A global profile of replicative polymerase usage

Yasukazu Daigaku1,5, Andrea Keszthelyi1,5, Carolin A Müller2, Izumi Miyabe1, Tony Brooks3, Renata Retkute4, Mike Hubank3, Conrad A Nieduszynski2 & Antony M Carr1

Three eukaryotic DNA polymerases are essential for genome replication. Polymerase (Pol) α–primase initiates each synthesis event and is rapidly replaced by processive DNA polymerases: Polε replicates the leading strand, whereas Polδ performs lagging-strand synthesis. However, it is not known whether this division of labor is maintained across the whole genome or how uniform it is within single replicons. Using Schizosaccharomyces pombe, we have developed a polymerase usage sequencing (Pu-seq) strategy to map polymerase usage genome wide. Pu-seq provides direct replication-origin location and efficiency data and indirect estimates of replication timing. We confirm that the division of labor is broadly maintained across an entire genome. However, our data suggest a subtle variability in the usage of the two polymerases within individual replicons. We propose that this results from occasional leading-strand initiation by Polε followed by exchange for Polδ.

Accurate DNA replication is fundamental to life, and errors that occur during replication underpin the genome instability that is the hallmark of cancer development1,2. In most eukaryotes, bidirectional replication is initiated stochastically, with distinct regions of the genome showing varying initiation efficiencies and distinct temporal regulation3. In budding yeast, specific DNA consensus sequences define the binding of the origin recognition complex (ORC) to DNA throughout the cell cycle4. Each region of replication initiation is thus defined by a single DNA sequence or origin. In higher eukaryotes, ORC association with the chromosomes varies through the cell cycle, and the mechanisms defining where the ORC binds are not understood. Initiation zones in higher eukaryotes are probably composed of numerous low-efficiency origins clustered together3.

In exponentially growing budding yeast, the different origins are activated with different efficiencies. Thus, times at which different initiation regions are replicated (the population average) are distinct5. In higher eukaryotes, growing cultures of individual cell types display reproducible replication-timing profiles, thus indicating that ORC association and/or the likelihood of replication initiation from ORC-associated regions are stable characteristics of specific cell types6. Interestingly, replication-timing profiles for different mammalian cell types correlate well with three-dimensional chromosome interaction maps, thus suggesting a link between replication timing and chromatin organization within the nucleus (reviewed in ref. 3).

The ORC attracts the MCM complex in the G1 phase of the cell cycle, licensing the site for initiation7. The six-subunit MCM complex is the core of the replicative helicase, which is subsequently activated by the loading of two additional components: Cdc45 and the four-subunit GINS complex. The resulting active helicase is known as CMG8. An ancillary replisome component, the Ctf4 trimer, links Polα–primase to CMG9,10, coordinating the necessary initiation events. The Polε holoenzyme interacts directly with GINS, an association also independently required for the initial formation of CMG11–13. Once CMG is formed, the Polε holoenzyme-GINS interaction is not required for CMG helicase activity. It is not known whether the Polδ holoenzyme interacts directly with CMG. Once DNA replication is initiated, each fork synthesizes the leading strand continuously and the lagging strand discontinuously.

Certain DNA polymerase mutations introduce a biased mutation spectrum. This has allowed assignment of the source of mutations due to mispairing on one or the other DNA strand14. From studies using these mutant polymerases, Polε was genetically assigned as the leading-strand DNA polymerase at several loci in Saccharomyces cerevisiae. Similarly, Polδ was assigned as the major lagging-strand polymerase15. These data led to the model that the labor of replication is shared: Polε replicates the leading strand and Polδ the lagging strand. An equivalent experiment using S. pombe similarly assigned Polδ to the lagging strand16, demonstrating evolutionary conservation of polymerase usage. An S. pombe mutant Polε that incorporated ribonucleotides into DNA at increased frequency was used to physically assign Polδ to leading-strand synthesis16. These experiments relied on the increased incorporation of rNTPs into the leading strand, thus causing that specific strand to be fragmented by alkalai treatment, which cleaves the phosphate backbone at ribonucleotides but not deoxyribonucleotides.

To establish whether the division of labor between Polε and Polδ is consistent across an entire genome and to ascertain whether there is variation in the usage between the two polymerases within a single replicon, we set out to physically map, genome wide, the division of labor between these polymerases. We devised a strategy to identify,
by high-throughput sequencing, the position of ribonucleotides in the genome and combined this with Polε and Polδ mutants that incorporate excess ribonucleotides to establish a Pu-seq methodology that allowed us to map the division of labor genome wide. We confirm that the division of labor is broadly maintained across an entire genome. We also demonstrate that a single Pu-seq experiment, which consists of two library samples for deep sequencing (one each from asynchronous cultures of the respective polymerase mutants) delivers a direct and extremely high-resolution genome-wide map of RNA replication initiation and allows the indirect calculation of robust genome-wide replication-timing data. The resolution of our data revealed evidence for subtle variability in the usage of the two polymerases within individual replicons. We suggest that this results from occasional leading-strand initiation by Polδ.

RESULTS

At physiological dNTP and rNTP concentrations, S. cerevisiae replicative DNA polymerases incorporate, in vitro, ribonucleotides at frequencies ranging from 1:650 bp (Polε) to 1:5,000 bp (Polδ).17 Ribonucleotides are efficiently removed from duplex DNA by ribonucleotide excision repair (RER). RNase H2 nicks 5′ to the ribonucleotide, Polδ (or Polε) initiates strand-displacement synthesis, and Fen1 (or Exo1) removes the resulting flap before ligation completes repair.18 In the absence of RER, single ribonucleotides persist (although some are removed by Top1 (refs. 19–21)). Ribonucleotides can template DNA synthesis, albeit with a reduction in processivity.22,23 We previously exploited an S. pombe cdc20-M630F (Polε) allele to introduce excess ribonucleotides into DNA replicated by Polε. Southern blot analysis in an RNase H2–deficient (rnh201Δ) background previously provided physical evidence that Polε performed the majority of leading-strand synthesis.16 To facilitate mapping the division of labor genome wide, we have generated an equivalent mutation for Polδ, cdc6-L591G. DNA prepared from cells containing this mutation showed lagging strand–specific degradation when alkali gels were probed for sequences flanking an efficient origin (Fig. 1a, b). This is complementary to the DNA prepared from cells containing the previously characterized cdc20-M630F (Polε) allele, which demonstrated leading strand–specific degradation (Fig. 1b). Both the cdc20-M630F (Polε) and the cdc6-L591G (Polδ) mutant strains in the rnh201Δ background incorporated similar levels of ribonucleotides,24 grew with similar kinetics and displayed similar flow-cytometry profiles (Fig. 1c).

Mapping polymerase usage across the genome

Alkali treatment of duplex ribonucleotide–containing DNA results in phosphate-backbone cleavage 3′ to the ribose to result in a 5′-OH (Fig. 1d). If the denatured DNA is used to template random hexamer primer extension, 5′-to-3′ synthesis results in a flush end adjacent to the initial ribose (Fig. 1e). By generating a library from single-stranded DNA and placing distinct index primers at each end, deep-sequencing reads can be mapped to individual strands, locating with base accuracy the original ribonucleotide. To map replication polymerase usage across the genome, we therefore grew two RNaseH2–deficient cultures containing either cdc20-M630F (Polε) or cdc6-L591G (Polδ) mutations, prepared DNA, treated with alkali

**Figure 1** rNMP incorporation into DNA in Polδ (cdc6-L591G) and Polε (cdc20-M630F) cells. (a) Schematic representation of the region flanking ARS3006 and ARS3007. Leading and lagging strands are represented by red and blue lines, respectively. (b) Southern blot of digested and alkali-treated genomic DNA hybridized with probes indicated in a. Uncropped blot image is in Supplementary Data Set 1. (c) Top, proportion of high-mobility product from rnh201Δ cells in experiments equivalent to that in b. Bottom, flow-cytometry analysis of wild-type, rnh201Δ, cdc20-M630F (Polε) and rnh201Δ cdc6-L591G (Polδ) cells with population-doubling times in parentheses. 2C, DNA content equivalent to two copies of the genome; 4C, DNA content equivalent to four copies of the genome (two replicated sister cells still attached at the septum). (d) Hydrolysis at the misincorporated RNA molecule. The 2′-OH group of the rNMP is susceptible to nucleophilic attack (left), causing cleavage of the sugar backbone and the generation of a cyclic 2′,3′-phosphate and a 5′-OH group. (e) Schematic of library preparation. Position of incorporated ribonucleotides is shown as ‘r′.
and created two independent libraries. We mapped approximately 10 million paired-end sequence reads for each strain to 300-bp bins across the genome (Fig. 2a). We then calculated the relative ratio of reads from the Polɛ and Polδ data sets (Fig. 2b) and smoothed the data to provide frequency scores representative of relative Polɛ and Polδ usage for the Watson (+) and Crick (−) strands (Fig. 2c).

**Polymerase usage transitions define initiation sites**

Bidirectional initiation and the division of polymerase labor predicts a reciprocal demarcation on both the Watson and the Crick strands between Polɛ (leading) and Polδ (lagging) usage for each initiation zone. Efficient origins should manifest as sharp reciprocal changes in the polymerase usage ratios. Less efficient origins, which are replicated passively in most cells, should present as reciprocal inflections in otherwise uniform gradients. We thus used the two independent data sets to calculate Polɛ usage on the Watson strand or Polδ usage on the Crick strand (Fig. 3a) and plotted the differential of each neighboring data point (Fig. 3b). Where a reciprocal positive peak was identified (i.e., there was a change in polymerase usage in both data sets), we derived maxima and minima and plotted the average of their differences (Fig. 3c; further explanation in Fig. 3e). Peak heights reflect relative origin efficiency: the highest peaks correspond to the most-efficient origins (distribution of origin efficiencies in Supplementary Fig. 1a; identified origins and their relationship to previous studies in Supplementary Table 1).

To account for experimental variation, we analyzed four additional independent experiments and annotated how often each origin was identified (Supplementary Table 1). To independently visualize origins in a manner concordant with the literature23, we synchronized wild-type cells in G2, released them into S phase in the presence of hydroxyurea (HU) plus the nucleotide analog bromodeoxyuridine (BrdU) and quantified replication by using BrdU immunoprecipitation (BrdU-IP) plus deep sequencing (Fig. 3d). This identified 421 origins, >90% of which correspond to Pu-seq origins (Supplementary Table 1 and Supplementary Fig. 2).

**A map of replication timing by marker frequency analysis**

Although Pu-seq provides a direct assay for replication-initiation efficiency, it can also indirectly provide information about relative replication timing (described below). To validate replication-timing data calculated from the Pu-seq experiments, we first wished to generate a direct replication-timing map for *S. pombe* that is not biased by cell synchronization or treatment with replication inhibitors25. We thus mapped replication profiles of cells synchronized by elutriation, using marker frequency analysis (Fig. 4a). We examined aliquots of an elutriated culture over time for mitotic index, septation and DNA content. From these data, we calculated percentages of G2-phase, M-phase, S-phase and post-S phase cells for each time point, on the basis of the known cell-cycle behavior of *S. pombe* (Fig. 4b and Online Methods). We calculated the fraction of DNA replicated for each time point and set boundaries for the beginning and end of S phase (Fig. 4c). Next, we extracted DNA from the indicated aliquots spanning S phase and prepared libraries for deep sequencing. We compared the proportion of reads for each 1-kb bin across the genome to a fully replicated (G2) control and calculated the percentage of replication at each locus for each time point sequenced (Fig. 4d).

Because elutriation can cause cellular perturbation due to centrifugation26, we validated that elutriation did not distort replication profiles by performing fluorescence-activated cell sorting (FACS) and deep-sequence analysis (Sort-seq) in which S-phase cells

---

**Figure 2** Polymerase usage across the fission-yeast genome. (a) Total counts of the flanking 5′ nucleotide of the sequenced reads assigned to 300-bp bins plotted for a representative region. Polɛ (cdc20-M630F), red; Polδ (cdc6-L591G), blue. (b) Ratio of the relative reads in each bin for Polɛ (cdc20-M630F, e[e + δ]) in red and Polδ (cdc6-L591G, δ/δ + e) in blue, plotted for the same region. (c) Smoothed data providing a map of polymerase usage. Supplementary data sets to visualize the whole genome are listed in Supplementary Table 2.

**Figure 3** Identification of replication origins. (a) Usage of Polɛ on the Watson strand (blue) and Polδ on the Crick (red) strand. (b) Differential (diff.) of the polymerase usage plots from a. (c) Origin efficiencies (E_{f0}) calculated from Pu-seq data. (d) Comparative map of origins generated by deep sequencing of DNA immunoprecipitated by anti-BrdU antibodies. YD18 cells were synchronized by cdc25 (G2) block and release into HU. Supplementary data sets to visualize the whole genome are listed in Supplementary Table 2. (e) Example of how origin efficiencies were quantified. Polymerase usage ratio, established minima and maxima (yellow triangles) around the reciprocal peaks (yellow dots) identified from a. Differentials, example region of differentials from b. Origin efficiency, differences between the above identified maxima and minima (E_{f0} and E_{f0}). Average origin efficiency, averaged differences producing the relative origin efficiency (E_{f0}).

---

**Table 1** Polɛ data to provide frequency scores representative of relative Polɛ reads from the Polɛ minima (whole genome are listed in Supplementary data sets to calculate Polɛ in otherwise uniform gradients. We thus used the two independent data sets to calculate Polɛ usage on the Watson strand or Polδ usage on the Crick strand (Fig. 3a) and plotted the differential of each neighboring data point (Fig. 3b). Where a reciprocal positive peak was identified (i.e., there was a change in polymerase usage in both data sets), we derived maxima and minima and plotted the average of their differences (Fig. 3c; further explanation in Fig. 3e). Peak heights reflect relative origin efficiency: the highest peaks correspond to the most-efficient origins (distribution of origin efficiencies in Supplementary Fig. 1a; identified origins and their relationship to previous studies in Supplementary Table 1).

To account for experimental variation, we analyzed four additional independent experiments and annotated how often each origin was identified (Supplementary Table 1). To independently visualize origins in a manner concordant with the literature23, we synchronized wild-type cells in G2, released them into S phase in the presence of hydroxyurea (HU) plus the nucleotide analog bromodeoxyuridine (BrdU) and quantified replication by using BrdU immunoprecipitation (BrdU-IP) plus deep sequencing (Fig. 3d). This identified 421 origins, >90% of which correspond to Pu-seq origins (Supplementary Table 1 and Supplementary Fig. 2).

**Polymerase usage transitions define initiation sites**

Bidirectional initiation and the division of polymerase labor predicts a reciprocal demarcation on both the Watson and the Crick strands between Polɛ (leading) and Polδ (lagging) usage for each initiation zone. Efficient origins should manifest as sharp reciprocal changes in the polymerase usage ratios. Less efficient origins, which are replicated passively in most cells, should present as reciprocal inflections in otherwise uniform gradients. We thus used the two independent data sets to calculate Polɛ usage on the Watson strand or Polδ usage on the Crick strand (Fig. 3a) and plotted the differential of each neighboring data point (Fig. 3b). Where a reciprocal positive peak was identified (i.e., there was a change in polymerase usage in both data sets), we derived maxima and minima and plotted the average of their differences (Fig. 3c; further explanation in Fig. 3e). Peak heights reflect relative origin efficiency: the highest peaks correspond to the most-efficient origins (distribution of origin efficiencies in Supplementary Fig. 1a; identified origins and their relationship to previous studies in Supplementary Table 1).

To account for experimental variation, we analyzed four additional independent experiments and annotated how often each origin was identified (Supplementary Table 1). To independently visualize origins in a manner concordant with the literature23, we synchronized wild-type cells in G2, released them into S phase in the presence of hydroxyurea (HU) plus the nucleotide analog bromodeoxyuridine (BrdU) and quantified replication by using BrdU immunoprecipitation (BrdU-IP) plus deep sequencing (Fig. 3d). This identified 421 origins, >90% of which correspond to Pu-seq origins (Supplementary Table 1 and Supplementary Fig. 2).

**A map of replication timing by marker frequency analysis**

Although Pu-seq provides a direct assay for replication-initiation efficiency, it can also indirectly provide information about relative replication timing (described below). To validate replication-timing data calculated from the Pu-seq experiments, we first wished to generate a direct replication-timing map for *S. pombe* that is not biased by cell synchronization or treatment with replication inhibitors25. We thus mapped replication profiles of cells synchronized by elutriation, using marker frequency analysis (Fig. 4a). We examined aliquots of an elutriated culture over time for mitotic index, septation and DNA content. From these data, we calculated percentages of G2-phase, M-phase, S-phase and post-S phase cells for each time point, on the basis of the known cell-cycle behavior of *S. pombe* (Fig. 4b and Online Methods). We calculated the fraction of DNA replicated for each time point and set boundaries for the beginning and end of S phase (Fig. 4c). Next, we extracted DNA from the indicated aliquots spanning S phase and prepared libraries for deep sequencing. We compared the proportion of reads for each 1-kb bin across the genome to a fully replicated (G2) control and calculated the percentage of replication at each locus for each time point sequenced (Fig. 4d).

Because elutriation can cause cellular perturbation due to centrifugation26, we validated that elutriation did not distort replication profiles by performing fluorescence-activated cell sorting (FACS) and deep-sequence analysis (Sort-seq) in which S-phase cells
Figure 4 Genome replication timing in *S. pombe*. (a) Flow-cytometry profiles of cells synchronized in G2 by elutriation, washed into fresh media and allowed to progress through mitosis and into S phase. (b) Percentage of cells in G2, M, S and post-S phases. (c) Population-average genome copy number calculated for each time point. The period in which cells in the population are in S phase is shaded gray. (d) Visualization of DNA copy number during the S-phase time course across a representative region. White circles define origins.

Pu-seq provides timing and termination information

Mathematical analysis of the Pu-seq data provides a measure of replication timing: the proportion of reads mapping to each strand from the marker frequency analysis of the elutriated culture demonstrated a good correlation (Fig. 5a). This confirms that elutriation does not perturb replication timing.

Figure 5 Characterization of DNA-replication profiles. (a) Comparison of median replication time ($T_{rep}$, the time at which 50% of the locus is replicated) calculated from the synchronous culture by marker frequency analysis (Fig. 4d) in red and the normalized copy number of each locus from a single population of cells sorted by FACS from an asynchronous culture (Sort-seq) in blue. Open circles define origins. (b) Percentage of leftward-moving forks calculated from the Pu-seq data. (c) Comparison of $T_{rep}$ derived from marker frequency analysis and Sort-seq with $T_{rep}$ determined by Pu-seq. Red lines represent $T_{rep}$ calculated from the Pu-seq data. Light-blue line at top represents $T_{rep}$ calculated from the marker frequency analysis. Blue line at bottom shows the copy numbers derived from Sort-seq. (d) Calculated percentage of replication-termination events from the Pu-seq data for each locus. Supplementary data sets to visualize the whole genome are listed in Supplementary Table 2.

are recovered by FACS from an asynchronous culture and are subjected to deep sequencing. Plotting the normalized copy number for the Sort-seq against the calculated median replication time from the marker frequency analysis of the elutriated culture demonstrated a good correlation (Fig. 5a). This confirms that elutriation does not perturb replication timing.

Pu-seq provides timing and termination information

Mathematical analysis of the Pu-seq data provides a measure of replication timing: the proportion of reads mapping to each strand from the cdc6-L591G (Polδ) and cdc20-M630F (Polε) data sets provides two independent and direct measurements of the proportion of replication forks moving leftward (or rightward) throughout the genome (Fig. 5b). Such fork-direction data allow a direct calculation of relative replication times. From a mean replication-fork velocity of 1.5 kb/min, we calculated a relative replication-time map from Pu-seq data that is superimposable on direct replication-time measurements derived from the time course and Sort-seq analysis (Fig. 5c). Changes in mean fork direction across a chromosome are a consequence both of replication-origin activity and of replication-termination events: even close to an efficient origin, the proportion of moving forks always decreases with distance. This is the consequence of both the initiation and replication-termination
events in the population. We can thus also calculate the percentage of termination events occurring within a defined window. Although we observe that replication origins result in sharp transitions in fork direction, indicating discrete and efficient initiation sites, replication-termination events are dispersed stochastically across large termination zones (Fig. 5d), with no evidence of programmed termination regions (Supplementary Fig. 1b).

**Observed polymerase usage variation within a replicon**

Potential differences in the ribonucleotide incorporation rates between cdc20-M630F (Polε) and cdc6-L591G (Polδ) preclude accurate establishment of the absolute fraction of DNA synthesized by Polε and Polδ. Without considering the minor contribution of Polα, the anticipated division of labor plus coupled leading-strand and lagging-strand synthesis predicts that ~50% of the genome is replicated by Polδ and ~50% by Polδ. Using this assumption, we plotted polymerase usage of the duplex for each 300-bp bin across the genome (Fig. 6a). Genome wide, the division of labor was largely uniform, although small fluctuations were evident. The majority of these correspond to efficient origins. Therefore, we computationally identified interorigin regions of >30 kb where the directionality of replication forks was not appreciably perturbed by less efficient origins (Fig. 6b) and determined the average use of Polε and Polδ across replicons. A substantial bias toward Polδ was evident proximal to origins and declined toward the center of the interorigin region. This effect was not influenced by either global replication timing or by the absence of the Rad18 ubiquitin ligase (Supplementary Fig. 1c), which prevents PCNA ubiquitination and thus compromises noncanonical polymerase usage. Hence, when proximal to efficient origins, replicons exhibit an apparent bias toward usage of Polδ relative to Polε that is dependent on distance from the origin and is independent of postreplication repair.

**DISCUSSION**

We, along with other groups, have developed approaches to identify the genome-wide location of ribonucleotides incorporated into DNA. In a cdc20+cdc6+ (Polε+ and Polδ+); rnh201Δ background, we observed that the percentage of each ribonucleotide incorporated shows little bias relative to genomic sequence composition (Supplementary Fig. 3). This was not appreciably altered in the two polymerase-mutant backgrounds. We observed a moderate increase in the frequency of ribonucleotide incorporation in gene coding regions when compared to 5′ and 3′ untranslated regions and promoters, a bias that is not influenced by our polymerase mutations (Supplementary Fig. 4). Adaptations to this hydrolysis-dependent ribonucleotide mapping methodology will facilitate research into the causes of, and biological consequences arising from, ribonucleotide incorporation.

To study DNA replication, we combined this approach with ribonucleotide-discrimination mutations in the two main replicative polymerases to provide a Pu-seq strategy that allowed us to map polymerase usage genome wide. Our analysis demonstrated that the division of labor for Polε and Polδ is consistent across an entire genome. Although not unexpected, this is important to establish. Strikingly, Pu-seq provided a highly discriminatory data set that directly revealed the location and efficiency of replication origins at very high resolution. We compared our origin assignments to those previously collated from the literature in oriDB to locate potential overlap, we first identified the central nucleotide of the Pu-seq–identified origin and established whether it fell within ±900 bp of the reported origin region. Comparing the two data sets (741 origins from oriDB and 1,145 origins recognized by Pu-seq), we identified 97.5% of ‘confirmed’, 84.9% of ‘likely’ and 67.7% of ‘dubious’ oriDB origins (Supplementary Table 1).

Previous work in *S. cerevisiae* used replication-timing data to calculate termination frequencies across the genome and demonstrated that defined termination zones were not common: termination events per 1 kb fluctuated between approximately 0% and 4% per cell cycle across the genome. Applying this established mathematical analysis to the Pu-seq data similarly predicted that the distribution of termination frequencies in *S. pombe* is consistent with there not being defined termination zones between origins. This suggests that termination is largely defined by stochastic origin usage as opposed to the positioning of discrete replication-fork pausing elements.

The high definition provided by Pu-seq enabled us to identify an apparent bias toward Polδ close to the sites of efficient initiation, a phenomenon that is reproducible across multiple biological and experimental replicates (data not shown). This phenomenon is not influenced by either regional replication timing or by postreplication repair, thus implying that it is independent of noncanonical repair polymerases. Although we cannot exclude an unidentified prosaic explanation accounting for these data, one interpretation is that
a small fraction of leading-strand replication events, once started by Polε–primase, are initially extended by Polδ in place of Polε.

The interaction between the N-terminal 103 amino acids of the Dpb2 subunit of Polε and GINS is likely to position Polε for leading-strand synthesis. Although this same interaction is required for the formation of the CMG complex\(^1\), it is subsequently dispensable for CMG helicase activity, and loss of the interaction does not prevent replication progression if CMG formation is promoted by an ectopically expressed N-terminal region of Dpb1 (ref. \(^1\)). In such cells, replication is slow, and synthesis of the leading strand is probably completed by Polδ. Indeed, in yeasts, the entire genome can be replicated without the catalytic activity of Polε\(^1,36,37\), thus demonstrating substantial flexibility in the use of Polε and Polδ during DNA replication.

The choice of Polε for leading-strand synthesis is, at least in part, a function of the interaction of Polε helozenzyme and the core replication machinery discussed above. Polδ, although not apparently showing a strong interaction with the core replisome, does have a high affinity for PCNA and therefore potentially could compete for the leading-strand primer. Initiation of leading-strand synthesis by Polδ is likely to result in Polδ being subsequently displaced by Polε during elongation. Indeed, in vitro studies have shown that S. cerevisiae Polε helozenzyme is preferentially recruited to leading-strand substrates preloaded with CMG and that, although Polδ can load in the absence of Polε, it is displaced if Polε is added after DNA synthesis has initiated\(^38\). We thus propose that the apparent discrepancy in polymerase usage within a replicon reflects occasional recruitment of Polδ to leading-strand synthesis, with its subsequent displacement during progression by Polε. It will be interesting to test this proposition with further experiments in the future.

In summary, Pu-seq provides a simple yet powerful tool to explore genome replication in any eukaryote in which suitable polymerase mutants can be introduced in a background deficient (or depleted) for RNase H2. Unlike replication-timing data, Pu-seq data directly identify regions of replication initiation. We show here that it can also provide indirect but accurate evidence of relative replication timing and frequency of termination. Pu-seq will thus provide a useful tool for examining DNA replication.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Data files have been deposited in the Gene Expression Omnibus database under accession number GSE62108.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Murray and S. Mohebi for assistance with elutriation and T. Kunkel for advice on polymerase mutation. A.M.G. acknowledges UK Medical Research Council grant G1100074 and European Research Council grant 268788-SMI-DDR. C.A.N. acknowledges Biotechnology and Biological Sciences Research Council grant G1100074 and European Research Council grant 268788-SMI-DDR. Y.D. acknowledges a postdoctoral fellowship for research in human tumorigenesis. Nature 434, 864–870 (2005).

1. Bartkova, J. et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 434, 864–870 (2005).

2. Gorgoulis, V.G. et al. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. Nature 434, 907–913 (2005).

3. Rhind, N. & Gilbert, D.M. DNA replication timing. Cold Spring Harb. Perspect. Biol. 5, a010132 (2013).

4. Stillman, B. Origin recognition and the chromosome cycle. FEBS Lett. 579, 777–784 (2005).

5. Hawkins, M. et al. High-resolution replication profiles define the stochastic nature of genome replication initiation and termination. Cell Reports 5, 1132–1141 (2013).

6. Nakayama, T. et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. 20, 761–770 (2010).

7. O’Donnell, M., Langston, L. & Stillman, B. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. Cold Spring Harb. Perspect. Biol. 5, 1011013 (2013).

8. Moyer, S.E., Lewis, P.W. & Botchan, M.R. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc. Natl. Acad. Sci. USA 103, 10236–10241 (2006).

9. Grussius, A. et al. A key role for CHD in coupling the MCM2–7 helicase to DNA polymerase alpha within the eukaryotic replisome. EMBO J. 28, 2992–3004 (2009).

10. Simon, A.C. et al. A Ctf4 trimer couples the CMG helicase to DNA polymerase α at the eukaryotic replisome. Nature 510, 293–297 (2014).

11. Sengupta, S., van Deursen, F., de Piccoli, G. & Labib, K. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. Curr. Biol. 23, 535–542 (2013).

12. Handa, T., Kanke, M., Takashiki, T.S., Nakagawa, T. & Masukata, H. DNA polymerization-independent functions of DNA polymerase epsilon in assembly and progression of the replisome in fission yeast. Mol. Biol. Cell 23, 3240–3253 (2012).

13. Muramatsu, S., Hirai, K., Tak, Y.S., Kamimura, Y. & Araki, H. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol ε and GINS in budding yeast. Genes Dev. 24, 602–612 (2010).

14. Purser, E.F., Issaz, I., Lundstrom, E.B., Johansson, E. & Kunkel, T.A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. Science 317, 127–130 (2007).

15. Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M. & Kunkel, T.A. Division of labor at the eukaryotic replication fork. Mol. Cell 30, 137–144 (2008).

16. Miyabe, I., Kunkel, T.A. & Carr, A.M. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. PLoS Genet. 7, e1002407 (2011).

17. Nick McElhinny, S.A. et al. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc. Natl. Acad. Sci. USA 107, 4949–4954 (2010).

18. Sparks, J.L. et al. Rnase H2-initiated ribonucleotide excision repair. Mol. Cell 47, 980–986 (2012).

19. Kim, N. et al. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. Science 332, 1561–1564 (2011).

20. Williams, J.S. et al. Topoisomerase I-mediated removal of ribonucleotides from nascent leading-strand DNA. Mol. Cell 49, 1010–1015 (2013).

21. Sekiguchi, J. & Shuman, S. Site-specific ribonucleoside activity of eukaryotic DNA topoisomerase I. Mol. Cell 1, 89–97 (1997).

22. Watt, D.L., Johansson, E., Burgers, P.M. & Kunkel, T.A. Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. DNA Repair (Amst.) 10, 897–902 (2011).

23. Williams, J.S. & Kunkel, T.A. Ribonucleotides in DNA: origins, repair and consequences. DNA Repair (Amst.) 19, 27–37 (2014).

24. Lujan, S.A., Williams, J.S., Clausen, A.R., Clark, A.B. & Kunkel, T.A. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. Mol. Cell 50, 437–443 (2013).

25. Xu, J. et al. Genome-wide identification and characterization of replication origins by deep sequencing. Genome Biol. 13, R27 (2012).

26. Shiokazi, K., Shiokazi, M. & Russell, P. Heat stress activates fission yeast Spc1/StyI MAPK by a MEKK-independent mechanism. Mol. Biol. Cell 9, 1339–1349 (1998).

27. Möller, C.A. et al. The dynamics of genome replication using deep sequencing. Nucleic Acids Res. 42, e3 (2014).

28. Retsuke, R., Nieduszynski, C.A. & de Moura, A. Mathematical modeling of genome replication. Phys. Rev. E 86, 031916 (2012).

29. Clausen, A.R. et al. Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. Nat. Struct. Mol. Biol. doi:10.1038/nsmb.2957 (26 January 2015).

30. Koh, K.D., Balachander, S., Hesselberth, J.R. & Storici, F. Ribose-seq: global mapping of ribonucleotides embedded in genomic DNA. Nat. Methods doi:10.1038/nmeth.3259 (26 January 2015).
31. Reijns, M.A.M. et al. Lagging strand replication shapes the mutational landscape of the genome. Nature doi:10.1038/nature14183 (26 January 2015).
32. Siow, C.C., Nieduszynska, S.R., Muller, C.A. & Nieduszynski, C.A. OriDB, the DNA replication origin database updated and extended. Nucleic Acids Res. 40, D682–D686 (2012).
33. Fachinetti, D. et al. Replication termination at eukaryotic chromosomes is mediated by Top2 and occurs at genomic loci containing pausing elements. Mol. Cell 39, 595–605 (2010).
34. Lang, G.I. & Murray, A.W. Mutation rates across budding yeast chromosome VI are correlated with replication timing. Genome Biol. Evol. 3, 799–811 (2011).
35. Ulrich, H.D. & Takahashi, T. Readers of PCNA modifications. Chromosoma 122, 259–274 (2013).
36. Dua, R., Levy, D.L. & Campbell, J.L. Analysis of the essential functions of the C-terminal protein/protein interaction domain of Saccharomyces cerevisiae pol epsilon and its unexpected ability to support growth in the absence of the DNA polymerase domain. J. Biol. Chem. 274, 22283–22288 (1999).
37. Feng, W. & D’Urso, G. Schizosaccharomyces pombe cells lacking the amino-terminal catalytic domains of DNA polymerase epsilon are viable but require the DNA damage checkpoint control. Mol. Cell. Biol. 21, 4495–4504 (2001).
38. Georgescu, R.E. et al. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. Nat. Struct. Mol. Biol. 21, 664–670 (2014).
**ONLINE METHODS**

**Genetics and mutation.** Standard *S. pombe* genetic and molecular techniques were used as described previously. The cdc6-L591G (Polβ) mutant was constructed by site-directed mutagenesis and introduced into the *S. pombe* genome by recombination-mediated cassette exchange (RMCE). Southern blotting to detect alkali-sensitive sites in genomic DNA was performed as described previously. A list of strains used is given in Supplementary Table 3.

**Identification of a Polβ mutant that incorporates rNTPs.** DNA containing ribonucleotides is alkali labile; this causes strand fragmentation after alkali treatment. Exploiting an *rnh201*-null mutation (in which RNase H2 activity is missing), alkali degradation and Southern blot analysis were assessed for over 1,000 mutants. 

**Library production and sequencing.** Cells from early log phase IM642, IM855, IM654, YAK139 and YAK138 were harvested by centrifugation, and genomic DNA was prepared with a Qiagen 100/G Genomic-tip. For Pu-sequ analysis, 20 µg of genomic DNA was alkali treated in 0.23 M NaOH at 55 °C for 2 h. 10 µg of the single-stranded DNA (ssDNA) was loaded onto a 2% TBE gel and was run for 2 h at 100 V. The gel was stained with acridine orange (final concentration 5 µg/ml) for 2 h at room temperature with gentle shaking; this was followed by overnight destaining in water. Fragments of 300–500 bp were excised from the gel and isolated with a gel-extraction kit (Macherey-Nagel, Nucleospin Gel and PCR Clean-up). The experimental design for strand-directed high-throughput DNA sequencing was adapted from ref. 42: 100 ng of purified ssDNA fragments were converted to dsDNA with the BioPrime DNA Labeling system (Invitrogen) according to the manufacturer’s instructions, with dNTPs in which dTTP was substituted by dUTP. Converted dsDNA was purified with Ampure XP beads (Beckman Coulter), and concentration was determined by spectrometry (Pico green; Life Technologies) and size distribution examined with an Agilent Bioanalyzer. All DNA (20–60 ng) was used for Illumina library preparation with NEBNext Ultra DNA Library Prep Kit with the following modified protocol: end-cleaning of DNA fragments and adaptor ligation were performed as instructed by the manufacturer but without USER treatment and were followed by size-selection of insert (250–600 bp) with Ampure XP beads. Purified DNA was then treated with USER enzyme and subjected to subsequent PCR (13 cycles) with multiplexing index primers to generate Illumina libraries. After purification with Ampure XP beads, libraries were subjected to 100- or 150-bp paired-end sequencing with an Illumina HiSeq2500 or NextSeq 500 platform, respectively.

**Analysis of polymerase usage.** Paired-end reads of high-throughput sequencing were aligned to the *S. pombe* genome sequence (ASM294v2.23: chromosomes I, II and III, downloaded from PomBase, http://www.pombase.org/) with bowtie2-2.2.2. With alignment data, the position of the 5′ end of each R1 read, which corresponds to the 5′ end of ssDNA hydrolyzed by alkali treatment, was determined, and the number of reads in 300-bp bins across the genome were counted separately for the Watson and Crick strands. This generated the four data sets: at the chromosome coordinate x, $N_\delta(x)$, the count for cdc6–L591G (Polβ) on the Watson strand; $N'_\delta(x)$, for cdc6–L591G (Polβ) on the Crick strand; $N'_\epsilon(x)$, for cdc20–M630F (Polε) on the Watson strand; and $N'_\epsilon(x)$, for cdc20–M630F (Polε) on the Crick strand. The data sets were normalized with the total number of reads: $D_\delta(x) = N_\delta(x)/\Sigma N_\delta$, for the Polδ mutant on the Watson strand; $D_\delta(x) = N_\delta(x)/\Sigma N_\delta$, for the Polδ mutant on the Crick strand; $D_\epsilon(x) = N'_\epsilon(x)/\Sigma N_\epsilon$ for the Polε mutant on the Watson strand; and $D_\epsilon(x) = N'_\epsilon(x)/\Sigma N_\epsilon$ for the Polε mutant on the Crick strand. Making the assumption that each part of the duplex genome is replicated by Polδ and Polε, the ratio of DNA synthesis catalyzed by Polδ (D) and Polε (E) were calculated as: $D_\epsilon(x) = D_\delta(x)/D_\epsilon(x) + E_\epsilon(x)$, ratio of Polδ synthesis on the Watson strand; $E_\epsilon(x) = E_\delta(x)/D_\delta(x) + E_\epsilon(x)$, ratio of Polε synthesis on the Watson strand; $D_\epsilon(x) = D_\delta(x)/D_\epsilon(x) + E_\epsilon(x)$, ratio of Polε synthesis on the Crick strand; and $E_\epsilon(x) = E_\delta(x)/D_\delta(x) + E_\epsilon(x)$, ratio of Polδ synthesis on the Crick strand. With the assumption that 50% of the genome is replicated by Polδ and 50% by Polε, the ratios of Pol usage were optimized: when n is the total number of bins, $D_\epsilon(x) = D_\delta(x)/n \times \Sigma D_\epsilon(x)+D_\epsilon(x)$; $D_\epsilon(x) = D_\delta(x)/n \times \Sigma D_\epsilon(x)+D_\epsilon(x)$; and $E_\epsilon(x) = E_\delta(x)/n \times \Sigma E_\epsilon(x)+E_\epsilon(x)$.

**Identification of replication origins.** Custom R scripts (available on request) were used to identify origins: polymerase usage ratio data from each strand (calculated without the assumption that 50% of the genome is replicated by Polδ and 50% by Polε) were smoothed with a moving average of 3; the data point for each bin is an average of seven points: the point at the origin and the three points either side. Computational analysis was performed with the Apollo cluster computer at the University of Sussex.

**Mapping origins by BrdU ChIP-seq.** YD18 cells were grown to exponential phase (0.2 × 10^6/ml) at 25 °C and synchronized at G2 phase by incubation at 36 °C for 3.5 h. After addition of bromodeoxyuridine (0.5 µM) and hydroxyurea (10 mM), cells were further incubated at 25 °C for 90 min, and 1 × 10^5 cells were pelleted by centrifugation and were subjected to genomic DNA extraction. Subsequently Brdu-IP was performed as described in ref. 25.

**Replication timing by marker frequency analysis.** Cells (strain 501) were synchronized in G2 by elutriation (considered the least physiologically stressful method of synchronization for fission yeast), concentrated into a volume of 200 ml and grown in fresh medium at 27 °C. Samples were taken at 5-min intervals through S phase and were analyzed for DNA content by flow cytomtery and for mitotic index and separation by staining with DAPI and Calcofluor. The population-averaged fraction of the genome replicated at each time point was calculated from flow-cytometry and separation-index data. During flow-cytometry sample preparation, post–S phase S. pombe cells can separate. Consequently, during early time points after elutriation the 2N peak (in flow-cytometry data) is predominantly pre–S phase cells, but in later time points the 2N peak starts to include post–S phase cells. We determined the proportion of pre–S phase (G2 and M), S-phase and post–S phase (septum pinched in or two cells together) cells from the separation data. Flow cytometry was used to quantify the fraction of cells in the 2N peak with a DNA content greater than 2N. Then the separation-index data were used to determine the proportion of the 2N peak that represented post–S phase cells. Briefly, if the proportion of G2- and M-phase cells (separation data) was less than the proportion of cells in the 2N peak, this difference could be attributed to either very early S-phase or post–S phase cells. The post–S phase separation data allowed us to distinguish between these alternatives. In early time points (20–85 min), the small proportion of post–S phase cells (57%) were assumed to contribute to the 2N peak. In later time points (90–120 min), the small proportion of G2- and M-phase cells (≤10%) were used to infer that the remaining cells in the 2N peak were post S phase. Once the proportion of pre– and post–S phase cells in the 2N peak has been estimated, the flow-cytometry data were used to determine the population-averaged fraction of the genome replicated at each time point. The DNA-content signal from the 2N peak was assumed to correspond to a haploid genome content (copy number 1) and the signal from the 4N peak to a diploid genome content (copy number 2). This permitted calculation of the relative population-averaged genome copy number throughout the time course. The reference sample was taken prereplication, 45 min after elutriation.
DNA was prepared from the elutriated reference sample and samples from within S phase, and libraries were prepared and subjected to high-throughput sequencing as previously described\textsuperscript{27}. The relative representation of each locus in the S-phase samples was normalized to the percentage of total replication and to the unreplicated reference sample to provide an average percentage replication for each locus for each time point. To provide an unbiased replication-timing map, S-phase cells from an unperturbed exponentially growing culture were collected by FACS after fixation with 70% ethanol and were subjected to marker frequency analysis with the Sort-seq protocol previously described\textsuperscript{27}).

**Calculation of relative and absolute replication timing.** The time-course data were used to calculate a median absolute replication time ($T_{\text{rep}}$) for each genomic locus as described previously\textsuperscript{27}. Briefly, a sigmoidal function was fitted to the population-averaged fraction of the genome replicated at each time point for each genomic locus, and $T_{\text{rep}}$ was determined as a time when the population-averaged fraction of the genome replicated was equal to 0.5. Times are shown relative to the approximate start of S phase, 50 min post elutriation. Relative replication times and the distribution of replication-termination sites were calculated from the Pu-seq fork-direction data with custom scripts described previously\textsuperscript{5,28}. Briefly, relative replication time was calculated from the integral of the percentage of leftward-moving forks, assuming a constant average fork velocity across the genome. Termination frequency was calculated with a finite difference approximation for estimating the derivative of the percentage of leftward-moving forks.

39. Moreno, S., Klar, A. & Nurse, P. Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. Methods Enzymol. 194, 795–823 (1991).
40. Watson, A.T., Garcia, V., Bone, N., Carr, A.M. & Armstrong, J. Gene tagging and gene replacement using recombinase-mediated cassette exchange in Schizosaccharomyces pombe. Gene 407, 63–74 (2008).
41. Lipkin, D., Talbert, P.T. & Cohn, M. The mechanism of the alkaline hydrolysis of ribonucleic acids. J. Am. Chem. Soc. 76, 2871–2872 (1954).
42. Zhang, Z., Theurkauf, W.E., Weng, Z. & Zamore, P.D. Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence 3, 9 (2012).