G1 are sufficient to drive observed patterns of replication timing in S phase observed in all three organisms, we modeled Mcm2-7 binding and DNA replication *in silico* using the assumption that DNA replication initiates in each cell at random from sites at which Mcm2-7 is bound in that cell. This model was remarkably accurate in its prediction of replication timing, and furthermore it was robust to changes in additional assumptions about how limitations of origin firing factors restrict the number of origins that can fire simultaneously and how freely those limiting firing factors can diffuse.

Models for differences in replication timing have long assumed differences in the propensity of individual licensed origins to fire, with some origins having an intrinsic tendency to fire earlier than others [5, 6], although models that instead emphasize the role of stochastic firing of licensed origins in determining replication timing have also been proposed [12-14]. Our results demonstrate that differences in replication timing can be largely attributed to differences in MCM binding in G1.

Figure legends

Figure 1. Tagging MCM2, MCM4 and MCM6 subunits with MNase in budding yeast reveals binding MCM helicase binding sites. (A) Streaking out strains on rich medium (YEPD) indicates that tagging MCM2, MCM4 and MCM6 with MNase did not affect growth rates. (B) Plot of

read depths across chromosome I shows that binding sites for MCM2 (black, top panels), MCM4 (green, bottom left) and MCM 6 (blue, bottom right) were largely congruent. (C) Fragment sizes in libraries from strains with tagged MCM2, MCM4 and MCM6 are similar to each other and are consistent with the 62 base pairs of DNA shown to be protected by the MCM helicase complex by cryo-EM [7]. D. Binding sites of MCM subunits can be visualized by chromosomal location on the X axis, fragment size on the Y axis and read depth according to color intensity, creating images reminiscent of agarose gels. All three subunits generate similar footprints, with ARS1103 on chrXI shown.

Figure 2. MCM distribution and replication in budding yeast (chrXI), fission yeast (chrII) and mouse (chr16). Left 3 panels show budding yeast, middle three panels show fission yeast and right 3 panels show mouse. For MCM2-ChEC experiments (top 2 rows), budding yeast were arrested for 2 hours with alpha factor, fission yeast were transferred from log phase to 15 mM HU for 3 hours, and Patski cells were arrested with serum starvation. Top row: MCM2 binding visualized by plotting chromosomal location on the x axis, fragment size on the y axis, and read depth indicated by color intensity. The budding yeast image corresponds to ARS1103. Middle row: MCM2-ChEC read depths plotted across 240 kb (both yeasts) or 100 Mb (mouse). Vertical dark gray lines in both yeasts indicate known origins of replication, as listed in oridb.org. Red arrows indicate regions of correspondence between the top and middle rows. Bottom row: Replication as determined by S to G1 ratio of flow sorted cells for budding yeast [15], fission yeast [10] and mouse (this publication) is plotted in black. Replication simulation based on MCM2-ChEC data, with each loaded MCM having an equal probability of firing, plotted in red.

Supplemental Figure 1. Fragment size distributions for MCM2-MNase libraries prepared from fission yeast (A) and mouse (B).

Supplemental Figure 2. Close up view of MCM2-MNase signal at fission yeast origin on chrI. Coordinates (kb) on chrI are plotted on the x axis, fragment size is plotted on the y axis, read depth is indicated by color intensity.

Supplemental Figure 3. Anti-Mcm western showing cell without (left) and with (right) MNase-tagged Mcm2.

Acknowledgements

This work was supported by the National Institute of Health Grants R01-GM117446 (to A.B.).

Bibliography

1. Bell, S.P. and K. Labib, *Chromosome Duplication in Saccharomyces cerevisiae.* Genetics, 2016. **203**(3): p. 1027-67.
2. Bleichert, F., M.R. Botchan, and J.M. Berger, *Mechanisms for initiating cellular DNA replication.* Science, 2017. **355**(6327).
3. Fragkos, M., et al., *DNA replication origin activation in space and time.* Nat Rev Mol Cell Biol, 2015. **16**(6): p. 360-74.
4. Riera, A., et al., *From structure to mechanism-understanding initiation of DNA replication.* Genes Dev, 2017. **31**(11): p. 1073-1088.
5. Dileep, V. and D.M. Gilbert, *Single-cell replication profiling to measure stochastic variation in mammalian replication timing.* Nat Commun, 2018. **9**(1): p. 427.
6. Rhind, N. and D.M. Gilbert, *DNA replication timing.* Cold Spring Harb Perspect Biol, 2013. **5**(8): p. a010132.
7. Abid Ali, F., et al., *Cryo-EM structure of a licensed DNA replication origin.* Nat Commun, 2017. **8**(1): p. 2241.
8. Zentner, G.E., et al., *ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo.* Nat Commun, 2015. **6**: p. 8733.
9. Foss, E.J., et al., *Sir2 suppresses transcription-mediated displacement of Mcm2-7 replicative helicases at the ribosomal DNA repeats.* PLoS Genet, 2019. **15**(5): p. e1008138.
10. Daigaku, Y., et al., *A global profile of replicative polymerase usage.* Nat Struct Mol Biol, 2015. **22**(3): p. 192-198.
11. Muller, C.A. and C.A. Nieduszynski, *Conservation of replication timing reveals global and local regulation of replication origin activity.* Genome Res, 2012. **22**(10): p. 1953-62.
12. Hawkins, M., et al., *High-resolution replication profiles define the stochastic nature of genome replication initiation and termination.* Cell Rep, 2013. **5**(4): p. 1132-41.
13. Hyrien, O., *Up and Down the Slope: Replication Timing and Fork Directionality Gradients in Eukaryotic Genomes*, in *The Initiation of DNA Replication in Eukaryotes*. 2016. p. 65-85.
14. Kelly, T. and A.J. Callegari, *Dynamics of DNA replication in a eukaryotic cell.* Proc Natl Acad Sci U S A, 2019. **116**(11): p. 4973-4982.

15. Foss, E.J., et al., *SIR2 suppresses replication gaps and genome instability by balancing replication between repetitive and unique sequences.* Proc Natl Acad Sci U S A, 2017. **114**(3): p. 552-557.

Figure 1

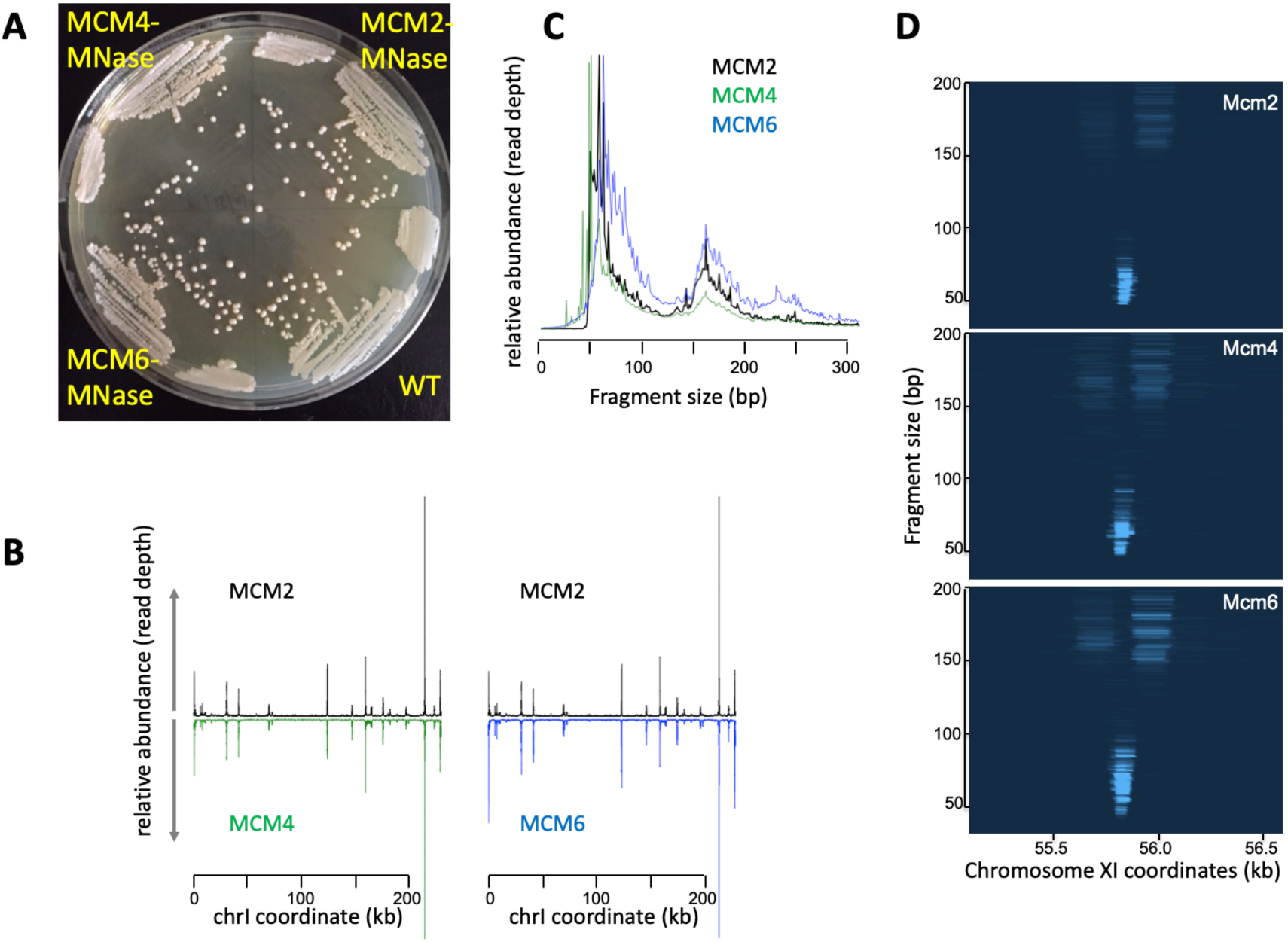
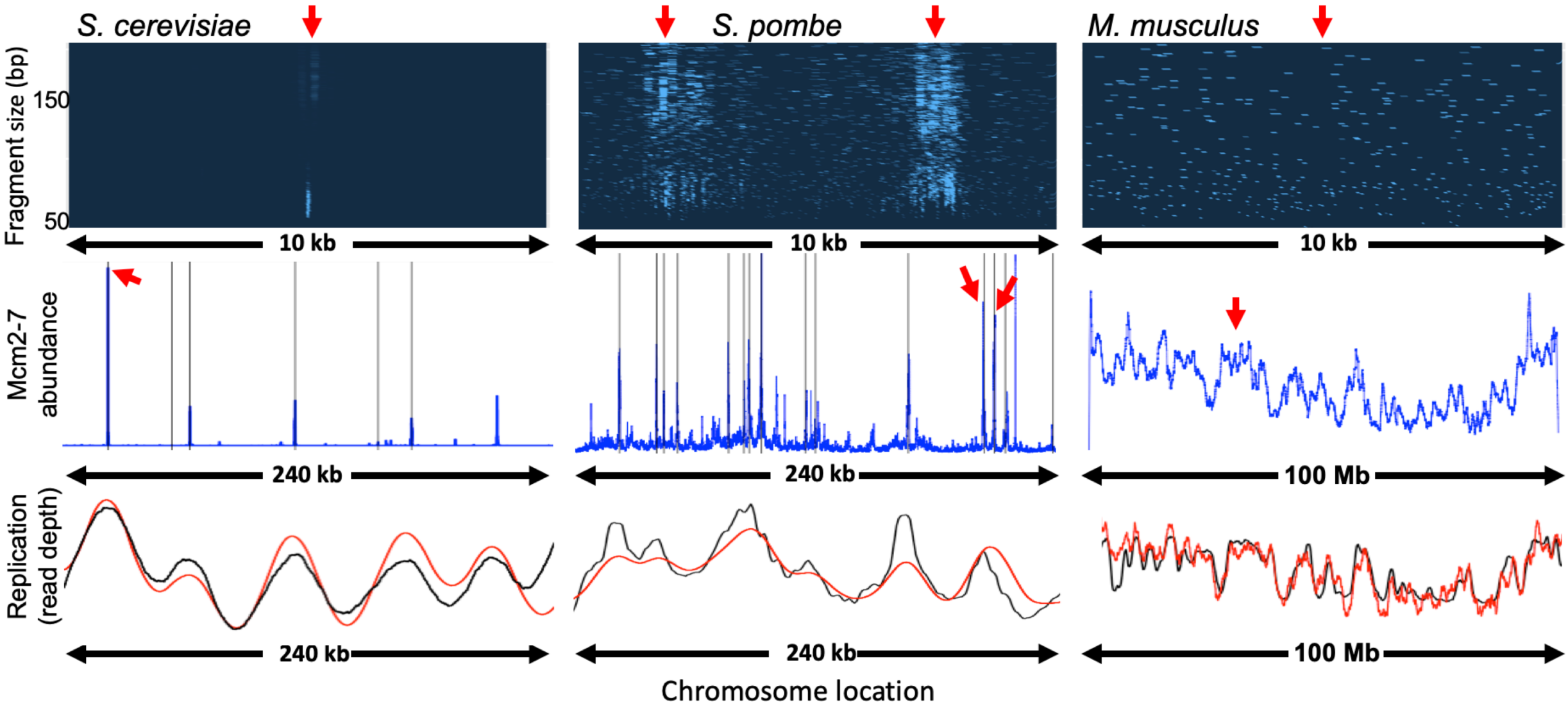
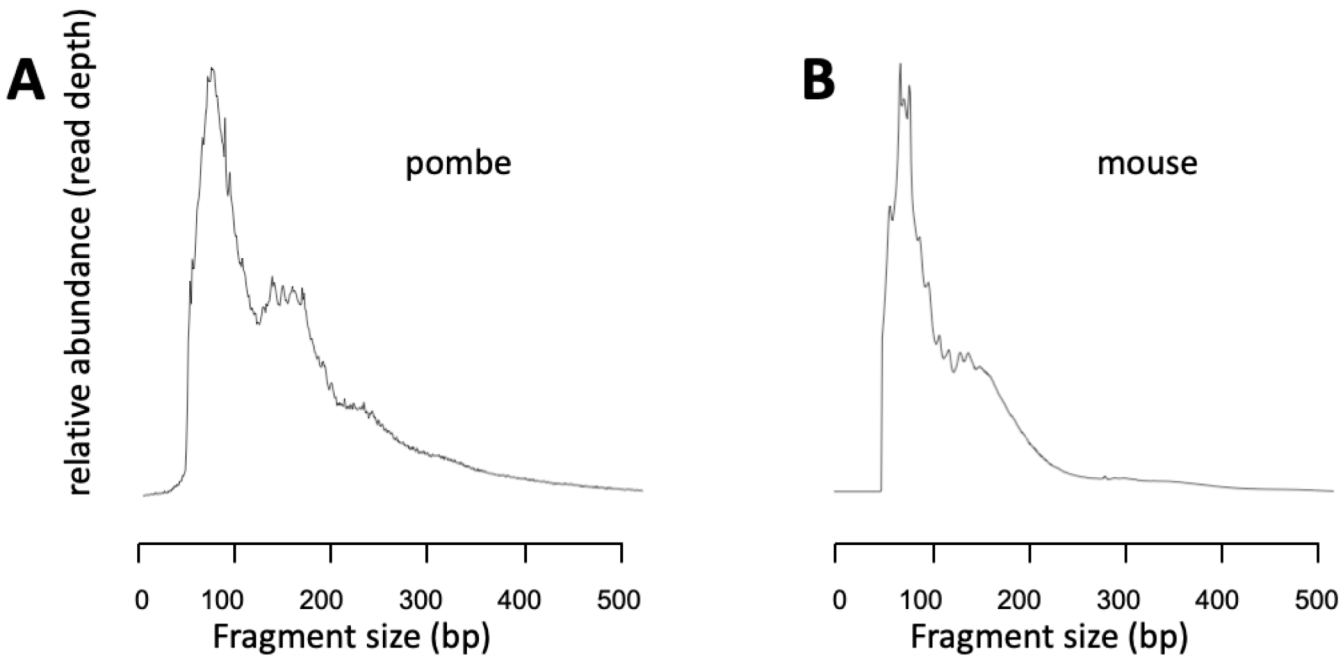

A
MCM4-
MNase
MCM2-
MNase
MCM6-
MNase
WT
C
MCM2
MCM4
MCM6
relative abundance (read depth)
0
100
200
300
Fragment size (bp)
D
Mcm2
Mcm4
Mcm6
Fragment size (bp)
200
150
100
50
55.5
56.0
56.5
Chromosome XI coordinates (kb)
B
relative abundance (read depth)
MCM2
MCM4
MCM2
MCM6
0
100
200
chrI coordinate (kb)

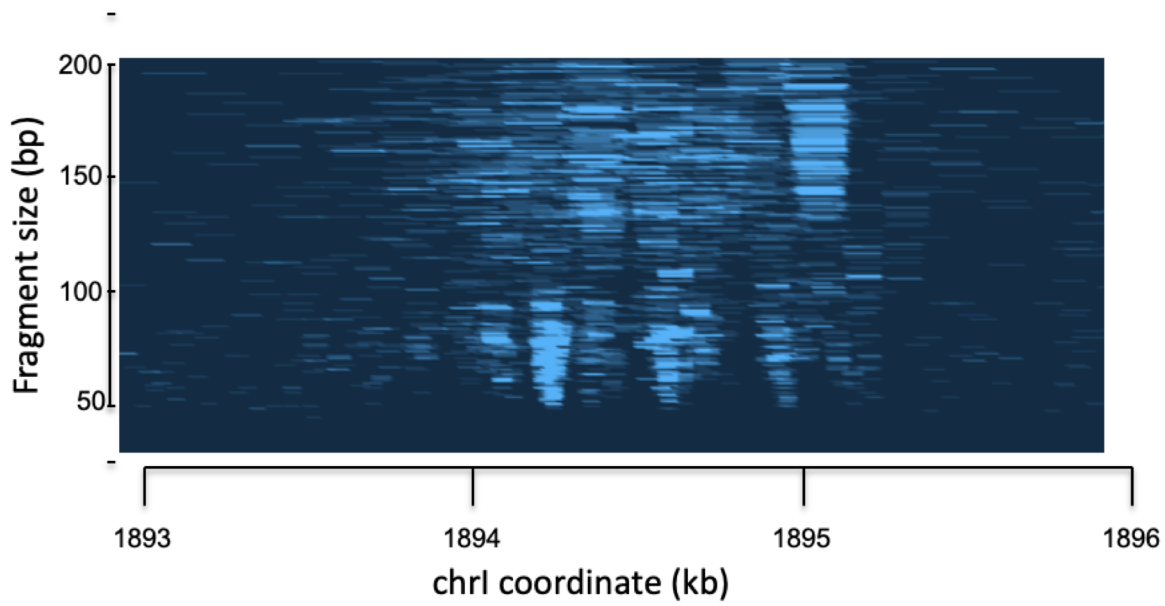
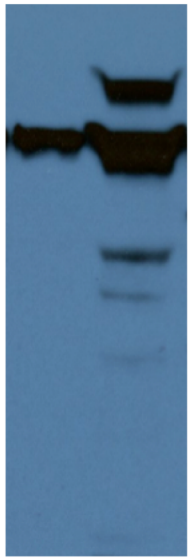
Figure 2


S. cerevisiae
S. pombe
M. musculus
Fragment size (bp)
150
50
10 kb
10 kb
10 kb
Mcm2-7 abundance
240 kb
240 kb
100 Mb
Replication (read depth)
240 kb
240 kb
100 Mb
Chromosome location

Supplemental Figure 1



Supplemental Figure 2



Supplemental Figure 3