



Published in final edited form as:

Nat Struct Mol Biol. 2015 March ; 22(3): 185–191. doi:10.1038/nsmb.2957.

Tracking replication enzymology *in vivo* by genome-wide mapping of ribonucleotide incorporation

Anders R. Clausen¹, Scott A. Lujan¹, Adam B. Burkholder², Clinton D. Orebaugh¹, Jessica S. Williams¹, Maryam F. Clausen³, Ewa P. Malc³, Piotr A. Mieczkowski³, David C. Fargo², Duncan J. Smith⁴, and Thomas A. Kunkel¹

¹Laboratory of Molecular Genetics and Laboratory of Structural Biology, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina

²Integrative Bioinformatics, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina

³Department of Genetics, High Throughput Sequencing Facility, University of North Carolina, Chapel Hill, North Carolina

⁴Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York

Abstract

Ribonucleotides are frequently incorporated into DNA during eukaryotic replication. Here we map the genome-wide distribution of these ribonucleotides as markers of replication enzymology in budding yeast, using a new 5'-DNA end-mapping method, **Hydrolytic End Sequencing**. HydEn-Seq of DNA from ribonucleotide excision repair-deficient strains reveals replicase- and strand-specific patterns of ribonucleotides in the nuclear genome. These patterns support the role of DNA polymerases α and δ in lagging strand replication and of DNA polymerase ϵ in leading strand replication. They identify replication origins, termination zones and variations in ribonucleotide incorporation frequency across the genome that exceed three orders of magnitude. HydEn-Seq also reveals strand-specific 5'-DNA ends at mitochondrial replication origins, suggesting unidirectional replication of a circular genome. Given the conservation of enzymes that incorporate and process ribonucleotides in DNA, HydEn-Seq can be used to track replication enzymology in other organisms.

Reprints and permission information is available at www.nature.com/reprints

Corresponding author: Phone: 919-541-2644, kunkel@niehs.nih.gov.

ACCESSION CODES

Sequencing data have been deposited in the Gene Expression Omnibus under accession number GSE62181.

AUTHOR CONTRIBUTIONS

ARC, DJS and TAK designed the experiments, ARC, CDO, JSW, MFC and EPM performed the experiments, ARC, SAL, ABB, DCF, PAM and TAK analyzed the data, TAK wrote the manuscript, and all authors edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests. The authors declare no competing financial interests.

Keywords

ribonucleotide incorporation; replication origins; polymerase switching; mitochondrial replication; biomarkers

INTRODUCTION

Among the many eukaryotic DNA polymerases (Pols), e.g., 17 in humans and eight in budding yeast, three replicate the bulk of the nuclear genome^{1,2}. Synthesis at replication origins is initiated when an RNA primase synthesizes an RNA primer that is extended by limited DNA synthesis by Pol α ³. Pol ϵ is then proposed to catalyze the majority of leading strand replication⁴⁻⁶ in a largely continuous manner. In contrast, the nascent lagging strand is synthesized as a series of ~180 nucleotide Okazaki fragments that are initiated by RNA primase followed by limited synthesis by Pol α . This is followed by extensive synthesis catalyzed by Pol δ ⁶⁻⁸ and subsequent maturation of Okazaki fragments into a continuous nascent lagging strand⁹. The exact locations of polymerase switching during leading and lagging strand replication are under investigation^{10,11} but remain uncertain. Equally uncertain is polymerase use after replication forks encounter difficult circumstances that may require switching to a different replicase or a more specialized DNA polymerase, e.g., to copy unusual DNA sequences or to bypass lesions^{12,13}. Replication enzymology differs for the mitochondrial genome, where both DNA strands are replicated by the same replicase, Pol γ , by mechanisms that also remain uncertain¹⁴⁻¹⁶.

We have been using mutator alleles of yeast Pols ϵ , α and δ to infer their roles in nuclear DNA replication *in vivo*. These mutator alleles, *pol2-M644G* (Pol ϵ), *pol1-L868M* (Pol α) and *pol3-L612M* (Pol δ), generate single base replication errors at higher rates than their wild type parents. In the absence of mismatch repair (MMR), these errors remain in the genome and mark where each replicase synthesized DNA during replication. The results (see⁶ and references therein) imply that in unstressed yeast cells, Pol ϵ is the primary leading strand replicase and Pols α and δ are primarily responsible for lagging strand replication. However, the resolution of this approach for tracking replication enzymology *in vivo* is limited by the high fidelity of replication. For example, the average genome-wide replication error rates of the mutator replicases are $1-2 \times 10^{-7}$ ⁶, such that single base replication errors are low-density markers of replication enzymology.

In the present study, we set out to track replication enzymology *in vivo* at much higher resolution using ribonucleotides rather than mutations. This approach takes advantage of several facts. The presence of an oxygen atom on the 2'-position of a ribose increases the sensitivity of the phosphodiester bond in nucleic acids to alkaline hydrolysis by five orders of magnitude. The active sites of Pols α , δ and ϵ can be engineered to increase the probability of ribonucleotide incorporation into DNA, to frequencies as high as 10^{-2} to 10^{-3} ¹⁷. Disabling Ribonucleotide Excision Repair (RER) prevents removal of ribonucleotides from both the nascent leading strand and the nascent lagging strand¹⁷⁻¹⁹. RER defective yeast cells are viable, including those encoding replicases that are promiscuous for ribonucleotide incorporation. These facts led us to propose²⁰ that

ribonucleotides can be used as high-density markers of DNA polymerization reactions *in vivo*. Here we demonstrate that this is indeed the case, using a newly developed method to map ribonucleotides in the yeast genome at single nucleotide resolution. Initial results support the strand assignments for the nuclear replicases, confirm nuclear replication origins and identify new origins, reveal the locations of replication termination zones, quantify ribonucleotide incorporation for each of the four bases, establish that the distribution of ribonucleotides across the genome is non-uniform, and provide new information that is likely to be relevant to mitochondrial DNA replication.

RESULTS

Genome-wide mapping of ribonucleotides in DNA by HydEn-Seq

The new genome-wide mapping method, which we call HydEn-Seq (for Hydrolytic DNA End Sequencing (Fig. 1a, Supplementary Table 1), has been used to map ribonucleotides in five pairs of RER-deficient (*rnh201*⁻) versus RER-proficient (*RNH201*) yeast strains (Supplementary Table 2). One pair encodes wild type Pols α , δ and ϵ . A second pair encodes *pol2-M644L*, a Pol ϵ variant that incorporates fewer ribonucleotides than does a wild type strain¹⁸. A third pair encodes *pol2-M644G*, a Pol ϵ variant that is promiscuous for ribonucleotide incorporation^{18,21}. A fourth pair encodes a *pol3-L612G* variant in which leucine 612 in the Pol δ active site is replaced with glycine, based on the prediction that like the analogous *pol2-M644G* (Pol ϵ) variant¹⁸, the *pol3-L612G* variant would be even more promiscuous for ribonucleotide incorporation than our previously studied *pol3-L612M* allele²². The fifth pair encodes a *pol1-Y869A* variant with alanine substituted for the “steric gate” tyrosine in the Pol α active site that normally prevents ribonucleotide incorporation²³. This allele is used to increase the frequency of ribonucleotide incorporation by Pol α over that observed in our previously studied *pol1-L868M* variant¹⁷.

Alkaline hydrolysis of genomic DNA followed by electrophoresis in an alkaline agarose gel reveals that the genomes of all five *rnh201*⁻ mutant strains contain more alkali-sensitive sites than their *RNH201*⁺ parents (Fig. 1b). Importantly, the genomes of the double mutant *pol1-Y869A rnh201*⁻, *pol2-M644G rnh201*⁻ and *pol3-L612G rnh201*⁻ strains contain many more alkali-sensitive sites than the strains with either single mutation alone (Fig. 1b,c,d), such that most of the 5'-DNA ends in these strains result from alkaline hydrolysis of ribonucleotides incorporated during replication by the variant derivatives of Pols α , δ or ϵ , respectively. This contrasts with the *pol2-M644L rnh201*⁻ mutant strain, which harbors fewer ribonucleotides than the other *rnh201*⁻ strains with variant replicases.

The locations of the 5'-DNA ends in the genomes of these strains were mapped by HydEn-Seq (Fig. 1a). Genomic DNA samples were hydrolyzed with 0.3M KOH²⁴, libraries were prepared from the resulting single-stranded DNA fragments, and 50 base, paired-end sequencing was performed on an Illumina HiSeq2500 instrument to identify the location of the 5'-DNA ends. Ribonucleotides are located immediately adjacent to the 5'-DNA ends (Fig. 1a). Two or more independent libraries were analyzed for each strain (Supplementary Table 3), with replicate libraries yielding similar results (Supplementary Table 4). Alignment of the fragments to a well-annotated reference genome⁶ identified the DNA strand to which fragments align, and the location and identity of ribonucleotides in the

genome. Read counts, scaled using the number of 5'-ends at the ends of chromosomes (Methods), confirm the relative ribonucleotide densities anticipated by agarose gel electrophoresis.

Strand specificity and origin identification

DNA fragments from the *pol2-M644G rnh201* strain align with the two DNA strands in the nuclear genome in an alternating pattern complementary to alignments for the *pol1-Y869A rnh201* and *pol3-L612G rnh20* strains (Fig. 2, chromosome 10; Supplementary Fig. 1, 5'-DNA end read counts corresponding the Fig. 2, bottom; Fig. 3, all 16 chromosomes; Fig. 4a, heat maps). In contrast, a strand-specific pattern is not observed in the *pol2-M644L rnh201* strain (Fig. 4a) or in RER-proficient strains (Supplementary Fig. 2). Thus the majority of 5'-DNA ends in the *pol2-M644G rnh201*, *pol1-Y869A rnh201* and *pol3-L612G rnh20* strains are due to ribonucleotides incorporated during replication that are not removed because RER is defective. Comparing ribonucleotide maps in these three strains reveals numerous strand-specific transitions (diamonds in Figs. 2 and 3). Among these are 294 transitions that correspond to confirmed replication origins in the yeast origin database²⁵. Transitions are also observed at 72 locations (Fig. 3, listed in Supplementary Table 5) that have not yet been reported to be origins, but may be origins that are used in some cells in the population.

The ribonucleotide maps in the three *rnh201* strains encoding the variant replicases strongly support earlier interpretations based on replication errors⁴⁻⁷, that Pol ϵ synthesizes the majority of the nascent leading strand and Pols α and δ synthesize the majority of the nascent lagging strand of the budding yeast nuclear genome. Thus HydEn-Seq confirms a fundamental aspect of yeast replication enzymology. The evolutionary conservation among eukaryotic nuclear replicases and among type 2 RNases H in all three kingdoms of life (Supplementary Figure 3), suggests that the HydEn-Seq strategy used here may be applicable to tracking replication enzymology and identifying origins in other organisms. The ribonucleotide map in *pol1-Y869A rnh201* strain further demonstrates that when RER is deficient, some DNA synthesized by Y869A Pol α survives Okazaki fragment maturation and resides in the mature lagging strand. This same conclusion was reached in earlier studies^{6,7,26-28} that monitored replication errors rather than ribonucleotides using *pol3-L612M* strains that were either proficient or deficient in MMR but were RER proficient. It remains to be determined if DNA synthesized by Pol α survives Okazaki fragment maturation in wild type yeast.

Polymerase use at replication origins and termination zones

Heat maps (Fig. 4a) and meta-analyses of 5'-DNA ends in 50 base pair bins (Fig. 4b) reveal where strand switches at origins occur in all three replicase variant backgrounds. These transitions occur over several hundred base pairs centered on the Autonomously Replicating Sequences (ARS) Consensus Sequences (ACS, orange line in Fig. 4b). The results in the *pol1-Y869A rnh201* strain are consistent with a role for Pol α in initiating synthesis on both strands at origins. The breadth of the strand transition at origins in this strain suggests that in a cell population, initiation occurs within a zone rather than at a single base pair. Deeper coverage of 5'-DNA ends in the future should allow higher resolution mapping, to

investigate whether initiation occurs at a single base pair at some origins, as previously reported for one origin on chromosome 4²⁹. The strand transition at origins is much sharper in the *pol2-M644G rnh201* strain than in the *pol3-L612G rnh201* strain. Investigating this difference may eventually provide information that complements recent biochemical studies of initiation of leading and lagging strand replication (see² and references therein).

HydEn-Seq also reveals where mergers occur between forks arriving from adjacent origins and moving in opposite directions. The results suggest that termination occurs in zones that vary in location and breadth. In some cases (e.g., Fig. 2a, bottom right), the termination zone is broad and equidistant from adjacent origins, while in other cases (e.g., Fig. 2a, bottom left) the zone is narrower and or closer to one origin than the other. HydEn-Seq offers the opportunity to explore the mechanisms and genetic controls underlying these variations.

Ribonucleotide incorporation in wild type yeast

Studies of ribonucleotide incorporation *in vitro* by wild type Pols α , δ and ϵ predict that there should be 2.3 times more ribonucleotides incorporated into the nascent leading strand as compared to the nascent lagging strand²⁴. This prediction is supported by results in the RER-defective (*rnh201*) strain encoding wild type replicases. In this strain, the strand-specific heat map (Fig. 4a) and the transition from one strand to the other as analyzed by meta-analysis (Fig. 4b) match those of the *pol2-M644G rnh201* strain, and are opposite to those in the *pol3-L612G 201* or *pol1-Y869A rnh201* strains.

The observation that ribonucleotides are preferentially incorporated into the nascent leading strand in the strain encoding wild type replicases is relevant to the genome instability reported in the wild type replicase background when RER is defective. In this strain³⁰, the specificity of 2 to 5 base pair deletion mutations resulting from topoisomerase1 (Top1) cleavage at unrepaired ribonucleotides is indistinguishable from the 2 to 5 base pair deletion specificity in the *pol2-M644G rnh201* strain¹⁸ that primarily contains ribonucleotides in the nascent leading strand^{5,21,22}. The fact that ribonucleotides preferentially map to the nascent leading strand is also relevant to recent studies^{22,31} suggesting that nicks generated by RNase H2 at ribonucleotides in the continuously replicated nascent leading strand may direct mismatch repair (MMR) to correct replication errors in that strand. This idea, when combined with the non-uniform distribution of ribonucleotides in the genome discussed below, implies that the potential contribution of this MMR signaling mechanism may vary across the genome. The preferential presence of ribonucleotides in the nascent leading strand may also be relevant to other suggested signaling functions for ribonucleotides in DNA^{20,24}.

Variations in ribonucleotide incorporation by base identity

Wild type Pols α , δ and ϵ have different preferences for incorporating each of the four different ribonucleotides *in vitro*²⁴. To determine if this is also true during replication *in vivo*, we analyzed fragments close to replication origins where (as explained previously⁶) leading and lagging strand assignments can be made with the greatest confidence. Despite the fact that the 12 million base pair budding yeast genome is 62% A+T, the most abundant ribonucleotide present in the genome of the *pol2-M644G rnh201* strain is rC, followed by

rG, then rA and then rU (Fig. 5). These preferences re-capitulate the rank order for ribonucleotide incorporation by M644G Pol ϵ *in vitro*²⁴. The same rank order for ribonucleotide incorporation (rC > rG > rA > rU) is observed in the *pol3-L612G rnh201* strain, but the proportions of the four rNTPs incorporated are different. For example, more rC and less rU are present in the *pol3-L612G rnh201* genome as compared to the *pol2-M644G rnh201* genome (Fig. 5). A different rank order is seen in the *pol1-Y869A rnh201* strain, where after correcting for genome composition, the preference in the *pol1-Y869A* strains is rA \approx rC \approx rG > rU. In this strain, the non-rU rankings changed slightly among the three replicates examined.

The low abundance of rU seen in all three genomes is consistent with the fact that, among the four dNTPs, dTTP is present at the highest concentration in strains encoding either wild type replicases²⁴ or the *pol2-M644G* variant¹⁸, thereby reducing the probability of incorporating rU more than the other ribonucleotides. However, the dATP:rATP ratio is the lowest among the four ratios²⁴, yet rATP is only the most frequent ribonucleotide in one of the three strains. Thus, in addition to competition for incorporation within the polymerase active site based on mass action, other parameters may modulate ribonucleotide incorporation probability during replication *in vivo*. This includes the effect of DNA sequence context, as predicted by sequence context effects of ribonucleotide incorporation probability during DNA synthesis *in vitro*^{24,32,33}.

Non-uniform distribution of ribonucleotide in the genome

Several HydEn-Seq libraries contain an average of less than one 5'-DNA end read per base pair in the nuclear genome (Supplementary Table 3). It is therefore striking that end read counts vary from zero at many base pairs to more than 1,000 at others. This non-uniform distribution of ribonucleotides in the genome has implications for MMR signaling mentioned above, and for a second mechanism of genome instability wherein Top1 incises ribonucleotides in DNA to initiate the deletion of 2 to 5 base pairs within repetitive sequences^{18,34}. This instability is highly dependent on the DNA strand and sequence context in which the ribonucleotide resides¹⁸. Variations in the location and density of ribonucleotides in DNA may also be relevant to recombination³⁵ and gross chromosomal rearrangements in yeast³⁶ and to chromosomal abnormalities in RNase H2-defective mouse cells^{37,38}.

In certain regions of the genome, strand-specific ribonucleotide density also deviates from the expectations of a simple division of labor among the three replicases. Initial analyses indicate that these “excursions” fall into at least two classes, those that show unexpected Pol α , δ or ϵ correspondence (e.g., purple bar designated III in Fig. 2a, middle) and those that show unexpected Pol α or δ divergence (designated IV). These excursions may result from ribonucleotides remaining in the genomes of these RER-deficient cells, which can lead to replicase pausing during DNA synthesis^{18,39,40}. Such pausing may elicit template switching or bypass synthesis (e.g., Pol ζ or Pol η) in a subsequent round of replication⁴¹ or DNA synthesis associated with DNA repair or recombination after Top1 incision at ribonucleotides^{21,34,35}. An example is in *Schizosaccharomyces pombe*, where mating type switching occurs by recombination posited to be initiated by pausing of leading strand

replication upon encountering a di-ribonucleotide imprint⁴². Additional possibilities for some “excursions” detected by HydEn-Seq include events unrelated to ribonucleotides, such as encounters of replication forks with transcription complexes, a bulky lesion, repetitive DNA or non-B form DNA or tightly bound proteins. There is no obvious limitation to monitoring the distance over which a newly recruited DNA polymerase may operate, e.g., within a short repair or lesion bypass patch or to the end of a chromosome during break-induced recombination⁴³.

Ribonucleotide distribution relative to nucleosome dyads

Meta-analysis using nucleosome positioning data⁶ reveals that ribonucleotide densities are elevated at positions corresponding to the nucleosome dyad (Fig. 6a–d). The elevations are subtle (note the scale on the Y axis). They may partly reflect a bias in sequence composition because nucleosome dyads are slightly enriched for G and C content (Fig. 6e), the preferred ribonucleotides incorporated during replication in the *pol2-M644G rnh201* and *pol3-L612G rnh201* strains. However, this may not be the sole explanation because (1) the peaks at the dyad are more prominent in the *pol3-L612G rnh201* strain (Fig. 6d) as compared to the *pol2-M644G rnh201* strain (Fig. 6b), yet these two strains have similar G +C versus A+T ribonucleotide incorporation preferences (Fig. 5), (2) the peaks are more prominent in the *pol3-L612G rnh201* and *pol1-Y869A rnh201* strains as compared to the *pol2-M644G rnh201* strain, and (3) the peak in both strands in the *pol3-L612G rnh201* strain are symmetrical around the dyad, whereas the peaks in the two strands in the *pol1-Y869A rnh201* strain are offset and on opposite sides of the dyad. In the latter cases, lagging strand replicase features could be signatures of polymerization by Pol α and Pol δ during Okazaki fragment maturation, a process that is proposed to preferentially occur at the nucleosome dyad and to be phased according to the nucleosome repeat¹⁰.

Ribonucleotides at mitochondrial DNA replication origins

HydEn-Seq also reveals that the yeast mitochondrial genome contains large numbers of 5'-DNA ends generated by alkaline hydrolysis (Fig. 7). Most of these ends are in discrete, strand-specific (red and blue) peaks that span multiple base pairs. Eight of these peaks correspond to previously identified⁴⁴ mitochondrial replication origins (shaded green). Interestingly, the relative proportions of 5'-DNA ends in the major peaks are similar in all yeast strains examined. Thus the peaks are independent of the status of the nuclear replicases, which have no known role in mitochondrial replication, and they are also independent of the status of RNase H2, which has not been found in mitochondria⁴⁵.

These observations are consistent with at least three hypotheses. The peaks may represent the ends of linear chromosomes, similar to the high density of 5'-ends observed at the ends of linear nuclear chromosomes. This possibility cannot yet be eliminated, but it seems unlikely because the most prominent peaks largely map to either the plus or minus strand, but not to both. Also, when the mitochondrial genome is rearranged *in silico* (Methods) to join the chromosome “ends” that were arbitrarily assigned and numbered when the genome was sequenced⁴⁴, no drop in the depth of coverage of fragments is observed at the junction relative to immediately adjacent regions. Thus, like mammalian mitochondrial genomes, the

S. cerevisiae mitochondrial genome may largely, albeit not necessarily exclusively, be circular.

The second hypothesis is that some mtDNA fragments generated may be due to lesions other than ribonucleotides, such as strand-specific nicks or alkali-sensitive abasic sites resulting from oxidative stress. We cannot exclude this possibility, but it is currently disfavored by the fact that the 5'-DNA ends are distributed in a highly non-uniform and largely strand-specific manner.

The third hypothesis stems from previous studies showing that mammalian mtDNA contains ribonucleotides^{46,47}, and that human mitochondrial replicase (Pol γ) incorporates ribonucleotides during DNA synthesis *in vitro*⁴⁸. Moreover, eight of the most prominent, strand-biased 5'-DNA end peaks in the yeast mitochondrial genome correspond to previously identified⁴⁴ mitochondrial replication origins. These peaks, and perhaps similar strand-specific peaks detected within open reading frames and in sequences encoding RNAs, could reflect the presence of unrepaired residues of RNA primers made by mtRNA polymerase and used to initiate mtDNA replication, as has been reported in mammalian cells^{16,49–52}. The results suggest that the terminal ribonucleotides of RNA primers for mtDNA replication may not always be removed, either by mitochondrial RNase H1⁵³, which cannot incise at an RNA-DNA junction⁴⁵, or by strand displacement and flap cleavage, as for Okazaki fragment maturation during nuclear DNA replication⁹. If this explanation holds, then the fact that 5'-DNA ends at the origins of mtDNA replication preferentially map to one strand or the other favors a unidirectional replication model for mtDNA in budding yeast.

DISCUSSION

This study demonstrates that ribonucleotides can be used to track replication enzymology at high resolution using a simple, 5-step library preparation procedure involving minimal use of enzymes and requiring less than two days to execute. While HydEn-seq is used here to map 5'-DNA ends primarily generated by alkaline hydrolysis at ribonucleotides, it can also be used to study other lesions in DNA, and it is not limited to spontaneous chemical hydrolysis but can be adapted to map 5'- and 3'-DNA ends generated by enzymatic hydrolysis. In addition to normal replication enzymology, HydEn-Seq should be useful to study polymerization changes in response to endogenous and exogenous environmental stress. The ability of HydEn-Seq to identify replication origins, termination zones and polymerase usage during replication should be applicable to other organisms in which replicases can be engineered to enhance ribonucleotide incorporation and RER can be inactivated. Polymerase structure-function studies have advanced to the point where it is now feasible to engineer replicases (e.g., see Supplementary Figure 3a), and more specialized polymerases in most polymerase families, to retain catalytic efficiency yet render them promiscuous for ribonucleotide incorporation. Theoretically, this may permit a variety of DNA synthesis reactions in cells to be studied by HydEn-Seq.

Following the idea that the high-density peaks in the mitochondrial genome may be due to unrepaired residues of RNA primers made by mtRNA polymerase, HydEn-Seq may also be

useful to study RNA primers synthesized by RNA polymerases, RNA primases or Prim-Pols. The ability to map genomic locations that contain a high density of ribonucleotides can be used to explore the idea that ribonucleotides in DNA provide selective advantages to cells²⁴. Relevant here are ribonucleotides that may persist in certain locations even in RER-proficient cells, as exemplified by the di-ribonucleotide imprint used for mating type switching in *Schizosaccharomyces pombe*⁵⁴.

ONLINE METHODS

Materials

Oligonucleotides and yeast strains used in this study are listed in Supplementary Tables 1 and 2, respectively. The *pol1-Y869A* and *pol3-L612G* strains and their *rmh201* derivatives were constructed as described earlier for *pol1-L868M* and *pol3-L612M* strains⁷.

HydEn-Seq protocol

Yeast strains were grown to mid log phase ($OD_{600}=0.6$) at 30°C in YPDA medium supplemented with 0.25 mg/ml adenine. DNA was isolated using the MasterPure™ Yeast DNA Purification Kit (Epicentre) without RNase A treatment. HydEn-Seq (Fig. 1) was performed by hydrolyzing one µg of genomic DNA with 0.3 M KOH for 2 hours at 55°C²⁴. Following ethanol-precipitation, the DNA fragments were treated for three minutes at 85°C, phosphorylated with 10 units of 3'-phosphatase-minus T4 polynucleotide kinase (New England Biolabs) for 30 minutes at 37°C, heat inactivated for 20 minutes at 65°C and purified using HighPrep™ PCR beads (MagBio). Phosphorylated products were treated for three minutes at 85°C, ligated to oligo ARC140 (Supplementary Table 1) overnight at room temperature using 10 units of T4 RNA ligase, 25% PEG8000 and 1 mM $CoCl_3(NH_3)_6$, and purified using HighPrep™ PCR beads (MagBio). Ligated products were treated for 3 minutes at 85°C. The ARC76–ARC77 adapter was annealed to the second strand for five minutes at room temperature. The second strand was synthesized using four units of T7 DNA polymerase (New England Biolabs) and purified using HighPrep™ PCR beads (MagBio). Libraries were PCR amplified using primer ARC49 and primer ARC79 or ARC84 to ARC107, using KAPA HiFi Hotstart ReadyMix (KAPA Biosystems). Libraries were then purified using HighPrep™ PCR beads (MagBio) and pooled for sequence analysis. Paired-end sequencing was performed on a HiSeq2500 sequencer (Illumina) to identify the location of the 5'-DNA ends generated by alkaline hydrolysis.

HydEn-Seq trimming, filtering and alignment

All reads were trimmed for quality and adapter sequence using cutadapt 1.2.1 (-m 15 -q 10 -- match-read-wildcards)⁵⁵. Pairs with one or both reads shorter than 15 nucleotides were discarded. Mate 1 of the remaining pairs was aligned to an index containing the sequence of all oligos utilized in the preparation of these libraries using bowtie 0.12.8 (-m1 -v2), and all pairs with successful alignments were discarded. Pairs passing this filter were subsequently aligned to the L03 *S. cerevisiae* reference genome⁶ (-m1 -v2 -X10000 --best). Single-end alignments were then performed using mate 1 of all unaligned pairs (-m1 -v2). The count of 5' ends of all unique paired-end and single-end alignments were determined for all samples, per-strand, across all chromosomes, combining all technical replicates, and shifted one base

upstream to the location of the hydrolyzed ribonucleotide as summarized (Supplementary Table 2). These counts were converted to bigWig format for visualization on the UCSC browser. The distributions of counts per-nucleotide were determined using these values.

End count scaling and background subtraction

Two modes of end count scaling were used, depending on several factors. For analyses resulting in visual comparisons of individual libraries (i.e. heat maps and meta-analyses), end counts were normalized to counts per million uniquely mapped reads (divided by the values listed in Supplementary Table 2 under “Uniquely mapped ends” and then multiplied by 1,000,000). For analyses that required weighted averaging of multiple libraries and background subtraction (i.e. strand bias maps, origin predictions, and genomic ribonucleotide density estimates), end counts were scaled using n chromosomal 5'-end counts as internal standards (counts were divided by the values listed in Supplementary Table 2 under “Telomere End-Derived Scale Factor”; see below). To ensure that ends in these latter analyses originated only from replicase-inserted genomic ribonucleotides, scaled end counts from polymerase “L” variant strains (*rnh201*⁺, *rnh201*⁻, and *pol2-M644L rnh201*⁻) were subtracted from the scaled end counts of corresponding promiscuous-replicase strains (*pol1-Y869A rnh201*⁻, *pol3-L612G rnh201*⁻, and *pol2-M644G rnh201*⁻, respectively).

Calculating telomere end-derived scale factors and genomic ribonucleotide densities

The genomic ribonucleotide density (R_{bulk}) is

$$R_{bulk} = \frac{N_{bulk}}{2 \times L_{genome} - L_{telomere}},$$

where N_{bulk} is the bulk end count, L_{genome} is the length of the genome, and $L_{telomere}$ is the total length of all telomeric repeats in the reference genome:

$$L_{telomere} = \sum_{i=1}^{2 \times N_C} \max(0, L_{i,telomere} - L_{read}),$$

where N_C is the chromosome count (16 in *S. cerevisiae*). The mean number of 5' chromosome ends per telomere ($N_{telomere}$) is similar, but should be corrected for R_{bulk} in order to account for ribonucleotides found in telomeric repeats but not at chromosome ends (this correction never amounted to more than 2% of the final value):

$$N_{telomere} = \frac{N_{telomere} - L_{telomere} \times R_{bulk}}{2 \times N_C},$$

where $N_{telomere}$ is the unadjusted total telomeric end count. $N_{telomere}$ serves as a scaling factor, allowing conversion of end counts in any bin into counts per position per genome (Supplementary Table 2). Where the bin is the whole genome, this results in an estimate of

the mean fragment size ($L_{fragment}$; always larger than the median fragment size as reported in ^{21,22}:

$$L_{fragment} = \frac{N_{telomere}}{R_{bulk}},$$

and thence the mean number of ends per genome (N_{ends} ; always smaller than the median count per genome as reported in ^{21,22}:

$$N_{ends} = \frac{2 \times L_{genome}}{L_{fragment}}.$$

Predicting replication origins from HydEn-Seq maps

Replication origins were predicted from the weighted average fraction of scaled and background-subtracted ends (see above) mapping to the top strand in the Pol α , δ or ϵ variant strains. In each bin, the weighted average top-strand-fraction (f) was

$$\bar{f} = \begin{cases} \frac{\left(\begin{matrix} 0, & \alpha_f + \alpha_r = 0 \\ 1 - \frac{\alpha_f}{\alpha_f + \alpha_r}, & \text{otherwise} \end{matrix} \right) + \left(\begin{matrix} 0, & \delta_f + \delta_r = 0 \\ 1 - \frac{\delta_f}{\delta_f + \delta_r}, & \text{otherwise} \end{matrix} \right) + \left(\begin{matrix} 0, & \epsilon_f + \epsilon_r = 0 \\ \frac{2\epsilon_f}{\epsilon_f + \epsilon_r}, & \text{otherwise} \end{matrix} \right)}{\left(\begin{matrix} 0, & \alpha_f + \alpha_r = 0 \\ 1, & \text{otherwise} \end{matrix} \right) + \left(\begin{matrix} 0, & \delta_f + \delta_r = 0 \\ 1, & \text{otherwise} \end{matrix} \right) + \left(\begin{matrix} 0, & \epsilon_f + \epsilon_r = 0 \\ 2, & \text{otherwise} \end{matrix} \right)}, & \text{Null, } \alpha_f + \alpha_r + \delta_f + \delta_r + \epsilon_f + \epsilon_r = 0 \\ \text{otherwise} & \text{otherwise} \end{cases}$$

where α , δ , and ϵ are the background-subtracted end counts from *pol1-Y869A rnh201*, *pol3-L612G rnh201*, and *pol2-M644G rnh201* strains, respectively (each of which is itself the average of scaled counts from all replicate libraries). Parameters for origin calling were set based on results from a training set of OriDB confirmed replication origins on chromosome 11. Predicted origins were defined as regions where the bias changed abruptly over a defined distance in the weighted average curve in Fig. 4 (black; 200 bp bins; smoothed over 9 bins). In order for a position to be called an origin, either the average slope (the derivative) of the black curve had to exceed 0.00011 fractional units per bp in an 11-bin window (2.2 kbp) or 0.00016 fractional units per base pair in at least 3 of five surrounding bins (600 bp out of 1 kbp). These parameters attempt to define a sufficiently abrupt bias change over a region wide enough to exclude random noise.

Meta-analyses and preparation of heatmaps

Total counts of the per-strain 5' ends intersecting same- and opposite-strand bins centered on genomic features of interest were determined using custom tools, excluding all mitochondrial annotations. Heatmaps, generated using the Partek Genomics Suite depict counts in all bins (normalized to ends per 1,000,000 uniquely mapped reads; see Supplementary Table 2), while meta-analyses depict the sum across all features.

Ribonucleotide frequencies

The composition of uniquely mapped ends was tallied in defined windows on each genomic strand. Windows were set in regions of high strand bias (99%) to ensure that nearly all ends represented ribonucleotides inserted by a particular replicase, depending on strand (leading versus lagging; e.g., grey or black bars, respectively in Fig. 3). The frequency of nucleotides occurring at 5' ends intersecting these windows was determined using custom tools.

In silico mapping of the mitochondrial genome

The mitochondrial genome sequence was rearranged, such that the first 42,888 base pairs were removed from the start, and appended to the end. Read pairs from sample WT.1 (Supplementary Table 2) were aligned to this reordered mitochondrial genome, using bowtie 0.12.8 (-m1 -v2 -X10000). Based on start and end coordinates of the paired-end alignments, per-nucleotide coverage of HydEn-Seq fragments was determined using genomeCoverageBed (-d).

Analysis of polymerase and RNase H2 conservation

PSI-BLAST searches⁵⁶ (parameters in Supplementary Table 6) were conducted to find sequences homologous to *S. cerevisiae* replicases (catalytic subunit sequences from the Saccharomyces Genome Database) and the predicted *RNH201* of *Schizosaccharomyces pombe* 972h- (gi accession number 19114596). Environmental sequences were excluded. For the replicases, PSI-BLAST was iterated until no new eukaryotic sequences were found. For *RNH201*, PSI-BLAST was iterated until >5,000 hits were acquired. The top hits were selected until the cumulative e-value exceeded 1.47. In all cases, partial sequences were culled and the remainder were aligned with CLUSTAL X (2.0) default parameters⁵⁷. Sequences with obvious deletions spanning active sites were culled, the remainder re-aligned, and trees built from the results (neighbor-joining, default parameters). The tree in Supplementary Figure 3 was constructed, in part, with the Interactive Tree of Life version 2.2.2 (<http://itol.embl.de>)⁵⁸.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Matthew Young and Matthew Longley for helpful comments on the manuscript. This work was supported by the Division of Intramural Research of the US National Institutes of Health (NIH), National Institute of Environmental Health Sciences (Project Z01 ES065070 to TAK), and by NIH grant 2R01GM052319-16A1 to PAM.

References

1. Johansson E, Dixon N. Replicative DNA polymerases. Cold Spring Harb Perspect Biol. 2013; 5
2. Georgescu RE, et al. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. Nat Struct Mol Biol. 2014; 21:664–70. [PubMed: 24997598]
3. Burgers PM. Polymerase dynamics at the eukaryotic DNA replication fork. J Biol Chem. 2009; 284:4041–5. [PubMed: 18835809]

4. Pursell ZF, Isoz I, Lundstrom EB, Johansson E, Kunkel TA. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science*. 2007; 317:127–30. [PubMed: 17615360]
5. Lujan SA, et al. Mismatch repair balances leading and lagging strand DNA replication fidelity. *PLoS Genet*. 2012; 8:e1003016. [PubMed: 23071460]
6. Lujan SA, et al. Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Research*. 2014; 24:1751–64. [PubMed: 25217194]
7. Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA. Division of labor at the eukaryotic replication fork. *Mol Cell*. 2008; 30:137–44. [PubMed: 18439893]
8. Larrea AA, et al. Genome-wide model for the normal eukaryotic DNA replication fork. *Proc Natl Acad Sci U S A*. 2010; 107:17674–9. [PubMed: 20876092]
9. Balakrishnan L, Bambara RA. Okazaki fragment metabolism. *Cold Spring Harb Perspect Biol*. 2013; 5
10. Smith DJ, Whitehouse I. Intrinsic coupling of lagging-strand synthesis to chromatin assembly. *Nature*. 2012; 483:434–8. [PubMed: 22419157]
11. McGuffee SR, Smith DJ, Whitehouse I. Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell*. 2013; 50:123–35. [PubMed: 23562327]
12. Yeeles JT, Poli J, Mariani KJ, Pasero P. Rescuing stalled or damaged replication forks. *Cold Spring Harb Perspect Biol*. 2013; 5:a012815. [PubMed: 23637285]
13. Ghosal G, Chen J. DNA damage tolerance: a double-edged sword guarding the genome. *Transl Cancer Res*. 2013; 2:107–129. [PubMed: 24058901]
14. Wanrooij S, Falkenberg M. The human mitochondrial replication fork in health and disease. *Biochim Biophys Acta*. 2010; 1797:1378–88. [PubMed: 20417176]
15. Gerhold JM, Aun A, Sedman T, Joers P, Sedman J. Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol Cell*. 2010; 39:851–61. [PubMed: 20864033]
16. Reyes A, et al. Mitochondrial DNA replication proceeds via a ‘bootlace’ mechanism involving the incorporation of processed transcripts. *Nucleic Acids Res*. 2013; 41:5837–50. [PubMed: 23595151]
17. Williams, JS., et al. RNA-DNA damage elicited by ribonucleotides incorporated during DNA replication is leading strand-specific. personal communication.
18. Nick McElhinny SA, et al. Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol*. 2010; 6:774–81. [PubMed: 20729855]
19. Sparks JL, et al. RNase H2-initiated ribonucleotide excision repair. *Mol Cell*. 2012; 47:980–6. [PubMed: 22864116]
20. Williams JS, Kunkel TA. Ribonucleotides in DNA: Origins, repair and consequences. *DNA Repair (Amst)*. 2014
21. Williams JS, et al. Topoisomerase I-mediated removal of ribonucleotides from nascent leading-strand DNA. *Mol Cell*. 2013; 49:1010–5. [PubMed: 23375499]
22. Lujan SA, Williams JS, Clausen AR, Clark AB, Kunkel TA. Ribonucleotides are signals for mismatch repair of leading-strand replication errors. *Mol Cell*. 2013; 50:437–43. [PubMed: 23603118]
23. Pavlov YI, Shcherbakova PV, Kunkel TA. In vivo consequences of putative active site mutations in yeast DNA polymerases alpha, epsilon, delta, and zeta. *Genetics*. 2001; 159:47–64. [PubMed: 11560886]
24. Nick McElhinny SA, et al. Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc Natl Acad Sci U S A*. 2010; 107:4949–4954. [PubMed: 20194773]
25. Siow CC, Nieduszynska SR, Muller CA, Nieduszynski CA. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res*. 2012; 40:D682–6. [PubMed: 22121216]
26. Niimi A, et al. Palm mutants in DNA polymerases alpha and eta alter DNA replication fidelity and translesion activity. *Mol Cell Biol*. 2004; 24:2734–46. [PubMed: 15024063]
27. Nick McElhinny SA, Stith CM, Burgers PM, Kunkel TA. Inefficient proofreading and biased error rates during inaccurate DNA synthesis by a mutant derivative of *Saccharomyces cerevisiae* DNA polymerase delta. *J Biol Chem*. 2007; 282:2324–32. [PubMed: 17121822]

28. Nick McElhinny SA, Kissling GE, Kunkel TA. Differential correction of lagging-strand replication errors made by DNA polymerases {alpha} and {delta}. *Proc Natl Acad Sci U S A*. 2010; 107:21070–5. [PubMed: 21041657]
29. Bielinsky AK, Gerbi SA. Chromosomal ARS1 has a single leading strand start site. *Mol Cell*. 1999; 3:477–86. [PubMed: 10230400]
30. Clark AB, Lujan SA, Kissling GE, Kunkel TA. Mismatch repair-independent tandem repeat sequence instability resulting from ribonucleotide incorporation by DNA polymerase epsilon. *DNA Repair (Amst)*. 2011; 10:476–82. [PubMed: 21414850]
31. Ghodgaonkar MM, et al. Ribonucleotides misincorporated into DNA act as strand-discrimination signals in eukaryotic mismatch repair. *Mol Cell*. 2013; 50:323–32. [PubMed: 23603115]
32. Clausen AR, Zhang S, Burgers PM, Lee MY, Kunkel TA. Ribonucleotide incorporation, proofreading and bypass by human DNA polymerase delta. *DNA Repair (Amst)*. 2013; 12:121–7. [PubMed: 23245697]
33. Williams JS, et al. Proofreading of ribonucleotides inserted into DNA by yeast DNA polymerase varepsilon. *DNA Repair (Amst)*. 2012; 11:649–56. [PubMed: 22682724]
34. Kim N, et al. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science*. 2011; 332:1561–4. [PubMed: 21700875]
35. Potenski CJ, Niu H, Sung P, Klein HL. Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. *Nature*. 2014; 511:251–4. [PubMed: 24896181]
36. Allen-Soltero S, Martinez SL, Putnam CD, Kolodner RD. A *Saccharomyces cerevisiae* RNase H2 interaction network functions to suppress genome instability. *Mol Cell Biol*. 2014; 34:1521–34. [PubMed: 24550002]
37. Reijns MA, et al. Enzymatic removal of ribonucleotides from DNA is essential for Mammalian genome integrity and development. *Cell*. 2012; 149:1008–22. [PubMed: 22579044]
38. Hiller B, et al. Mammalian RNase H2 removes ribonucleotides from DNA to maintain genome integrity. *J Exp Med*. 2012; 209:1419–26. [PubMed: 22802351]
39. Watt DL, Johansson E, Burgers PM, Kunkel TA. Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. *DNA Repair (Amst)*. 2011; 10:897–902. [PubMed: 21703943]
40. Clausen AR, Murray MS, Passer AR, Pedersen LC, Kunkel TA. Structure-function analysis of ribonucleotide bypass by B family DNA replicases. *Proc Natl Acad Sci U S A*. 2013; 110:16802–7. [PubMed: 24082122]
41. Lazzaro F, et al. RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol Cell*. 2012; 45:99–110. [PubMed: 22244334]
42. Dalgaard JZ. Causes and consequences of ribonucleotide incorporation into nuclear DNA. *Trends Genet*. 2012; 28:592–7. [PubMed: 22951139]
43. Anand RP, Lovett ST, Haber JE. Break-induced DNA replication. *Cold Spring Harb Perspect Biol*. 2013; 5:a010397. [PubMed: 23881940]
44. Foury F, Roganti T, Lecrenier N, Purnelle B. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett*. 1998; 440:325–31. [PubMed: 9872396]
45. Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. *FEBS J*. 2009; 276:1494–505. [PubMed: 19228196]
46. Grossman LI, Watson R, Vinograd J. The presence of ribonucleotides in mature closed-circular mitochondrial DNA. *Proc Natl Acad Sci U S A*. 1973; 70:3339–43. [PubMed: 4202844]
47. Yang MY, et al. Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell*. 2002; 111:495–505. [PubMed: 12437923]
48. Kasiviswanathan R, Copeland WC. Ribonucleotide discrimination and reverse transcription by the human mitochondrial DNA polymerase. *J Biol Chem*. 2011; 286:31490–500. [PubMed: 21778232]
49. Shadel GS, Clayton DA. Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem*. 1997; 66:409–35. [PubMed: 9242913]
50. Bowmaker M, et al. Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. *J Biol Chem*. 2003; 278:50961–9. [PubMed: 14506235]

51. Fuste JM, et al. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol Cell*. 2010; 37:67–78. [PubMed: 20129056]
52. Holt IJ, Reyes A. Human mitochondrial DNA replication. *Cold Spring Harb Perspect Biol*. 2012; 4
53. Cerritelli SM, et al. Failure to produce mitochondrial DNA results in embryonic lethality in *Rnaseh1* null mice. *Mol Cell*. 2003; 11:807–15. [PubMed: 12667461]
54. Vengrova S, Dalgaard JZ. The wild-type *Schizosaccharomyces pombe* *mat1* imprint consists of two ribonucleotides. *EMBO Rep*. 2006; 7:59–65. [PubMed: 16299470]
55. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17:10–2.
56. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997; 25:3389–3402. [PubMed: 9254694]
57. Larkin MA, et al. Clustal W and clustal X version 2.0. *Bioinformatics*. 2007; 23:2947–2948. [PubMed: 17846036]
58. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*. 2011; 39:W475–W478. [PubMed: 21470960]

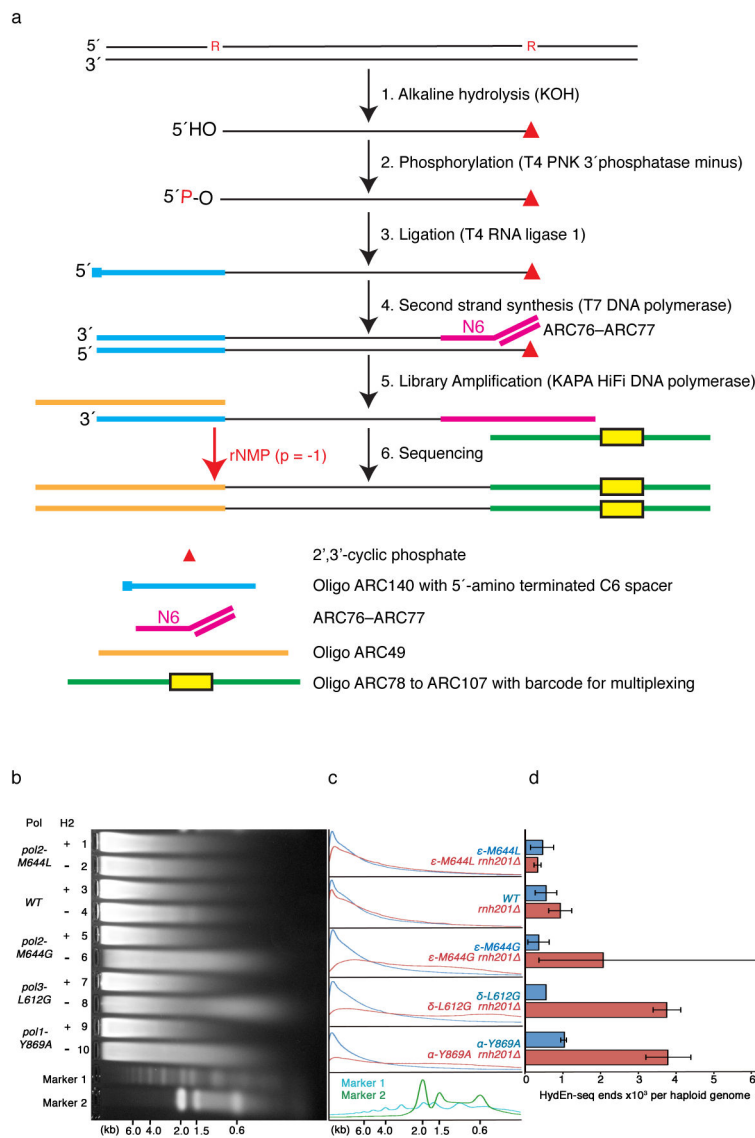


Figure 1. Mapping ribonucleotides by HydEn-Seq

(a) HydEn-Seq protocol. The procedure was performed as described in Methods, using the oligonucleotides listed in Supplementary Table 1. (b) Alkaline agarose gel electrophoresis. The analysis was performed as previously described¹⁸. Genomic DNA samples from the indicated yeast strains (*lanes 1–10*) were treated with alkali, separated by 1% alkaline agarose gel electrophoresis, and imaged after staining with SYBR Gold. Migration positions of two DNA size standards are indicated. (c) Densitometry scans of the gel image in (b). The Y-axis is scaled to maximum intensity for each pair of lanes. (d) Mean HydEn-seq end counts per haploid genome (see N_{ends} calculation in Methods; error bars represent ranges of two to four independent measurements).

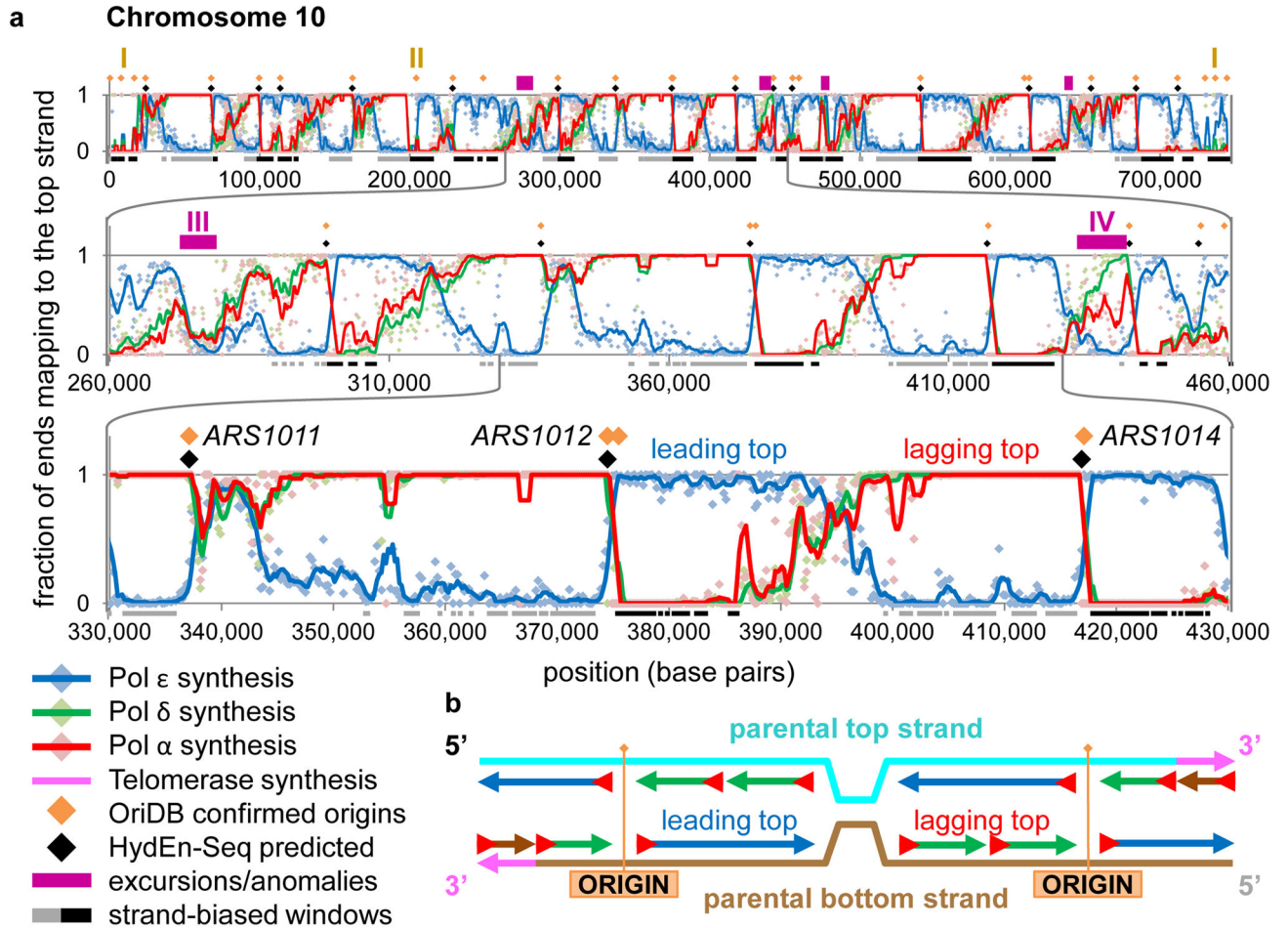


Figure 2. Strand-specific ribonucleotide mapping of chromosome 10

(a) Top, map of chromosome 10 showing the fraction of end reads mapped to the top strand in bins of 200 base pairs, after background subtraction (see Methods). Origin prediction was complicated at chromosome ends (I) and in other highly repetitive regions (II). Middle, an expanded 200 kbp region of chromosome 10. Excursions (in purple) from the simplest polymerase division of labor (Pol α or δ lagging, Pol ϵ leading) fall into two classes: unexpected Pol α , δ or ϵ correspondence (III) and Pol α or δ divergence (IV). Bottom, 100 kbp region of chromosome 10. Inter-origin regions are more (e.g. *ARS1012-ARS1014*) or less (e.g. *ARS1011-ARS1012*) symmetrical, depending on fork progression rates and origins firing times. Some origins in the origin database have little effect on ribonucleotide strand bias (e.g. *ARS1013*), indicating either an incorrect call, minority participation in normally growing cells, unidirectional origin firing, or simply later firing, such that forks proceeding from adjacent origins approach to within current detection thresholds. (b) A stylized chromosome with two replication origins, showing the division of polymerase labor, as predicted from the direction of strand bias transitions at origins (compare with the *ARS1012-ARS1014* region above). Roughly three quarters of previously confirmed replication origins (orange diamonds in panel (a); *S. cerevisiae* OriDB) align with abrupt transitions in strand preference (see Methods for quantitation). This allows algorithmic prediction of origins

(black diamonds). *ARS1013* was not detected via HydEn-seq. It is indicated (orange diamond) but not labeled.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

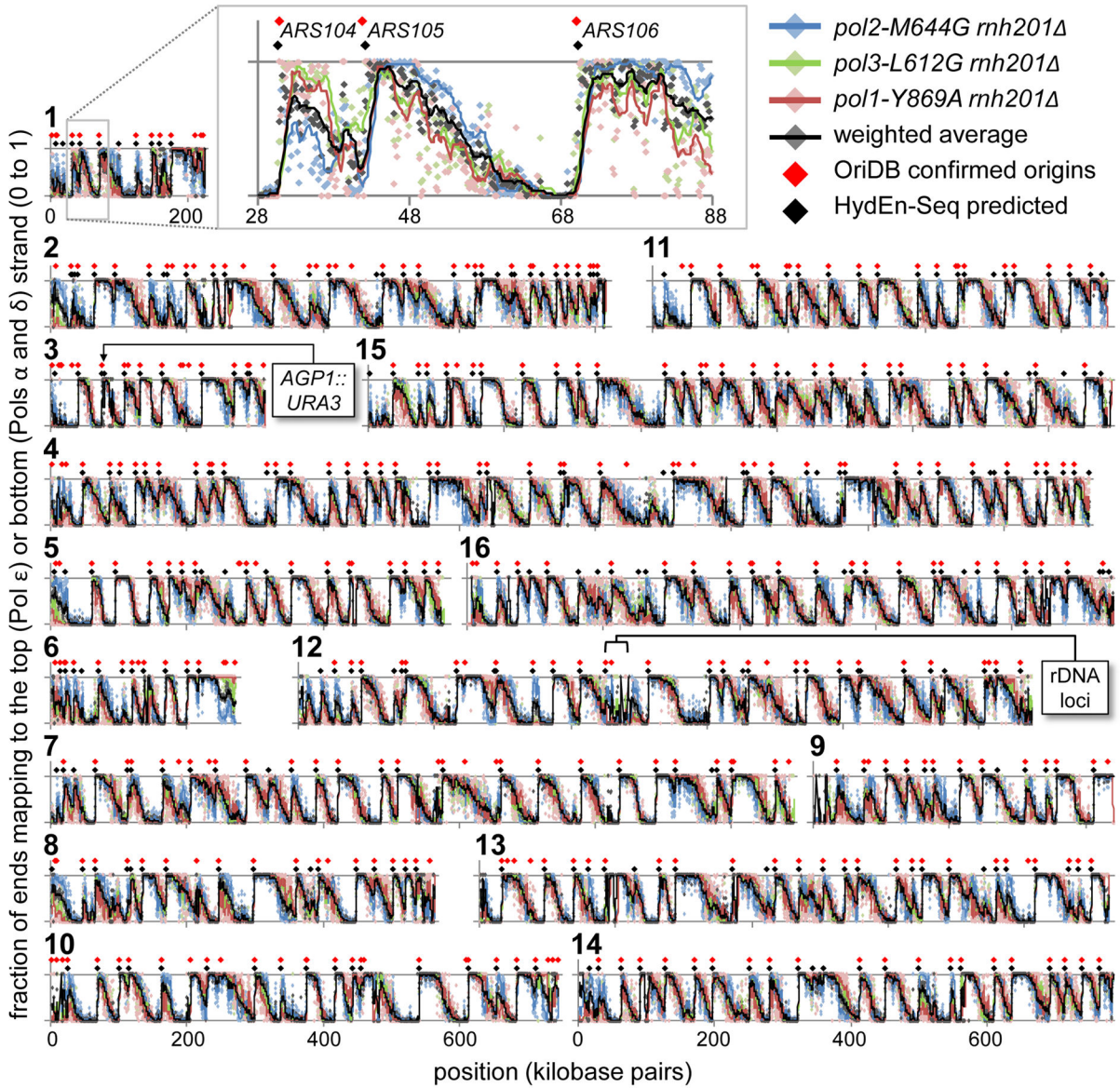


Figure 3. Genome-wide replication origins located by HydEn-Seq

(a) A 60 kbp segment of chromosome 1 showing the fraction of ends, after background subtraction, that mapped to the top strand from Pol ϵ data (*pol2-M644G rh201 Δ* ; blue) and the fraction mapped to the bottom strand for Pol α and δ data (*pol1-Y869A* and *pol3-L612G*, in red and green, respectively). Grey points are the weighted average of the other three data sets in each bin (see Methods). All curves are trend lines smoothed over 10 bins. (b) As per (a), but for all 16 *S. cerevisiae* chromosomes. Shown for reference are the locations of the *URA3* mutational reporter gene (near *ARS306*; used in our previous studies of leading and lagging strand replication fidelity⁵, and the rDNA locus in chromosome 12 (not drawn to scale; the highly repetitive sequence precludes read mapping).

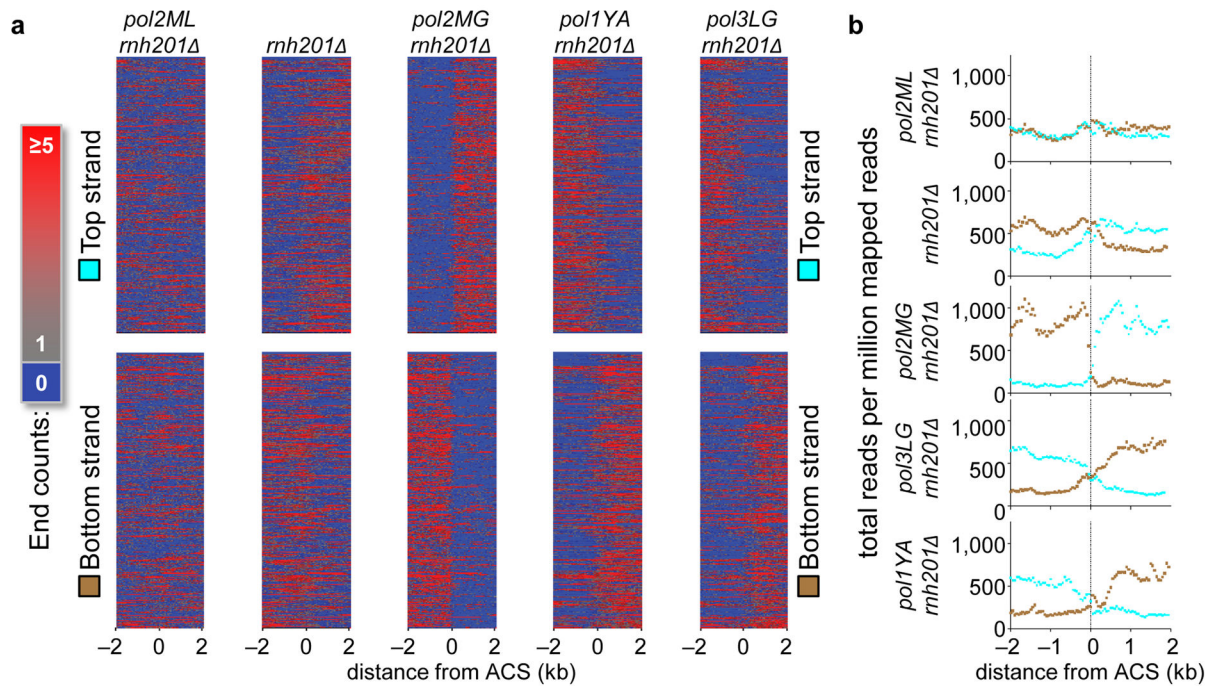


Figure 4. Distribution of ribonucleotides near origins in RER-deficient strains

(a) Heat maps for the top and bottom strands of the nuclear genome in five different *rnh201* strains, scaled per million reads and centered across a 4 kbp window of the 394 replication origins reported in the yeast origin database²⁵. **(b)** Meta-analysis of strand-specific ribonucleotides at 214 replication origins analyzed in a previous study¹⁰, again scaled per million reads, in bins of 50 bp.

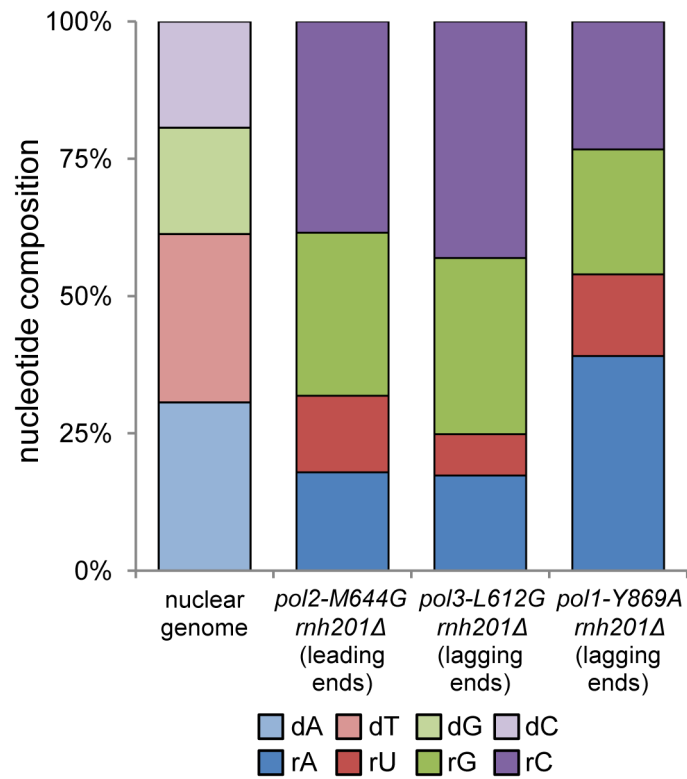


Figure 5. Ribonucleotide base identity

The proportion of each ribonucleotide base present in the nuclear genome of the three *rh201* strains encoding the indicated variant replicases. The base composition of the genome is shown on the left. Ribonucleotide proportions were calculated from the most highly strand-biased 10% of the genome (i.e. windows near replication origins; examples in Fig. 2a).

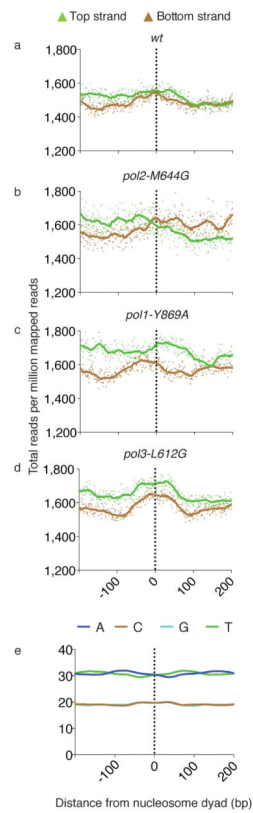


Figure 6. Meta-analysis of ribonucleotides at the nucleosome dyad

(a) Meta-analysis of strand-specific ribonucleotide mapping at 37,888 nucleosome dyads⁶ for the *rnh201* strain, scaled per million reads and centered within a 400 bp window. Each dot indicates the number of 5'-DNA ends reads at one base pair. The vertical dotted line indicates the dyad. (b) As in (a) but for the *pol2-M644G rnh201* strain. (c) As in (a) but for the *pol1-Y869A rnh201* strain. (d) As in (a) but for the *pol3-L612G rnh201* strain. The solid lines are the smoothed averages for a sliding window. (e) The base composition surrounding the nucleosome dyad.

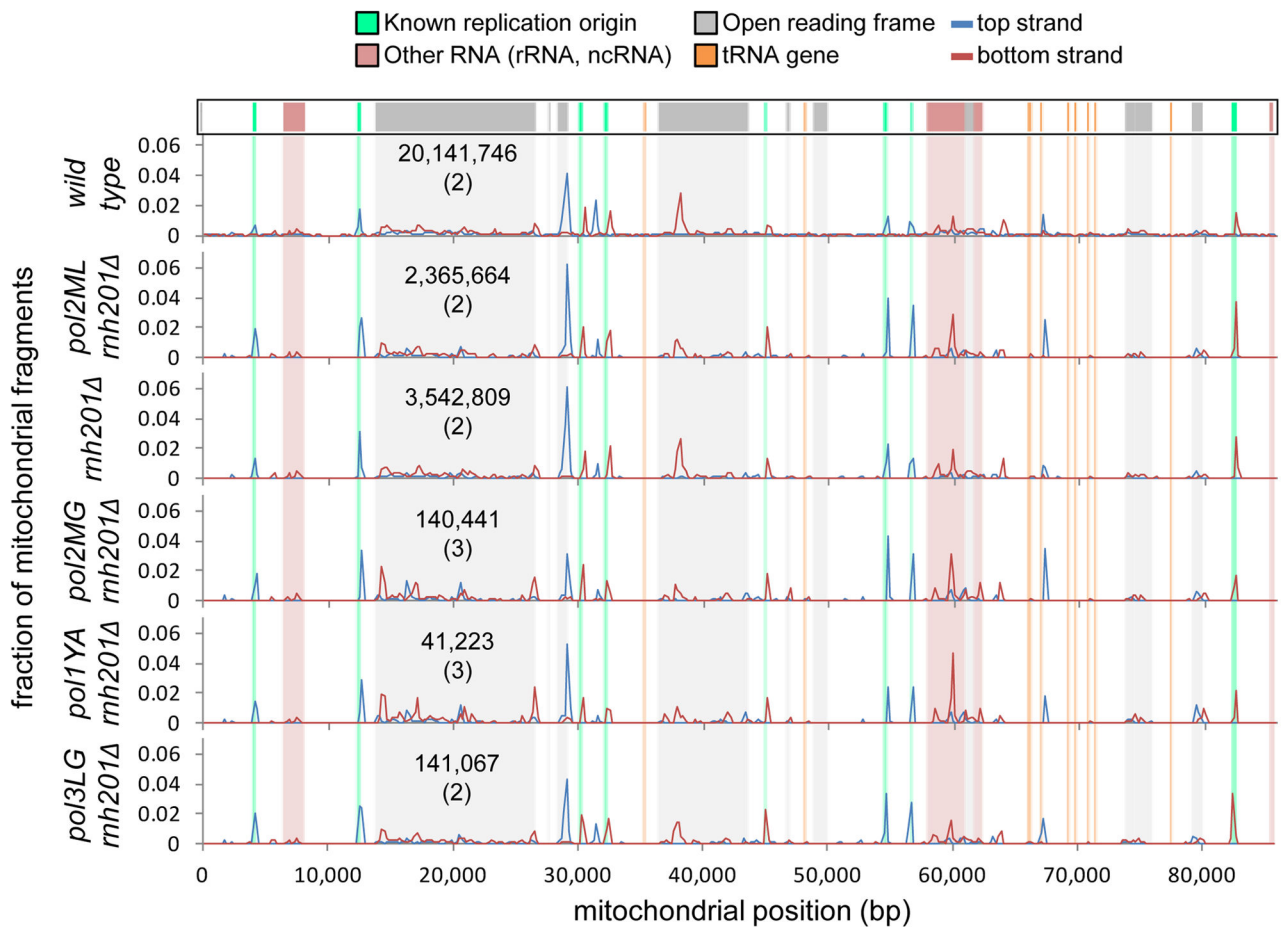


Figure 7. HydEn-Seq maps of mitochondrial DNA

Mitochondrial genomes for six strains are shown to indicate the base pair (bp) locations and proportions of strand-specific 5'-DNA ends detected by HydEn-Seq (blue for plus strand, red for minus). Previously assigned replication origins⁴⁴ are shaded in green, coding sequences in grey, tRNA genes in orange and genes for other non-coding RNAs in pink. Total mitochondrial end counts are shown for each strain with the number of replicate HydEn-Seq libraries for each in parentheses.