Genome duplication in *Leishmania major* relies on persistent subtelomeric DNA replication

Jeziel Dener Damasceno[1]*, Catarina A Marques[1], Dario Beraldi[1], Kathryn Crouch[1], Craig Lapsley[1], Ricardo Obonaga[2], Luiz RO Tosi[2], Richard McCulloch[1]*

[1]The Wellcome Centre for Integrative Parasitology, University of Glasgow, Institute of Infection, Immunity and Inflammation, Glasgow, United Kingdom; [2]Department of Cell and Molecular Biology, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil

Abstract DNA replication is needed to duplicate a cell's genome in S phase and segregate it during cell division. Previous work in *Leishmania* detected DNA replication initiation at just a single region in each chromosome, an organisation predicted to be insufficient for complete genome duplication within S phase. Here, we show that acetylated histone H3 (AcH3), base J and a kinetochore factor co-localise in each chromosome at only a single locus, which corresponds with previously mapped DNA replication initiation regions and is demarcated by localised G/T skew and G4 patterns. In addition, we describe previously undetected subtelomeric DNA replication in G2/M and G1-phase-enriched cells. Finally, we show that subtelomeric DNA replication, unlike chromosome-internal DNA replication, is sensitive to hydroxyurea and dependent on 9-1-1 activity. These findings indicate that *Leishmania*'s genome duplication programme employs subtelomeric DNA replication initiation, possibly extending beyond S phase, to support predominantly chromosome-internal DNA replication initiation within S phase.

*For correspondence:
jeziel.damasceno@glasgow.ac.uk (JDD);
Richard.McCulloch@glasgow.ac.uk (RMC)

Introduction

Once every cell cycle, a cell must completely duplicate its genome before segregating the two resulting copies into offspring cells. Genome duplication relies on DNA replication, a normally tightly controlled and high-fidelity reaction (*Burgers and Kunkel, 2017*). DNA replication is initiated at genomic loci termed origins, which are sequence-conserved features in prokaryotes (*Leonard and Méchali, 2013*). In contrast, with the exception of *Saccharomyces* and closely related yeasts (*Dhar et al., 2012*), replication origins in eukaryotes are not defined by conserved sequences. Instead, more elusive features, such as chromatin accessibility, transcription level and epigenetic elements (*MacAlpine et al., 2010*; *Cayrou et al., 2015*; *Deal et al., 2010*; *Dellino et al., 2013*; *Lombraña et al., 2013*; *Mesner et al., 2011*; *Chen et al., 2019*), are determinants of replication initiation activity. What is common to all known eukaryotic origins is that they are licensed through binding by the origin recognition complex (ORC), which recruits the replicative helicase, MCM2-7, during G1 (*Bleichert et al., 2017*). At the onset of S-phase origins are fired, initiating DNA synthesis that proceeds bi-directionally along the chromosomes.

Likely as a result of increasing genome size, DNA replication in eukaryotes is initiated at multiple origins per linear chromosome, with the number of origins being proportional to chromosome size (*Al Mamun et al., 2016*). To preserve genomic stability, origins licensed in G1 outnumber those that are fired in early S phase. Thus, in the event of failure of complete DNA synthesis from the fired origins, others can be activated to ensure complete genome duplication by completion of S phase

(*McIntosh and Blow, 2012*; *Alver et al., 2014*). However, unlike all previously characterised eukaryotes, mapping of DNA replication using Marker Frequency Analysis coupled with deep sequencing (MFA-seq) detected only a single clear region of replication initiation in each chromosome of Leishmania, a grouping of single-celled parasites (*Marques et al., 2015*). If these MFA-seq regions represent origins (as they do in *S. cerevisiae*) (*Müller and Nieduszynski, 2012*), these data would suggest a DNA replication programme in *Leishmania* that is unprecedented in eukaryotes and, indeed, contrasts with the multiple origins mapped in the chromosomes of *Trypanosoma brucei* (*Tiengwe et al., 2012a*), a kinetoplastid relative of *Leishmania* (see below). Moreover, this DNA replication programme is predicted to be insufficient to allow complete duplication of the larger *Leishmania* chromosomes during S phase (*Marques et al., 2015*), accounting for perhaps 50% of the chromosome complement, and is thus inadequate to secure complete genome duplication during prior to cell division. A further complication in the emerging understanding of *Leishmania* DNA replication is that a later study, which mapped short nascent DNA strands (SNS-seq) in asynchronous cells detected thousands of DNA synthesis initiation sites (hundreds per chromosome), therefore revealing a huge dichotomy with MFA-seq mapping (*Lombraña et al., 2016*). Indeed, 'DNA combing' analyses could detect DNA molecules with more than a single site of DNA synthesis (*Lombraña et al., 2016*; *Stanojcic et al., 2016*), though location within a chromosome could not be inferred and it could not be ruled out that extrachromosomal episomes, which arise frequently in *Leishmania* (*Ubeda et al., 2014*), were responsible for the DNA synthesis signals. These conflicting data raise questions, which we have sought to answer here, about the programme of DNA replication that *Leishmania* uses in order to effectively execute genome duplication.

Leishmania species are the causative agents of a spectrum of diseases, including skin-damage and fatal organ-failure leishmaniasis, affecting both humans and other animals worldwide (*Torres-Guerrero et al., 2017*). Leishmania belongs to the diverse kinetoplastid grouping (*Adl et al., 2019*; *Keeling and Burki, 2019*), which is evolutionarily distant from yeast, animals and plants, from where much of our understanding of eukaryotic DNA replication has emerged. Besides being of medical relevance (since several species and genera are parasites of insects and animals), kinetoplastids display unorthodoxies in core molecular processes. Notably, virtually all RNA polymerase (pol) II transcribed genes in Leishmania (*Martínez-Calvillo et al., 2004*) and *T. brucei* (*Kolev et al., 2010*) have been shown to be transcribed as multigene transcription units, each of which possesses a single poorly defined transcription start site that appears to be constitutively active (*Wedel et al., 2017*). The ubiquitous use of multigenic transcription is likely to be a universal feature of kinetoplastids (*Jackson et al., 2008*) and means that gene expression is perhaps exclusively regulated at the post-transcriptional level (*El-Sayed et al., 2005*; *Martínez-Calvillo et al., 2003*; *LeBowitz et al., 1993*; *Clayton, 2019*). Two further consequences arise from multigene transcription. First, one route for increased gene expression in the absence of transcriptional control is to increase gene copy number. In Leishmania this appears to be reflected in its remarkable genome plasticity (*Laffitte et al., 2016a*), manifested both as intra- and extra- chromosomal genome-wide gene copy-number variation (*Iantorno et al., 2017*; *Ubeda et al., 2008*) and as mosaic aneuploidy (*Rogers et al., 2011*; *Sterkers et al., 2011*; *Bussotti et al., 2018*; *Prieto Barja et al., 2017*). Second, in *T. brucei*, ORC and origin localisation is found at the boundaries of the multigene transcription units, where transcription initiates and/or terminates, perhaps limiting impediments to RNA pol movement (*Tiengwe et al., 2012a*). MFA-seq mapping also identified DNA replication initiation sites exclusively at transcription boundaries in Leishmania (*Marques et al., 2015*), whereas most SNS-seq mapped initiation sites are within the multigene transcription units, where they appear to coincide with trans-splicing and polyadenylation sites at which mature mRNAs are generated (*Lombraña et al., 2016*).

To attempt to clarify the current unclear picture of how DNA replication is executed in Leishmania, this study describes a modified MFA-seq strategy with which we demonstrate the existence of previously undetected sites of DNA replication initiation in the chromosomes of *L. major*. Consistent with previous MFA-seq mapping, DNA replication initiation in early S phase could only be clearly detected from a single internal region in each chromosome, which was marked by simultaneous accumulation of acetylated histone H3 (AcH3), a novel modified base, termed base J, and the putative kinetochore factor KKT1. In addition, we now describe DNA replication initiation that is proximal to the chromosome telomeres and only shows co-localisation with AcH3. DNA replication initiation from the subtelomeres is detected not only in early S phase cells, but also in cells enriched in late S, G2/M and G1 phases of the cell cycle. Furthermore, we show that telomere-proximal DNA

replication activity, unlike DNA replication from the core sites, is sensitive to replication stress and depends on RAD9 and HUS1, which are subunits of the 9-1-1 complex that acts in the replication stress response pathway (*Damasceno et al., 2013*; *Damasceno et al., 2016*; *Damasceno et al., 2018*). Thus, we reveal that two potentially distinct forms of DNA synthesis make up the genome replication programme of *L. major*, with a single predominant S-phase region of initiation in each chromosome supplemented by subtelomeric DNA synthesis activity that appears to extend beyond S phase.

Results

Detection of subtelomeric DNA replication throughout the cell cycle of *L. major*

Previous MFA-seq analysis, which detected a single region of replication initiation per chromosome in *L. major* and *L. mexicana* (*Marques et al., 2015*), is potentially limiting because it relies on cell sorting to enrich for S phase (replicating) and G2 or G1 (presumably non-replicating) cells, then inferring regions of DNA synthesis by comparing sequence read depth in the former relative to the latter. As a result, DNA replication very late in S phase, or that extends beyond the conventional S phase, would escape detection. To test this, we modified the MFA-seq approach in order that DNA content enrichment could be calculated in replicating cells relative to naturally occurring non-replicating cells. Mapping origins of replication by calculating DNA enrichment in exponentially growing cells versus cells in a stationary state has been used successfully in bacteria (*Ivanova et al., 2015*) and yeast (*Müller et al., 2014*). Thus, we reasoned that *L. major* in stationary phase could also serve as a non-replicative control for normalisation in MFA-seq analysis compared with cells that are growing exponentially. To test this, we used flow cytometry to compare DNA content of exponentially growing and stationary phase cells (*Figure 1A*) and, in addition, compared the capacity of the two cell populations to incorporate thymidine analogues (*Figure 1B,C*). Flow cytometry showed that the proportion of cells in S phase (with DNA content between 1C and 2C) was substantially lower in a population of stationary cells compared with exponentially growing cells (*Figure 1A*). Concomitantly, stationary phase cells, unlike exponentially growing cells, incorporated very low levels of IdU or EdU (*Figure 1B,C*). Thus, *L. major* cells in stationary phase perform little detectable DNA synthesis and were therefore deemed suitable to be used as a non-replicative sample in MFA-seq analysis.

Next, we compared MFA-seq profiles using read depth ratios from FACS-sorted early S (ES) and G2 cells, and from exponentially growing cells (EXP) and stationary (STA) cells (*Figure 1D*; *Figure 1—figure supplement 1* details the sequence data used and its processing for subsequent analyses). As reported previously (*Marques et al., 2015*), mapping the ratio of ES/G2 sequence reads revealed a single peak in each chromosome, with all peaks localising to the boundaries of the multigene transcription units. The same peaks were also apparent in EXP/STA read mapping but, strikingly, further regions of read enrichment were now detected that localised towards the chromosome ends (*Figure 1D*). Subtelomeric enrichment in EXP/STA could not be explained by loss of read coverage in these regions of the chromosomes in STA cells (*Figure 1—figure supplement 2*), and though the MFA-seq signal was more clearly detected when the data were analysed by Z score (*Figure 1—figure supplements 3*, *4* and *5*), the same pattern was discernible when analysing copy-number variation (*Figure 1—figure supplements 3* and *4*). To explore these new regions of predicted DNA replication initiation further, we compared read depth in exponentially growing cells sorted into G1, ES, late S (LS) and G2/M phases of the cell cycle (see gating strategy in *Figure 1A*) relative to STA cells. The pronounced ES/G2 MFA-seq peaks were only clearly seen in the ES/STA MFA-seq mapping (*Figure 1D*, *Figure 1—figure supplement 5*), whereas subtelomeric signal was seen in several chromosomes from the cell cycle-sorted DNA relative to STA. In ES/STA MFA-seq mapping, the amplitude of subtelomeric MFA-seq signal was lower than the more central MFA-seq peak in each chromosome and, indeed, was most clearly seen in the larger chromosomes. However, the subtelomeric MFA-seq signal was notably stronger in G1/STA and G2/STA MFA-seq mapping of all chromosomes, where the central MFA-seq peak was much less pronounced. Taken together, these data indicate two things. First, sites of subtelomeric DNA synthesis, predicted by change in sequence copy number, appear spatially separate from previously mapped, predominantly chromosome-central regions of DNA replication initiation (*Marques et al., 2015*) and can be found in virtually every
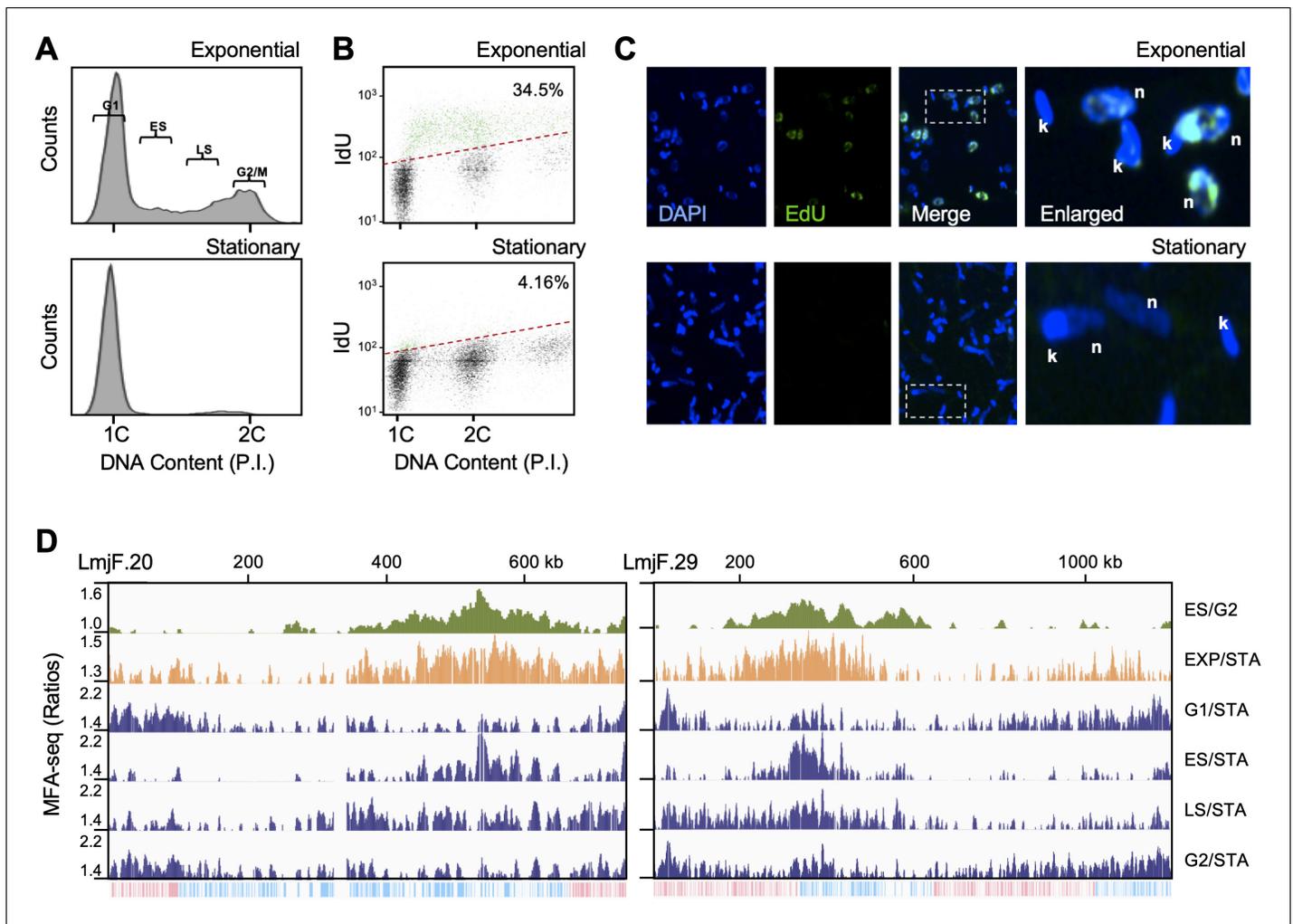
Figure 1. Detection of DNA replication initiation throughout the cell cycle in *L. major*. (A) DNA content analysis by FACS of cells in exponential growth or in stationary phase; DNA was stained with Propidium Iodide (PI). 1C and 2C indicate single and double DNA content, respectively; gates used to sort exponentially growing cells into G1, early S (ES), late S (LS) and G2/M enriched populations are indicated. (B) Representative density plots from flow cytometry analysis to detect DNA synthesis in exponentially growing and stationary phase cells; cells were incubated with IdU for 30 min and IdU fluorescence was detected under denaturing conditions; 30,000 cells were analysed; 1C and 2C indicate single and double DNA content, respectively; dashed red lines indicate the threshold used to discriminate negative (black dots) from IdU-positive (blue dots) events; inset numbers indicate total percentage of IdU-positive events relative to the whole population. (C) Cells in exponential growth or in stationary phase were pulsed with 10 µM EdU for 1 hr, subjected to click reaction and then visualised by confocal microscopy and DAPI staining; n and k indicate DNA from the nucleus and kinetoplast, respectively. (D) MFA-seq profile for the indicated chromosomes; from top to bottom: first track shows previously published MFA-seq profile in ES using G2/M cells for normalisation *Marques et al., 2015*; second track shows MFA-seq profile of exponentially growing cells, not sorted, using stationary (STA) cells for normalisation; the remaining tracks show MFA-seq profile for exponentially growing cells, sorted in the indicated cell cycle stages, using STA cells for normalisation; MFA-seq signal is shown as ratios of reads for all samples (y-axes; enriched populations indicated to the right), plotted in five kbp windows across the length of the chromosomes (x-axes; chromosome sizes are indicated); the bottom track in each panel indicates annotated CDSs (blue: transcribed from left to right; pink: transcribed from right to left).

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Distribution of reads per bin.

Figure supplement 2. Stationary phase sequencing coverage.

Figure supplement 3. Overview of data analysis.

Figure supplement 4. Comparing MFA-seq expressed as ratio and Z scores in the context of aneuploidy.

Figure supplement 5. MFA-seq profile for all chromosomes of *Leishmania major*.

L. major chromosome. Second, whereas a single replication initiation region in each *L. major* chromosome is activated early in S phase, predicted subtelomeric DNA replication initiation can be detected in all enriched stages of the *L. major* cell cycle.

MFA-seq reveals the programme of replication timing in the *L. major* genome

Analysis of the EXP/STA MFA-seq profile on individual chromosomes suggested the possibility of a programme of DNA replication initiation throughout the cell cycle (*Figure 1D*, *Figure 1—figure supplement 5*). In ES and LS, MFA-seq signal appeared to be found mainly in the internal regions of the chromosomes (completely overlapping previously mapped ES/G2 MFA-seq peaks), while signal was strongest proximal to the chromosome ends during G1 and G2/M. To test this possibility further, we first performed meta-analyses, comparing MFA-seq profiles across all chromosomes in EXP/STA cells (*Figure 2A*) and across different stages of the cell cycle (*Figure 2B*). In both EXP and ES cells, MFA-seq was seen as peaks of very consistent amplitude and width that fell into three clusters, one where the peak was found centrally in the chromosomes, and two others where the peak was displaced towards the right or left ends of the chromosomes. There was no obvious separation of peak localisation into the three clusters based on chromosome size (*Figure 2C*). In LS cells (*Figure 2B*), the MFA-seq peaks in each cluster had increased in width, again to a very consistent extent, with half of the peaks from the less central clusters having merged with the chromosome ends. These data are consistent with bidirectional progression of replication forks emanating from a single predominant region in each chromosome and suggest considerable consistency in rate or pattern of fork movement irrespective of initiation site. Finally, in G2/M and G1 cells, the three clusters overlapped to a much greater extent, with MFA-seq signal again showing a consistent pattern across the genome, but here the signal peaked at the chromosome ends and reduced in amplitude towards the interior of the molecules (*Figure 2B*). These data may be explained by DNA replication in *L. major* following a programme in which synthesis of a new chromosome initiates at a single predominant region in the interior of each chromosome in early S phase and progresses at a consistent rate towards the chromosome ends, but with DNA replication continuing even as cells navigate from LS through G2/M until they enter G1. Alternatively, the profiles may suggest a bimodal DNA replication programme, with all chromosomes initiating DNA replication from internal regions in ES phase as well as using DNA replication initiation that derives from the telomeres or from telomere-proximal regions, either late in S phase or in G2–G1 phases.

Since a single major MFA-seq peak was seen in each chromosome in ES cells, the majority of DNA replication is predicted to initiate from a single region in each case. Given the range of chromosome sizes in the *L. major* genome (~0.25–2.4 Mb), a prediction of the mapping is that the time to complete chromosome replication will be size-dependent. To test this prediction, we estimated the extent of each chromosome that is replicated in early S phase by comparing the ES/STA MFA-seq ratios with G2/STA ratios, revealing a greater extent of DNA replication in smaller chromosomes relative to the larger (*Figure 2D*). To explore this further, we separated the chromosomes into three size groupings and significant differences in replication timing were clearly seen (*Figure 2E*).

Chromatin and base composition differ at sites of core and subtelomere chromosome replication initiation

To date, our understanding of what DNA sequence or chromatin elements might dictate *Leishmania* DNA replication initiation is limited, and so we asked how the DNA replication pattern across the *L. major* cell cycle correlates with known chromatin and DNA features (*Figure 3*). Specifically, we examined the distribution of MFA-seq signal relative to sites of enrichment for acetylated histone H3 (AcH3; associated with transcription initiation) (*Thomas et al., 2009*), ß-D-glucosyl-hydroxymethyluracil (base J; associated with transcription termination) (*van Luenen et al., 2012*; *Reynolds et al., 2014*) and kinetoplastid kinetochore protein 1 (KKT1, which is most strongly enriched at a single site in each chromosome of Leishmania) (*Garcia-Silva et al., 2017*). This analysis revealed a difference between the chromosome-internal and subtelomere DNA replication reactions. As exemplified by chromosome 35 (*Figure 3A*), simultaneous co-localisation of AcH3, base J and KKT1 was found at the single MFA-seq peak seen in ES cells in every chromosome. Surprisingly, this co-localisation included MFA-seq peaks at six sites where the surrounding multigene transcription units converge,
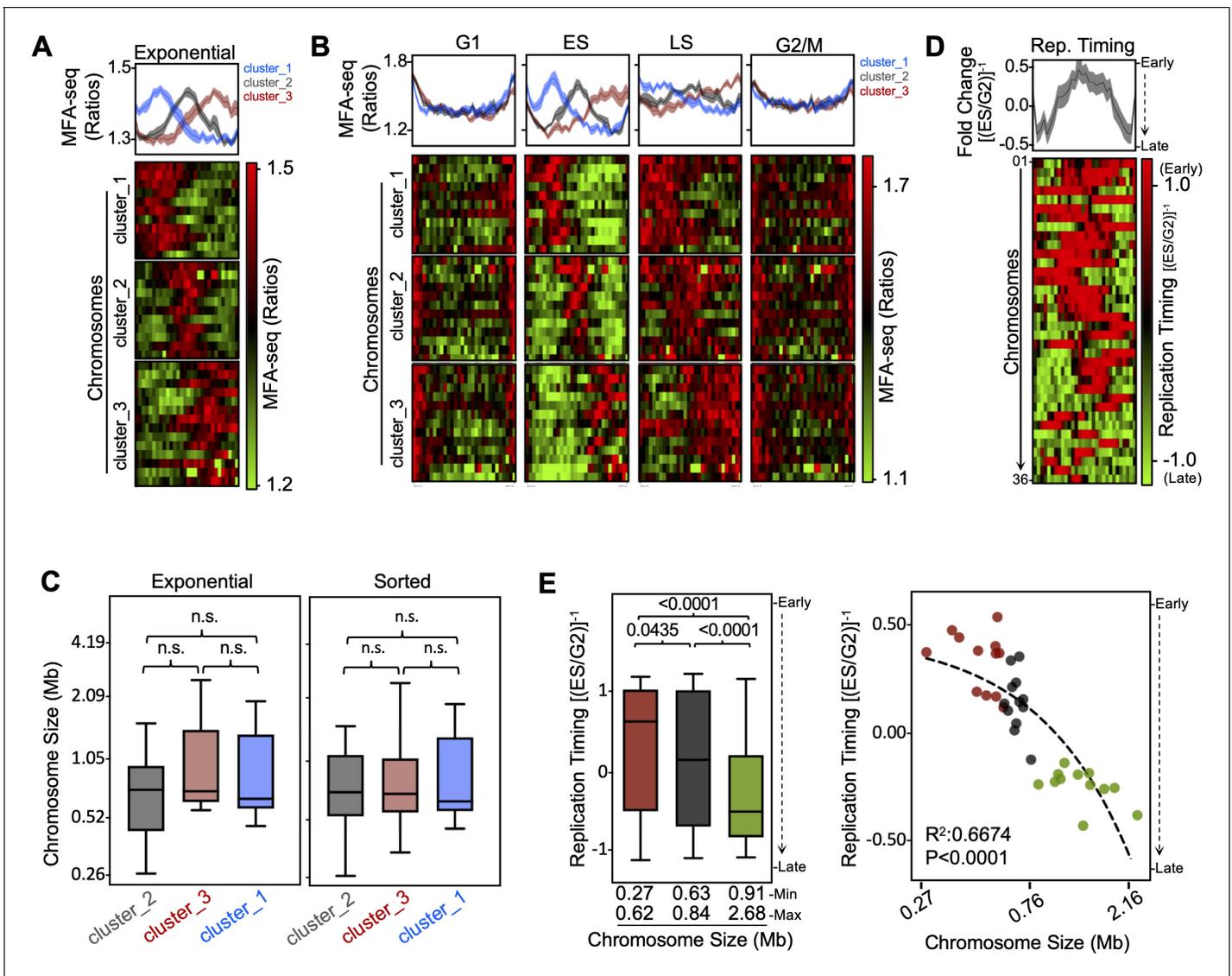
Figure 2. Changes in MFA-seq signal profile throughout the cell cycle. MFA-seq profile in (A) exponentially growing or (B) sorted cells in the indicated phases of cell cycle, in both cases relative to stationary phase (STA) cells; panels at the top represent global MFA-seq signal for the whole genome; colourmaps indicate MFA-seq profile for individual chromosomes (each row). Chromosomes were grouped by k-means clustering, using deepTools. Chromosomes in each cluster from exponential cells: *cluster_1*–01, 03, 05, 08, 19, 20, 22, 23, 27, 28, 32; *cluster_2*–02, 09, 10, 12, 15, 16, 17, 18, 24, 25, 26, 31, 33, 36; *cluster_3*–04, 06, 07, 11, 13, 14, 21, 29, 30, 34, 35; chromosomes in each cluster from sorted cells: *cluster_1*–01, 02, 03, 05, 08, 18, 19, 20, 22, 23, 25, 27,28,32; *cluster_2*–09, 10, 12, 15, 16, 17, 24, 26, 31, 33, 36; *cluster_3*–04, 06, 07, 11, 13, 14, 29, 21, 30, 34, 35. (C) Replication timing was estimated by calculating the fold change between MFA-seq signal (ratio) from ES cells and MFA-seq signal (ratio) from G2 cells relative to STA; panels at the top represent global MFA-seq signal for the whole genome; colourmaps indicate MFA-seq profile for individual chromosomes (each row). (D) Chromosome size is plotted for each cluster shown in A and B; differences were analysed with a Kruskal-Wallis test; n.s., not significant. (E) Mean replication timing (as calculated in C) from each chromosome is plotted; left, chromosomes were grouped according to their sizes (small, medium and large) and *P* values, analysed with a Kruskal–Wallis test, are shown for each indicated pair; right, mean values were plotted for each individual chromosome, with R^2 and *P* values for linear regression shown.

and at the telomere of chromosome 1; in each case these are locations of transcription termination, where AcH3 enrichment is not predicted (*Thomas et al., 2009*), but modest AcH3 enrichment was seen in all cases (*Figure 3B*, *Figure 1—figure supplement 5*). Meta-analysis of all chromosomes, testing correlation of MFA-seq signal in different cell cycle stages around loci where the AcH3, base J and KKT1 ChIP datasets overlap, revealed peaks of considerable consistency in amplitude and width in EXP and ES cells, which diminished as cells progressed to LS (*Figure 3C*). These data
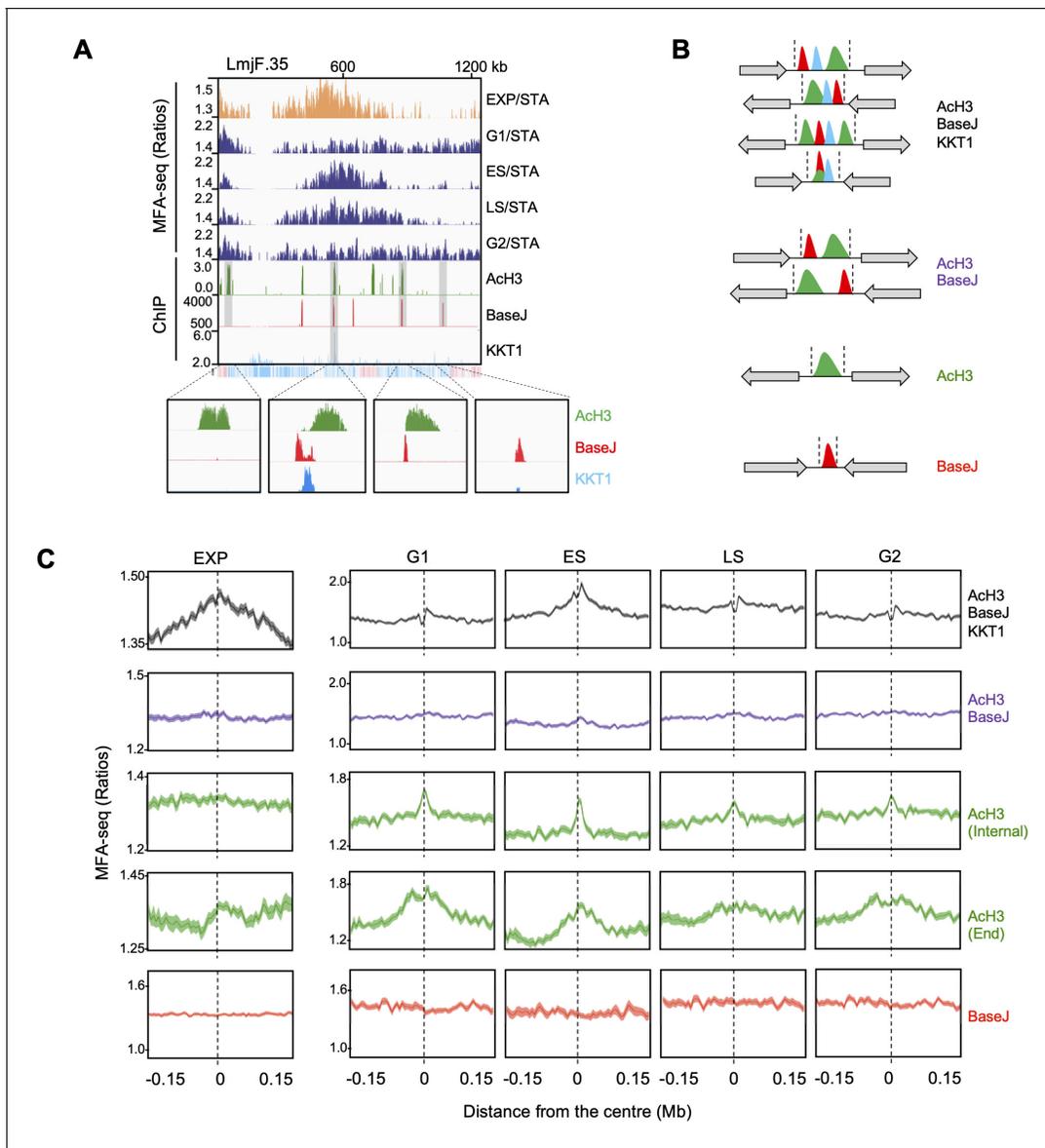
Figure 3. Changes in MFA-seq signal profile throughout the cell cycle are associated with distinct chromatin features. (A) A representative region of chromosome 35, showing MFA-seq signals in the indicated phases of cell cycle, as compared to the positioning of AcH3, BaseJ and KKT1 enriched sites; the bottom track indicates annotated CDSs (blue: transcribed from left to right; pink: transcribed from right to left); for improved visualisation, segments of interest are shaded and magnified at the bottom. (B) Schematic illustration of regions of transcription initiation or termination, containing the indicated combination of chromatin markers; gray arrows indicate all possible configurations of surrounding multigene transcription direction; coordinates defining the boundaries of these regions were taken from *Lombraña et al., 2016* and manually curated. (C) Metaplots of global MFA-seq signal, in exponential cells or at the indicated phases of the cell cycle, relative to STA cells ± 0.15 Mb from the centre of regions containing the indicated combination of chromatin marks; light-coloured areas around the lines indicate the standard error of the mean (SEM).

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Changes in MFA-seq signal profile throughout the cell.
Figure supplement 2. MFA-seq signal near chromosome's ends.
Figure supplement 3. MFA-seq of exponentially growing cells.
Figure supplement 4. DNA sequence within the regions containing the indicated combination of the chromatin markers was subject to MEME analysis; top three motifs (most abundant and lowest *p-value*) found within each region are shown.

suggest that the simultaneous presence of these three genome factors is a local driver for coordinated DNA replication initiation at a single region in each chromosome in ES but does not promote DNA replication initiation in other cell cycle stages.

Meta-analysis of MFA-seq at all sites where base J was found, alone or in combination with AcH3, showed that the modified base itself has little correlation with DNA replication in any cell cycle stage (*Figure 3C*; *Figure 3—figure supplement 1*). In contrast, MFA-seq signal was seen around sites of AcH3 enrichment in all cell cycle stages analysed (*Figure 3C*), suggesting that DNA replication in this chromatin context is ubiquitous throughout the cell cycle. Because chromatin environment and DNA sequence near telomeres might differ from the internal regions of chromosomes, we separately analysed MFA-seq signal around AcH3 sites near chromosome ends (<10 kb from telomeres) and those located elsewhere in the genome (>10 kb from telomeres). In both cases, MFA-seq signal was seen around sites of AcH3 enrichment in all phases of the cell cycle (*Figure 3C*; *Figure 3—figure supplement 1*). However, MFA-seq peaks around non-telomeric AcH3 regions were notably smaller in amplitude and narrower in width compared with the pronounced MFA-seq signal around telomere-proximal AcH3 sites. These data provide confirmation of a genome-wide association between AcH3 sites and DNA replication initiation throughout the cell cycle, but it appears that while AcH3 sites are associated with DNA replication initiation in the chromosome subtelomeres, fewer AcH3 sites are used in the chromosome interiors. Therefore, the genomic context where this histone modification is found correlates with distinct local replication initiation activity, perhaps because of unrecognised features at these sites in the subtelomeres. Alternatively, stronger subtelomeric MFA-seq signal may indicate DNA replication initiation emanating from the telomeres, and not necessarily from AcH3 sites.

To further explore the extent of DNA replication initiation near end-proximal AcH3 sites, we first mapped sequence ratios around the subtelomeres in ES, LS and G2-enriched cells relative to G1-enriched, rather than STA cells (*Figure 3—figure supplement 2A*). Conversely, sequence ratios in G1 were also calculated relative to ES, LS and G2 enriched populations (*Figure 3—figure supplement 2B*). In both cases, increased copy-number variation in the proximity of chromosome ends was seen in G1 relative to ES, LS and G2. This further demonstrates that these regions serve as DNA replication initiation sites, and that such sites seem to have a relatively higher usage during G1. Next, we compared MFA-seq signal in EXP/STA cells using both copy-number change and Z-score analyses. Irrespective of the analysis strategy, telomere-proximal signal, although less pronounced than the central enriched regions, was apparent in several chromosomes, though not in the majority (*Figure 3—figure supplement 3*). Consistent with this, meta-analysis of MFA-seq signal in EXP cells showed only minor signal enrichment around telomeric AcH3 sites (*Figure 3C*). Therefore, these data indicate that subtelomere DNA replication initiation is mainly not activated during ES but is a persistent activity throughout the enriched cell cycle populations. Altogether, these observations indicate the importance of local chromatin environment in modulating *L. major* DNA replication initiation in a temporal manner.

Distinct strand asymmetry and G4 distribution at sites of core and subtelomeric DNA replication initiation

To ask if genetic elements might be associated with DNA replication initiation in *L. major*, we next sought to analyse the patterns of sequence composition within and around the loci containing the above combinations of AcH3, base J and KKT1 enrichment. First, we tested for the existence of specific sequence motifs associated with the distinct patterns of predicted DNA replication initiation around these sites. As expected, MEME analysis retrieved motifs that have been associated with transcription initiation and termination, known to take place at these sites (G/T rich motifs and polypyrimidine tracks; *Figure 3—figure supplement 4*; *Clayton, 2019*). We could not find any sequence signature that correlated specifically with regions of DNA replication initiation, or that differed between the regions used during ES phase or throughout the cell cycle.

Next, we examined the patterns of intra-strand base composition skews (excess of G over C = [(G-C)/(G+C)]; excess of T over A = [(T-A)/(T+A)]). In many organisms, G skew and, to a lesser extent, T skew values switch polarity around replication initiation sites, with the leading strand usually associated with positive skews values (*Xia, 2018*; *Huvet et al., 2007*; *Baker et al., 2012*). With this in mind, we might expect to detect a predominance of negative G and T skew values upstream from the main DNA replication initiation regions mapped in ES cells by MFA-seq and co-localising with

AcH3/base J/KKT1 sites. Conversely, positive G and T skew values would predominate downstream from these sites. Our analysis did not indicate such a broad replication-associated pattern, but instead revealed that the major correlation of base skew in *L. major* is with transcription direction, in which the coding strand is mainly associated with positive G skew and negative T skew values (*Figure 4A*), as previously reported (*El-Sayed et al., 2005*; *McDonagh et al., 2000*). However, a closer examination of the skew patterns in the central area of the ES MFA-seq peaks revealed a highly localised pattern of polarity change, resembling the wider pattern seen around origins of replication in other organisms (see zoomed region in chromosomes 30 and 31; *Figure 4B*). This local
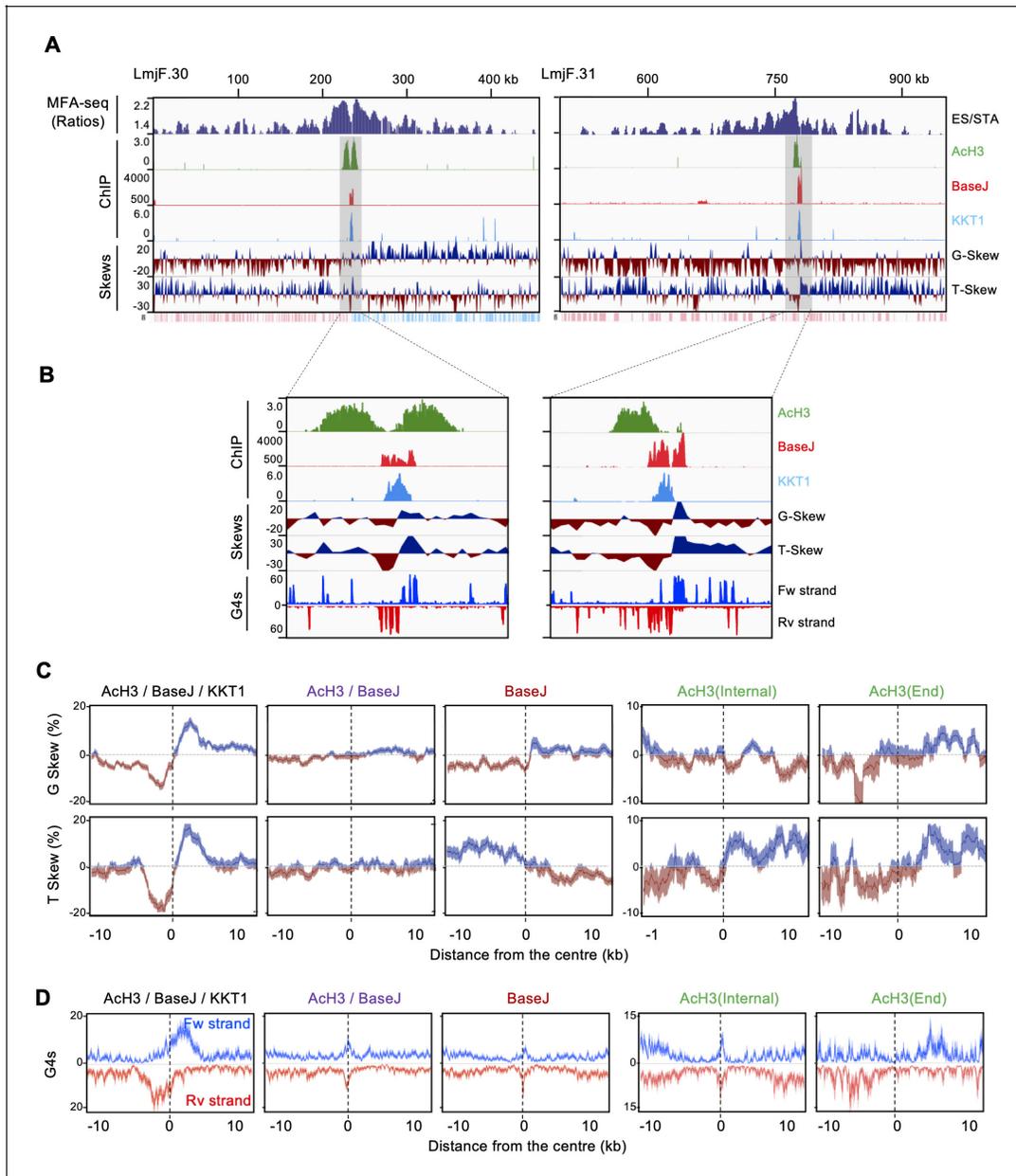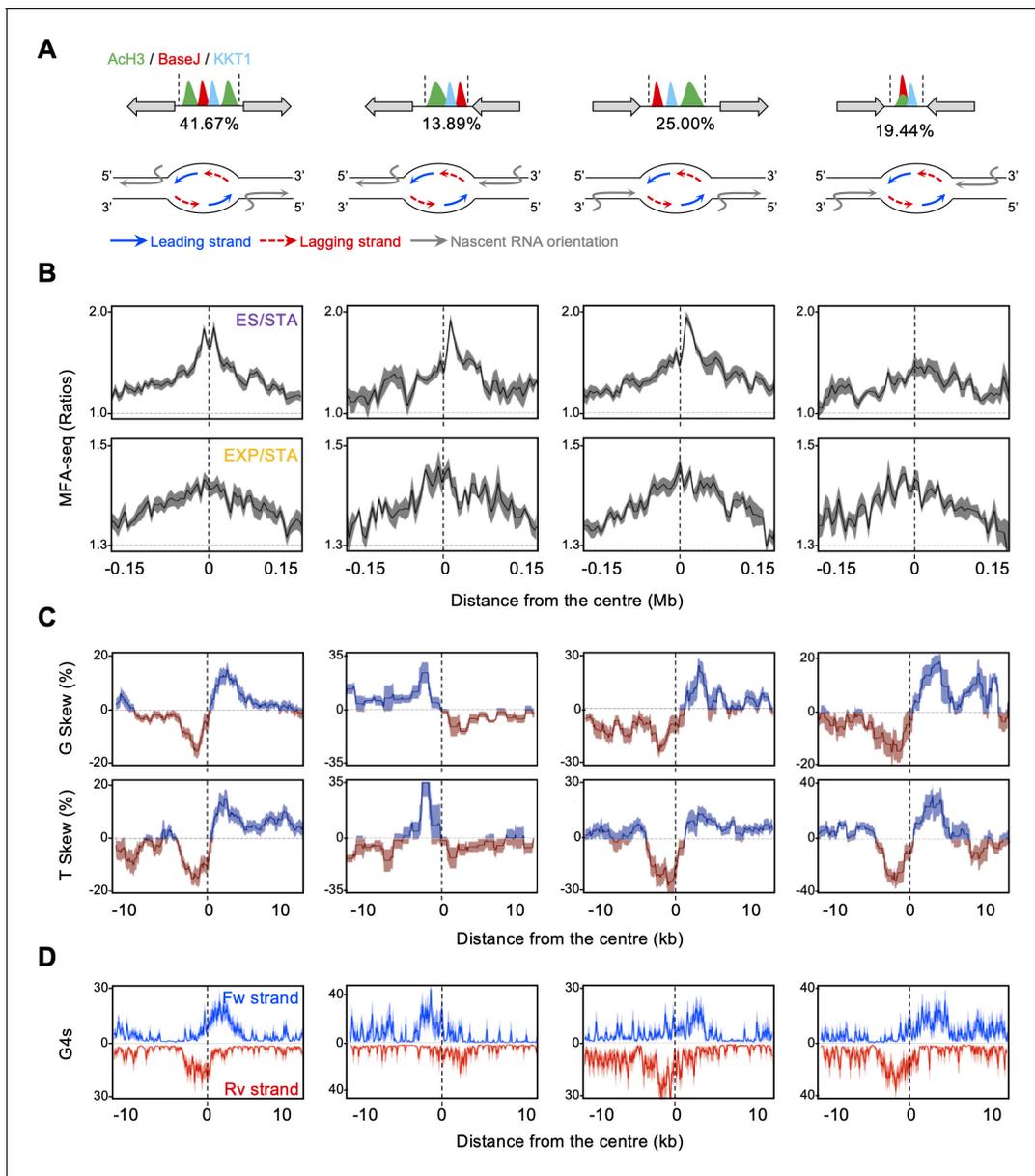

Figure 4. Predicted DNA replication initiation regions are related to distinct patterns of G skew, T skew and G4 accumulation. (**A**) Representative regions of chromosomes 30 and 31, showing MFA-seq signals in ES/STA, compared to the positioning of the indicated chromatin markers; G and T skews are also shown; the bottom track in each panel indicates annotated CDSs (blue: transcribed from left to right; pink: transcribed from right to left). (**B**) Shaded areas in A are magnified, and now include G4 peaks distribution for both forward (Fw strand) and reverse (Rv strand) DNA strands. (**C** and **D**) Metaplots of global G and T skews (**C**) and G4 distribution (**D**) ± 10 kb from the centre of regions containing the indicated combination of chromatin markers; light-coloured areas around lines indicates SEM.

change, from negative to positive in the G and T skew values across the centre of the regions of AcH3/base J/KKT1 enrichment, could be seen even when two adjacent transcription units are transcribed in the same direction (see zoomed region in chromosome 31, which is entirely transcribed unidirectionally; *Figure 4B*). Accordingly, meta-analysis revealed the same localised pattern of polarity change, for both G and T skew, when every ES MFA-seq site was compared (*Figure 4C*). A similar pattern was seen across AcH3 sites near chromosome ends, although both the amplitude and sharpness of the polarity change in base skew was less pronounced than that seen around AcH3/base J/KKT1 sites. In contrast, only T skew showed the same global polarity change around chromosome-internal AcH3 sites (*Figure 4B*). Finally, we did not find any obviously related profile of base skew polarity change around either AcH3/base J or base J sites (*Figure 4C*). These data provide clear evidence for base composition bias, consistent with DNA replication initiation, at a central region in each of the 36 *L. major* chromosomes. In addition, differences in base composition around core and subtelomeric AcH3 sites are indicative of extensive DNA replication initiation at the latter but not the former.

Next, we sought to examine the potential correlation between G-quadruplexes (G4s) and the distinct DNA replication initiation activities we see across the *L. major* genome. G4s are secondary DNA structures that can arise in single-stranded, guanine-rich DNA and have been implicated in various biological processes in eukaryotic cells, including DNA replication (*Cayrou et al., 2015*; *Besnard et al., 2012*). G4s have been predicted to colocalise with SNS-seq signal in *Leishmania* (*Lombraña et al., 2016*), but their relationship with DNA replication initiation regions mapped by MFA-seq has not been explored. Thus, we examined the distribution of experimentally mapped G4s (*Marsico et al., 2019*) around all sites with the differing combinations of AcH3, base J and KKT1 binding (*Figure 4D*). Our analyses showed a strong positional preference for G4 distribution around the midpoint where G and T skews change polarity across the AcH3/base J/KKT1 sites (see zoomed regions in chromosomes 30 and 31, *Figure 4B*). Upstream of the midpoint of these sites, a grouping of G4 peaks could be seen on the reverse DNA strand, while downstream of the midpoint a similar grouping was seen on the forward strand. Meta-analysis showed that these localised patterns of G4 distribution were reflected in every ES MFA-seq mapped peak, with a striking correlation relative to the base skew polarity change (*Figure 4D*). A similar global pattern of G4 distribution, although more diffuse, was observed around subtelomeric AcH3 sites (*Figure 4D*). At all these predicted sites of DNA replication the G4 distribution differed from the pattern seen at chromosome-internal AcH3 sites, as well as at AcH3/base J and base J sites, where G4s were arranged as sharp peaks on both the forward and reverse DNA strands that aligned at the central point of enrichment (*Figure 4D*). Altogether, these observations indicate that the region of *L. major* DNA replication initiation predicted by MFA-seq have imprinted, or have been shaped by, distinct local profiles of DNA intrastrand asymmetries and arrangement of G4s. Moreover, the patterns of these genome features reflect the predicted differences, in space and time, of DNA replication activity at chromosome-internal and subtelomeric regions.

Multigene transcription surrounding AcH3/base J/KKT1 sites can have three distinct configurations: the transcription units diverge (transcription initiation of both units within the locus), converge (transcription termination) or are arranged head to tail (combined transcription initiation and termination) (*Figure 5A*). To ask about the relationship between replication initiation and transcription activity, we plotted the G/T skews and G4 profile around each of these configurations. As expected, at all sites a pronounced MFA-seq peak was seen when comparing ES or EXP cells with STA cells, with no evidence for differential DNA replication initiation activity between them (*Figure 5B*). However, by separating these sites, it became clear that the patterns of G skew, T skew (*Figure 5C*) and G4 distribution (*Figure 5D*) were mainly dictated by the leading strand of the predicted bidirectional DNA replication fork, not by the lagging strand, irrespective of the surrounding transcription direction. Moreover, the analysis suggested that leading strand DNA replication and transcription are always co-directional on the same template DNA strand around these chromosome-central putative origins. Therefore, this analysis advances the correlation between DNA replication initiation activity and local DNA strand asymmetry and G4 distribution and suggests the existence of selective pressure to avoid clashes between DNA replisome and RNA polymerase activities at regions of Leishmania DNA replication initiation.

Figure 5. Local patterns of G and T skews at early S phase replication initiation are associated with co-directional transcription and DNA replication activity. (A) Top: a schematic illustration of sites containing simultaneous deposition of AcH3, BaseJ and KKT1, with gray arrows indicating all possible configurations of surrounding multigenic transcription direction; numbers below each scheme indicates the proportion of each configuration. Bottom: schematic illustration of transcription and replication organisation on the DNA strands at each of these sites. (B), (C and D) MFA-seq profile (ES/STA and EXP/STA), base skews (G and T) and G4 pattern (forward and reverse DNA strands) are shown, respectively, for each of these sites as metaplots.

DNA replication initiation around chromosome ends is more susceptible to replication stress

The data described above suggest differences in the timing and location of chromosome-internal and subtelomeric DNA replication initiation, but do not address if they use shared or separate machineries. To begin to answer this question, we asked if and how replication stress could affect the DNA replication programme in *L. major*. To do so, we performed MFA-seq analysis on cells sorted into ES and G2/M after release from replication arrest with 5 mM hydroxyurea (HU), which depletes the intracellular dNTP pool (*Figure 6A*). To examine DNA replication, we performed MFA-seq, comparing read depth relative to STA cells. We first searched for the appearance of new repli-cation initiation sites, perhaps indicative of activation of dormant origins. At only three sites could
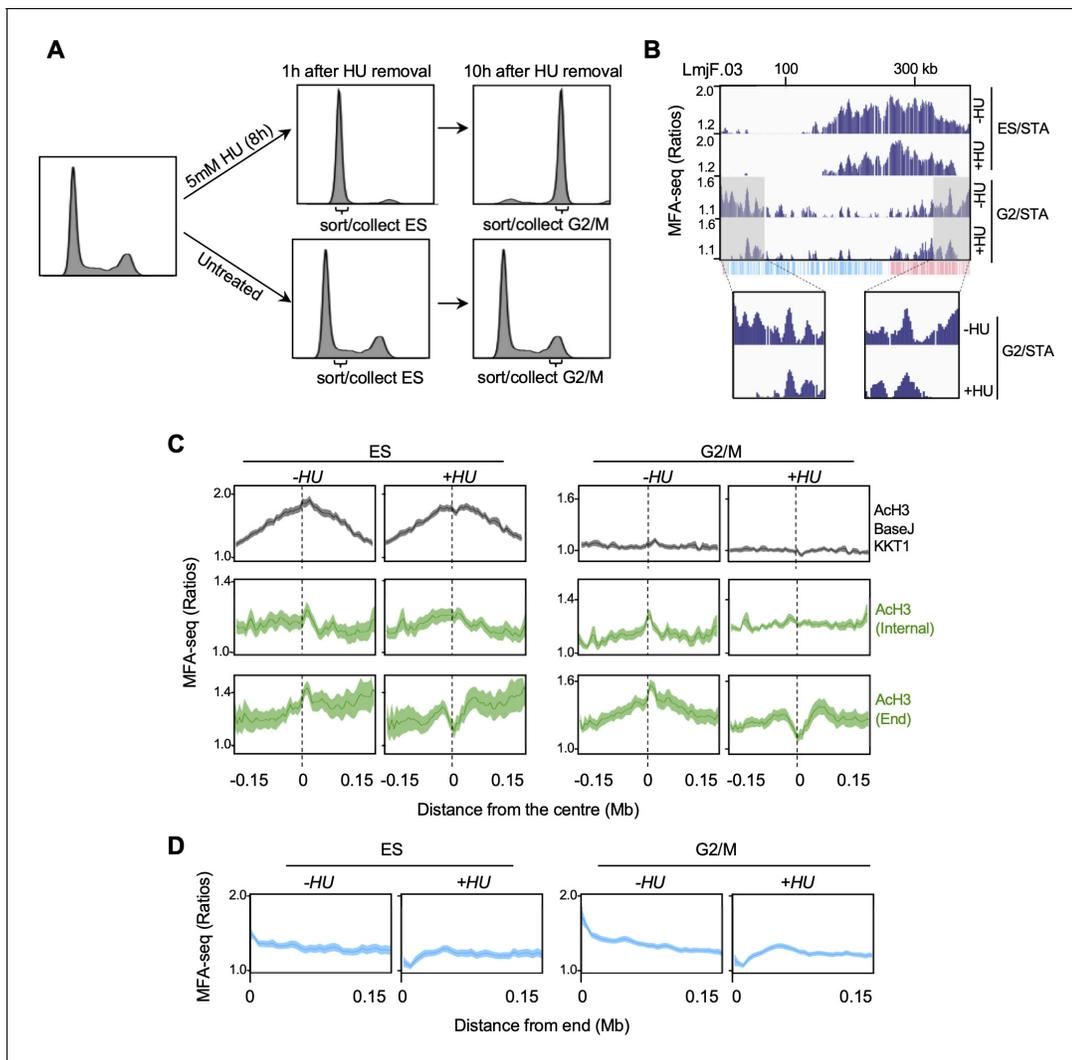
Figure 6. DNA replication close to chromosome ends is more sensitive to replication stress than replication at chromosome-central loci. (A) Cells were left untreated or treated with hydroxyurea (HU) and, at the indicated times after HU removal, sorted and the indicated cell cycle stages were collected; plots show FACS profiles of the cell populations after PI staining. (B) MFA-seq profile for the indicated phases of the cell cycle relative to STA, with or without prior HU treatment, in the indicated region of chromosome 3; shaded areas are magnified in the boxes at the bottom to highlight the differences in MFA-seq signal at regions close to chromosome end, upon HU treatment; the bottom track indicates annotated CDSs (blue: transcribed from left to right; pink: transcribed from right to left). (C) Metaplots of global MFA-seq signal, in the indicated phases of the cell cycle, ±0.15 Mb from the centre of regions containing the indicated combination of chromatin markers; lines represent mean read ratios and light-coloured areas indicate SEM. (D) Metaplots of global MFA-seq signal for the indicated phases of cell cycle relative to STA, with or without prior HU treatment, across 0.15 Mb of all chromosome ends.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. MFA-seq profile of cells at the indicated phases of cell cycle, with or without prior HU treatment; red arrows indicate sites with increased MFA-seq signal in HU treated cells, as compared with not treated cells.

Figure supplement 2. Extended data for *Figure 6*.

we detect an extremely mild increase in MFA-seq signal in ES or G2/M cells after HU treatment compared with non-HU treated cells (red arrows in *Figure 6—figure supplement 1*). Though all these sites were located at the boundaries of multigene transcription units, like the single predominant ES MFA-seq peak in each chromosome, these data suggest that *L. major* has a very limited capacity to initiate DNA replication initiation from other loci in ES phase, though when it occurs such a reaction also predominantly localises to sites of transcription initiation or termination.

Next, we looked for changes in the previously detected sites of DNA synthesis across the genome after HU treatment. First, we examined DNA replication in ES cells around the single AcH3/base J/KKT1 site in each chromosome. Mapping to each chromosome (*Figure 6B*), as well as meta-analysis of MFA-seq signal in all chromosomes (*Figure 6C*), did not reveal any change in MFA-seq peak profile in the HU treated cells, suggesting that DNA replication emanating from these regions is largely unaffected by this level of HU. In contrast, we saw pronounced changes in subtelomere replication, since a decrease in MFA-seq signal at the chromosome ends during G2/M after release from HU was detected in individual chromosomes (*Figure 6B*). Meta-analysis revealed the extent of this perturbation. First, a marked loss of MFA-seq signal, visible as early as in ES, was seen around the subtelomeric AcH3 sites upon HU treatment (*Figure 6C*; *Figure 6—figure supplement 2*), whereas no such clear effect was seen around chromosome-internal AcH3 sites (*Figure 6C*). Second, meta-analysis of MFA-seq signal at the ends of every chromosome (*Figure 6D*) showed that the HU-induced loss of the replication signal extended for ~50 kb. Unsurprisingly, given the lack of clearly detectable MFA-seq predicted DNA replication at AcH3/base J or base J sites, meta-analysis did not show any significant changes in global MFA-seq signal at either of these type of loci (*Figure 6—figure supplement 2*). These observations reveal that not all DNA replication initiation sites are equally affected by replication stress, with replication activity proximal to chromosome ends being more susceptible to HU treatment than replication from chromosome-internal regions. Moreover, the data reinforce the suggestion of differences in the nature of DNA replication around subtelomeric AcH3 sites and chromosome-internal AcH3/base J/KKT1 sites.

Replication initiation at chromosome ends requires RAD9 and HUS1

To begin to explore the factors needed for Leishmania genome duplication, we wondered if an explanation for the newly described subtelomeric DNA replication might be that, in this chromosome environment, DNA replication is under constitutively higher levels of stress when compared with DNA replication from the chromosome-internal AcH3/base J/KKT1 regions. If so, we reasoned that subtelomeric DNA replication might have a greater dependence on the replication stress response machinery. RAD9 and HUS1 are two components of the 9-1-1 clamp, which is a cell cycle checkpoint complex required for replicative DNA synthesis in *L. major* (*Damasceno et al., 2016*; *Damasceno et al., 2018*). As a result, we asked if RAD9 or HUS1 deficiency would alter the profile of *L. major* DNA replication. To do this, we isolated ES- and G2/M-phase cells from HUS1 and RAD9 heterozygous (+/-) mutants and performed MFA-seq as before. Strikingly, both visual inspection of individual chromosomes (*Figure 7A*) and meta-analysis (*Figure 7B and C*, *Figure 7—figure supplement 1*) showed a dramatic decrease in MFA-seq signal around subtelomeric AcH3 sites and proximal to the chromosome ends in RAD9$^{+/-}$ cells, in both ES and G2/M. In contrast, no obvious change in MFA-seq signal was seen around chromosome-internal AcH3 sites in G2/M RAD9$^{+/-}$ cells and, even more strikingly, no change in MFA-seq signal around AcH3/base J/KKT1 was detected in the RAD9$^{+/-}$ cells (*Figure 7B*). In HUS1$^{+/-}$ cells, a milder decrease in MFA-seq signal was observed, which was apparent only for the subtelomeric AcH3 sites during G2/M (*Figure 7B* and *Figure 7C*). Consistently, decreased MFA-seq signal near telomeres was also seen when directly comparing sequence abundance in G2-enriched populations from both mutants and wild type (WT) cells (*Figure 7—figure supplement 2*). No alteration in the MFA-seq signal was detected around AcH3/base J or base J sites in either of the mutants (*Figure 7—figure supplement 1*). These data indicate DNA replication initiation close to chromosome ends, but not at the distinct chromosome-internal initiation regions, requires the action of the 9-1-1 complex.

Discussion

In this study, we have analysed the orchestration of Leishmania genome replication in space and time. Our data show that genome duplication predominantly starts from a single region in every chromosome in early S phase, proceeding towards the chromosome ends, consistent with the observed size dependence of *L. major* chromosome replication timing. However, replication of the chromosomes is rarely completed by the end of S phase, and instead either continues from the same sites as cells transit through G2/M back to G1, or new sites of DNA replication initiation are activated in the chromosome subtelomeres during late S phase or in G2/M and G1 stages of the cell
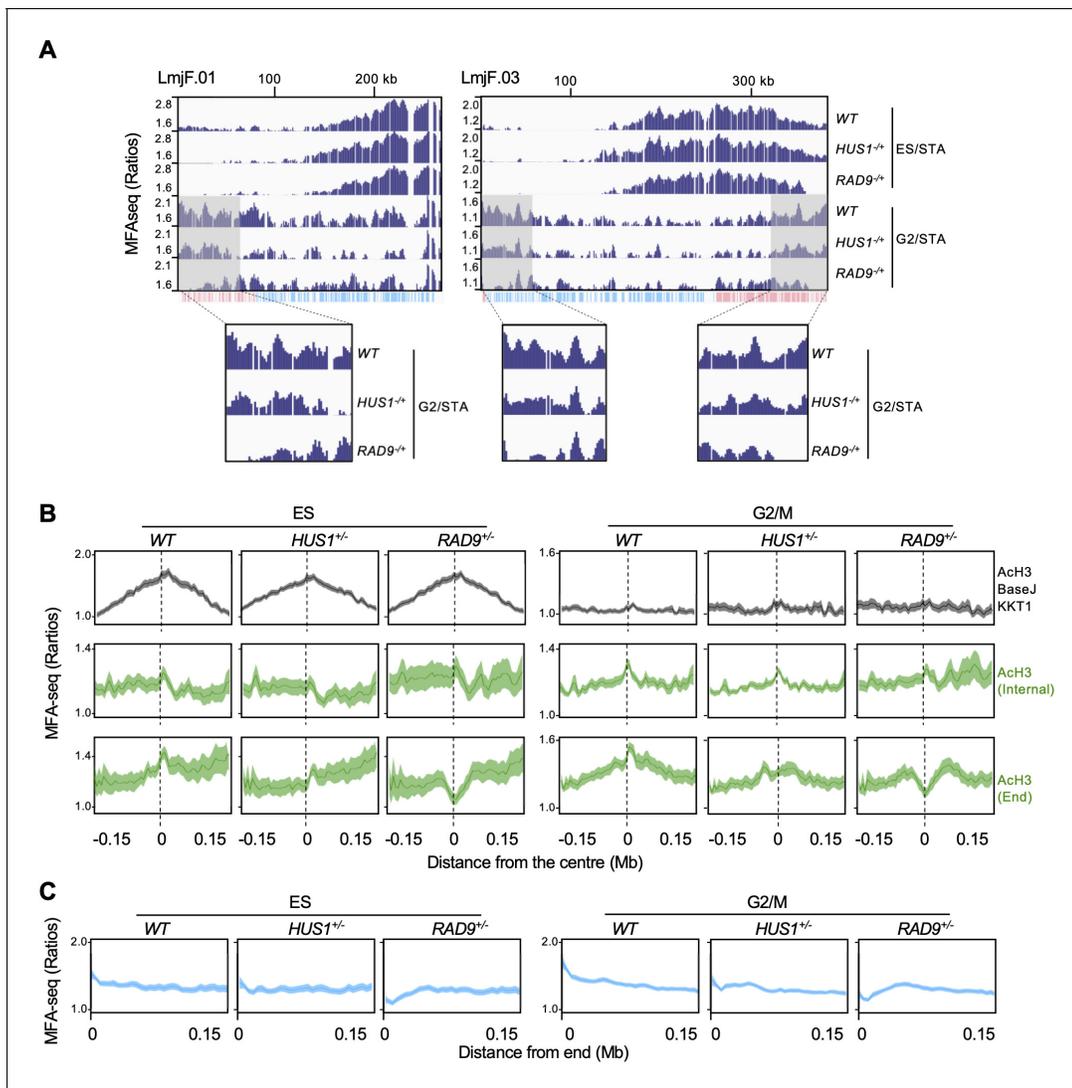
Figure 7. RAD9 and HUS1 are required for DNA replication near chromosome ends. (**A**) MFA-seq profile of wild type (*WT*), *RAD9*[+/-] and *HUS1*[+/-] cells at the indicated phases of cell cycle across chromosomes 1 and 3; shaded areas are magnified in the boxes at the bottom to highlight the differences in MFA-seq signal at regions close to chromosome ends; the bottom track indicates annotated CDSs (blue: transcribed from left to right; pink: transcribed from right to left). (**B**) Metaplots of global MFA-seq signal of *WT*, *RAD9*[+/-] and *HUS1*[+/-] cells, in the indicated phases of the cell cycle,±0.15 Mb from the centre of regions containing the indicated combination of chromatin markers. (**C**) Metaplots of global MFA-seq signal of *WT, RAD9+/-* and *HUS1+/-* cells, in the indicated phases of cell cycle, at the terminal 0.15 Mb of all chromosomes. In (**B** and **C**), lines represent mean read ratios and the lighter shaded areas indicate SEM.

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. Extended data for *Figure 7*.

Figure supplement 2. MFA-seq signal near chromosome ends upon deficiency of *HUS1* and *RAD9*.

cycle. In either model, the paucity of DNA replication initiation that occurs early in S phase is counterbalanced by DNA replication that can extend beyond conventional S phase.

The view that DNA replication activity is restricted to S phase is being increasingly challenged. Mitotic DNA repair synthesis (MiDAS) has been documented as a means to complete replication of hard-to-duplicate genome features such as common fragile sites and telomeres (*Özer and Hickson, 2018*). Moreover, in the unperturbed cell cycles of both yeast (*Ivanova et al., 2020*; *Torres-Rosell et al., 2007*) and mammalian cells (*Moreno et al., 2016*) onset of mitosis has been documented in the presence of under-replicated genomic regions, leading to DNA synthesis in offspring cells. These findings suggest that a temporal separation between DNA replication and chromosome

segregation does not seem to be a strict rule for eukaryotic cells. However, DNA synthesis outside S-phase is particularly prevalent in cells exhibiting aneuploidy (*Minocherhomji et al., 2015*; *Özer et al., 2018*), suggesting it may be less effective at maintaining genome integrity. Here, we provide evidence that *L. major*, a eukaryotic organism with constitutive mosaic aneuploidy (at least in parasite cells derived from, or proliferating in the insect vector) (*Prieto Barja et al., 2017*; *Dumetz et al., 2017*), achieves full genome duplication using DNA replication that can continue beyond S phase, including G2/M and G1. This finding, in a eukaryotic microbe that is evolutionarily distant from yeast and mammals, suggests that mitosis and cell division prior to complete genome duplication might not necessarily be an inherent feature of genome maintenance and transmission in eukaryotes. In addition, a reliance on DNA replication outside S phase may explain the widespread aneuploidy found in Leishmania. Despite segregation of under-replicated chromosomal regions being mutagenic (*Durkin and Glover, 2007*; *Debatisse et al., 2012*; *Song et al., 2014*), increased cell division rate in these circumstances may compensate for loss of fitness due to mutations. Also, if segregation of under-replicated loci is limited to specific genome compartments, such as the telomeres and subtelomeres, these may serve as a genetic playground, with higher mutation and copy-number variation allowing for increased genetic diversity, as has been proposed in yeast (*Ivanova et al., 2020*; *Brown et al., 2010*). Consistent with this idea, our data suggest that chromosome subtelomeric DNA replication is mainly detectable at the temporal extremes of the conventional Leishmania S phase (either very early in S phase, as seen with the G1 enriched population; or very late in S phase, as seen with the G2/M enriched populations), and such regions have been demonstrated to be particularly prone to copy- number variation (*Bussotti et al., 2018*). Whether post-S phase replication of subtelomeres is widely found in parasites is unclear and identification of molecular cell cycle stages markers will be crucial to determine this. Nonetheless, it is notable that subtelomeric compartments of the *T. brucei* genome share little homology between chromosome homologues, at least in part due to rearrangements among VSG genes (*Müller et al., 2018*). The subtelomeres of *Plasmodium* are also locations of immune response gene diversification, and it may therefore be valuable to dissect the timing of their replication. Indeed, it will be important to ask if core and subtelomere replication timing in any of these parasites, including Leishmania, is influenced by genome organisation in the nucleus (*Müller et al., 2018*; *Bunnik et al., 2019*).

This work also has implications about Leishmania cell cycle checkpoints. To ensure that genome replication is completed before cell division, cells may execute checkpoints to detect ongoing replication forks as cells enter mitosis. If such a checkpoint is operational at all in Leishmania, it must be permissive enough to allow cells to undergo mitosis with considerable ongoing DNA replication. In fact, even if such a cell cycle checkpoint is used, it may not be homogenous across the genome, given the pronounced sensitivity to replication stress at AcH3 sites closer to chromosome ends, contrasting with the mild sensitivity of internal AcH3 sites. This dichotomy may be due to the abundance of repetitive elements at parasite sub-telomeres (*Wickstead et al., 2003*), a known source of replication impediment and genomic instability (*Fouché et al., 2006*; *McMurray, 2010*; *Richard et al., 2008*). Perhaps, in *L. major* the 9-1-1 complex, or its subunits, participates in a specific DNA replication checkpoint responsible for detecting and protecting replication forks stalled by the subtelomeric repeats, allowing DNA synthesis to continue while undergoing mitosis. Such permissiveness would be compatible with the structural and functional diversification of the 9-1-1 subunits (*Damasceno et al., 2016*; *Damasceno et al., 2018*). In this view of the data, DNA replication detected at the subtelomeres may be an extension of replication forks emanating from chromosome-central initiation regions (see below), which are activated in early S phase and are unaffected by HU-induced replication stress. However, it is also possible that the subtelomeric and chromosome-central DNA replication reactions are distinct processes (as discussed below).

Another implication of these findings relates to the coordination of the Leishmania DNA replication process itself. Currently, we cannot establish, among the MFA-seq enriched regions we have mapped, which are *bona fide* origins of replication, recognised by ORC, and which, if any, might correspond with other DNA synthesis events (e.g. MiDAS), which may be ubiquitous throughout the cell cycle. Unlike in *T. brucei* (*Tiengwe et al., 2012a*), binding sites of ORC components have not been mapped in the *L. major* genome, and so it remains to be determined if and how they overlap with the MFA-seq signals we describe. Indeed, despite conservation of predicted ORC components (*Tiengwe et al., 2012b*; *Tiengwe et al., 2014*; *Marques and McCulloch, 2018*; *Marques et al., 2016*), no functional analysis of the initiator has been described in Leishmania (*da Silva et al., 2017*),

meaning we cannot rule out innovations in ORC that could provide S-phase-uncoupled licensing and firing of replication initiation sites, allowing DNA replication outside S phase. Nonetheless, this work shows that the single MFA-seq peak found in each chromosome – the only genomic region clearly activated early in S phase – coincides with KKT1 localised to the boundaries of the transcription units. Here, there is a parallel with *T. brucei*, where the earliest acting origin in each chromosome is bound by ORC and coincides with the mapped centromere (*Tiengwe et al., 2012a*), where at least two KKT factors localise (*Akiyoshi and Gull, 2014*). Moreover, many of the single S-phase MFA-seq peaks in Leishmania are positionally conserved in *T. brucei* (*Marques et al., 2015*), arguing that the putative centromere-focused MFA-seq peaks in Leishmania are ORC-defined origins. In contrast, the subtelomeric regions of intra- and potential post-S-phase DNA replication can only be correlated with AcH3, and we cannot be sure that this form of chromatin is actually a determinant of replication initiation; for instance, DNA replication may emanate from the telomeres. One possibility is that sub-telomeric DNA replication is not ORC-derived, but instead derives from DNA repair that is required for the stability of subtelomeric regions, by either controlling copy-number variation in these regions (*Laffitte et al., 2016b*) or by directly driving replication initiation. Such an activity could be related to MiDAS, but may also differ since it appears to be a constitutive component of the Leishmania genome replication programme and looks like it could be related to what has been proposed as a strategy for telomere maintenance (*Bussotti et al., 2018*). The loss of subtelomeric DNA replication in *L. major* cells deficient in RAD9 may provide an insight into the repair machinery involved, since there are parallels with the effects of these mutants and the need for DNA repair activity to direct DNA replication when origins are deleted (*Theis et al., 2010*) or in conditions of replication stress (*Bacal et al., 2018*). Further work will be needed to test if DNA repair, such as homologous recombination, which is known to direct gene amplification (*Laffitte et al., 2016a*), underlies subtelomeric and putative post-S-phase Leishmania DNA replication.

Similar to other eukaryotes, this and previous MFA-seq analysis (*Marques et al., 2015*), as well as SNS-seq mapping (*Lombraña et al., 2016*), have not found association between DNA replication initiation and any specific DNA sequence in *L. major*. Instead, these studies have consistently suggested that features related to transcription activity in *L. major* play a major role in DNA replication initiation. These findings reinforce the idea that Leishmania DNA replication may be initiated in an opportunistic fashion, coupled with transcription initiation and termination events: an association also seen in other eukaryotes (*Lombraña et al., 2013*; *Chen et al., 2019*), although the mechanisms might differ. It is unclear how epigenetic dynamics dictate DNA replication activity in *L. major*. AcH3 seems to correlate with persistent replication initiation throughout the cell cycle. However, when in combination with base J and KKT1, coordinated early S-phase DNA replication initiation is seen, while base J alone is clearly not sufficient for initiation. It should be noted, however, that the ChIP datasets for AcH3, base J and KKT1 lack cell cycle resolution. Thus, it is possible that the order by which each of these modifications is deposited during the cell cycle is a major determinant of the DNA replication initiation outcomes around them. Like in other eukaryotes (*Besnard et al., 2012*; *Comoglio et al., 2015*; *Prorok et al., 2019*), the specific arrangement of G4s around the sites of replication initiation may also be relevant, not only for the kinetics of these forms of chromatin deposition, but also other unknown epigenetic markers and replication-associated factors. Correlation between transcription and GC and AT skews in trypanosomatids has been reported before (*El-Sayed et al., 2005*; *McDonagh et al., 2000*) and is mainly a consequence of constitutive transcription in these organisms. Replication is more restricted in time, and thus only the most efficient replication initiation sites (those around AcH3/base J/KKT1 and, to a lesser extent, AcH3 alone) were locally imprinted with the skew patterns related to DNA replication initiation activity. Distinct GC and AT skews correlate with distinct DNA replication initiation patterns, indicating they may encode some regulatory information themselves, contributing to the space and time orchestration of the DNA replication programme. Alternatively, the skew differences between chromosome-internal and subtelomeric DNA replication initiation sites might have been shaped by the different nature of reactions driving DNA replication from each of these sites.

MFA-seq analysis, as performed previously in Leishmania (*Marques et al., 2015*), is a laborious, time-consuming and expensive technique, requiring purification of cells in different cell cycle stages by flow cytometry, or other techniques such as elutriation. Flow cytometry suggests that in exponentially growing cultures of *L. major*, 25–35% of cells are undergoing genome replication (proportion of cells in S-phase; *Figure 1A*). Thus, we anticipated that the inherent resolution of deep sequencing

should allow us to map DNA replication initiation sites by directly comparing unsorted exponentially growing cells with stationary cells. The success of this approach demonstrates its convenience, which represents a saving in time and cost, which may be particularly useful for comparing conditions or mutants predicted to change DNA replication dynamics. The strategy of genome replication used by Leishmania has been unclear, primarily because of significant discrepancies between previous MFA-seq (*Marques et al., 2015*) and SNS-seq (*Lombraña et al., 2016*) analyses, almost certainly due to differences in sensitivity between the two techniques. While MFA-seq may detect only the most frequent duplication events in the population, SNS-seq is highly sensitive, capable of detecting rarer events. In addition, while SNS-seq has been performed in asynchronous cells, previous MFA-seq used populations enriched for specific cell cycle stages. While the new approach for MFA-seq described here does not fully resolve the discrepancies in predictions of DNA replication activity by these different methodologies (discussed in recent reviews) (*Marques and McCulloch, 2018*; *da Silva et al., 2017*; *Rocha-Granados and Klingbeil, 2016*), the detection of subtelomeric DNA replication provides at least a partial explanation for DNA combing work that has observed two regions of DNA synthesis in a single molecule (*Lombraña et al., 2016*; *Stanojcic et al., 2016*). More importantly, this work suggests that subtelomeric, potentially post-S-phase replication is a core part of genome transmission in Leishmania, and understanding the nature of this form of DNA replication may have implications for understanding genome plasticity in this microbe, and others.

Materials and methods

Cell lines and culture
Promastigote forms of *Leishmania major* strain Friedlin were used for MFA-seq data presented in *Figures 1*, *2* and *3*, while *L. major* strain LT252 (MHOM/IR/1983/IR) were used for MFA-seq data presented in *Figures 6* and *7*. Generation of $RAD9^{+/-}$ and $HUS1^{+/-}$ cell lines was previously described (*Damasceno et al., 2013*; *Damasceno et al., 2016*). Cells were cultured at 25°C in HOMEM or M199 medium supplemented with 10% heat inactivated foetal calf serum and 1% penicillin/ streptomycin.

IdU incorporation
Cells were incubated with 150 µM IdU for 30 min and then fixed at −20°C with a mixture (7:3) of ethanol and 1x PBS for at least 16 hr. Next, cells were rinsed with washing buffer (1x PBS supplemented with 1% BSA) and DNA was denatured for 30 min with 2N HCL, followed by neutralisation with phosphate buffer (0.2 M Na_2HPO_4, 0.2 M KH_2PO_4, pH 7.4). Detection of incorporated IdU was performed with anti-BrdU antibody (diluted in washing buffer supplemented with 0.2% Tween-20) for 1 hr at room temperature. After washing, cells were incubated with anti-mouse secondary antibody conjugated with Alexa Fluor 488 (diluted in washing buffer supplemented with 0.2% Tween-20) for 1 hr at room temperature and then washed. Finally, cells were stained with 1xPBS supplemented with 10 µg.mL^{-1} Propidium Iodide (PI) and 10 µg.mL^{-1} RNAse A and filtered through a 35 µm nylon mesh. FACSCelesta (BD Biosciences) was used for data acquisition and FlowJo software for data analysis. Negative control (omission of anti-BrdU antibody during IdU detection step) was included in each experiment and used to draw gates to discriminate positive and negative events.

EdU incorporation
Cells were incubated for 1 hr with 10 µM of EdU (Click-iT; Thermo Scientific), then fixed with 3.7% paraformaldehyde for 15 min, adhered into poly-L-lysine coated slides, followed by permeabilisation with 0.5% TritonX100 for 20 min. After washing with PBS supplemented with 3% BSA, cells were subjected to Click-iT reaction, following manufacturers' instructions. DNA was stained with Hoechst 33342. Images were acquired with an SP5 confocal microscope (Leica) and processed with ImageJ software.

Fluorescent activated cell sorting (FACS) and genomic DNA extraction
FACS and DNA extraction were performed as previously described (*Marques et al., 2015*). Briefly, exponentially growing cells were collected by centrifugation and washed in 1 × PBS supplemented with 5 mM EDTA. Cells were fixed at 4°C in a mixture (7:3) of methanol and 1 × PBS. Prior to sorting,

fixed cells were collected by centrifugation and washed once in 1 × PBS supplemented with 5 mM EDTA, re-suspended in 1 × PBS containing 5 mM EDTA, 10 μg/ml PI and 10 μg/ml RNase A, and passed through a 35 μm nylon mesh. A FACSAria I cell sorter (BD Biosciences) was used to sort cells into lysis buffer (1 M NaCl; 10 mM EDTA; 50 mM Tris–HCl pH 8.0; 0.5% SDS; 0.4 mg/ml proteinase K; 0.8 μg/ml glycogen). Then, sorted cells were incubated for 2 hr at 55°C and genomic DNA was extracted using Blood and Tissue DNA extraction kit (Qiagen), by omitting the lysis step. Genomic DNA from non-sorted (exponentially growing or stationary) cells was also extracted using Blood and Tissue DNA extraction kit (Qiagen), following the manufacture instructions.

Whole genome sequencing

Whole genome sequencing of sorted cells used in *Figures 1*, *2* and *3* was previously described (*Marques et al., 2015*). Whole genome sequencing of exponentially growing, stationary and sorted cells (*Figures 4* and *5*) was performed by Eurofins (www.eurofinsgenomics.eu) using Illumina MiSeq paired-end 75 bp or 100 bp sequencing system (Illumina). In order to eliminate differences due to batch effects, each of the early S, late S, G1 and G2 samples per strain/cell line were multiplexed and sequenced in the same run. Sequencing data were uploaded to the Galaxy web platform (usegalaxy.org) for processing (*Afgan et al., 2018*). Quality control was performed with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimomatic (*Bolger et al., 2014*) was used to remove adapter sequences from reads. Reads were mapped to the *L. major Friedlin* reference genome, version 39 (Tritrypdb -http://tritrypdb.org/tritrypdb/), using BWA-mem (*Li and Durbin, 2009*). Reads with a mapping quality score <30 were discarded using SAMtools (*Li et al., 2009*). A summary of filtered reads distribution is shown in *Figure 1—figure supplement 1*.

Marker frequency analysis (MFA-seq)

After alignment and filtering, reads were compared using the method described previously (*Marques et al., 2015*), with modifications. Briefly, reads were binned in 1.0 kb windows along chromosomes. The number of reads in each bin was used to calculate the ratio between each cell cycle stage (or exponentially growing cells) versus stationary cells, scaled for the total size of the read library. Bins with less than 100 reads in either sample were discarded, as were reads that had a quality score below 30. To reduce noise due to collapsed regions, bins with a ratio above 2.8 were discarded. Also, bins overlapping with other problematic mapping regions, as previously described (*Lombraña et al., 2016*), were also removed. Due to strain differences, further aberrant mapping regions (available upon request) were excluded from samples from LT252 (MHOM/IR/1983/IR). Read depth ratios were normalised as described in *Batrakou et al., 2020*. Due to the prevalence of mosaic aneuploidy in Leishmania, ratio values were also converted into Z scores using in-house R-scripts and the equivalent plots are shown in figure supplements; in either case the data were plotted in 5 kb sliding windows for each individual chromosome, and MFA-seq profiles were represented in a graphical form using Gviz (*Hahne and Ivanek, 2016*). An overview of data analysis is shown in *Figure 1—figure supplement 3*. A comparison between data representation as ratio or Z scores is shown in *Figure 1—figure supplements 4* and *5*.

Datasets from other studies

ChIP data for AcH3, base J and KKT1 were previously published (*Thomas et al., 2009*; *van Luenen et al., 2012*; *Garcia-Silva et al., 2017*). Experimental mapping of G4-quadruplex in *L. major* was published previously (*Marsico et al., 2019*). Whole genome sequencing used for MFA-seq data presented in *Figures 1*, *2* and *3* was taken from *Marques et al., 2015*.

GC and AT skew analysis

G and T skews were calculated as (G-C)/(G+C) and (T-A)/(A+T), respectively. Calculations were performed in 1 kb windows using in-house python scripts and subsequently converted into bigwig files. Profiles were represented in a graphical form using Gviz (*Hahne and Ivanek, 2016*).

Meta-analysis

Unless otherwise stated, all underlying data for metaplots, heatmaps and clustering analyses were generated using deepTools (*Ramírez et al., 2016*), in the Galaxy web platform (usegalaxy. org) (*Afgan et al., 2018*). Metaplots were generated using Prism Graphpad.

Acknowledgements

We thank all current and previous members of the McCulloch and Tosi labs for input.

Additional information

Author contributions

Jeziel Dener Damasceno, Conceptualization, Resources, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Catarina A Marques, Conceptualization, Resources, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review and editing; Dario Beraldi, Data curation, Software, Formal analysis, Validation, Investigation, Methodology, Writing - review and editing; Kathryn Crouch, Resources, Data curation, Software, Formal analysis, Supervision, Investigation, Methodology, Writing - review and editing; Craig Lapsley, Formal analysis, Validation, Investigation, Writing - review and editing; Ricardo Obonaga, Formal analysis, Investigation, Methodology; Luiz RO Tosi, Conceptualization, Resources, Supervision, Funding acquisition, Investigation, Methodology, Project administration, Writing - review and editing; Richard McCulloch, Conceptualization, Formal analysis, Supervision, Funding acquisition, Investigation, Methodology, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Jeziel Dener Damasceno (iD) https://orcid.org/0000-0003-2077-3214
Catarina A Marques (iD) http://orcid.org/0000-0003-1324-5448
Kathryn Crouch (iD) http://orcid.org/0000-0001-9310-4762
Richard McCulloch (iD) https://orcid.org/0000-0001-5739-976X

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.58030.sa1
Author response https://doi.org/10.7554/eLife.58030.sa2

Additional files

Supplementary files
• Transparent reporting form

Data availability
Sequences used in this study have been deposited in the European Nucleotide Archive. Data can be accessed using the accession number PRJEB35027.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Damasceno JD, Marques CA, Beraldi D, Crouch K, Lapsley C, Obonaga R, Tosi LR, McCulloch R	2020	Genome duplication in Leishmania major relies on unconventional subtelomeric DNA replication	https://www.ebi.ac.uk/ena/data/view/PRJEB35027	European Nucleotide Archive, PRJEB35027

References

Adl SM, Bass D, Lane CE, Lukes J, Schoch CL, Smirnov A. 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *The Journal of eukaryotic microbiology* **66**:4–119.

Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* **46**:W537–W544. DOI: https://doi.org/10.1093/nar/gky379, PMID: 29790989

Akiyoshi B, Gull K. 2014. Discovery of unconventional kinetochores in kinetoplastids. *Cell* **156**:1247–1258. DOI: https://doi.org/10.1016/j.cell.2014.01.049

Al Mamun M, Albergante L, Moreno A, Carrington JT, Blow JJ, Newman TJ. 2016. Inevitability and containment of replication errors for eukaryotic genome lengths spanning megabase to gigabase. *PNAS* **113**:E5765–E5774. DOI: https://doi.org/10.1073/pnas.1603241113, PMID: 27630194

Alver RC, Chadha GS, Blow JJ. 2014. The contribution of dormant origins to genome stability: from cell biology to human genetics. *DNA Repair* **19**:182–189. DOI: https://doi.org/10.1016/j.dnarep.2014.03.012, PMID: 24767947

Bacal J, Moriel-Carretero M, Pardo B, Barthe A, Sharma S, Chabes A, Lengronne A, Pasero P. 2018. Mrc1 and Rad9 cooperate to regulate initiation and elongation of DNA replication in response to DNA damage. *The EMBO Journal* **37**:e99319. DOI: https://doi.org/10.15252/embj.201899319, PMID: 30158111

Baker A, Audit B, Chen CL, Moindrot B, Leleu A, Guilbaud G, Rappailles A, Vaillant C, Goldar A, Mongelard F, d'Aubenton-Carafa Y, Hyrien O, Thermes C, Arneodo A. 2012. Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLOS Computational Biology* **8**:e1002443. DOI: https://doi.org/10.1371/journal.pcbi.1002443, PMID: 22496629

Batrakou DG, Müller CA, Wilson RHC, Nieduszynski CA. 2020. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. *Nature Protocols* **15**:1255–1284. DOI: https://doi.org/10.1038/s41596-019-0287-7, PMID: 32051615

Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, Lemaitre JM. 2012. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Structural & Molecular Biology* **19**:837–844. DOI: https://doi.org/10.1038/nsmb.2339, PMID: 22751019

Bleichert F, Botchan MR, Berger JM. 2017. Mechanisms for initiating cellular DNA replication. *Science* **355**:eaah6317. DOI: https://doi.org/10.1126/science.aah6317, PMID: 28209641

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**:2114–2120. DOI: https://doi.org/10.1093/bioinformatics/btu170, PMID: 24695404

Brown CA, Murray AW, Verstrepen KJ. 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Current Biology* **20**:895–903. DOI: https://doi.org/10.1016/j.cub.2010.04.027, PMID: 20471265

Bunnik EM, Venkat A, Shao J, McGovern KE, Batugedara G, Worth D, Prudhomme J, Lapp SA, Andolina C, Ross LS, Lawres L, Brady D, Sinnis P, Nosten F, Fidock DA, Wilson EH, Tewari R, Galinski MR, Ben Mamoun C, Ay F, et al. 2019. Comparative 3D genome organization in apicomplexan parasites. *PNAS* **116**:3183–3192. DOI: https://doi.org/10.1073/pnas.1810815116, PMID: 30723152

Burgers PMJ, Kunkel TA. 2017. Eukaryotic DNA replication fork. *Annual Review of Biochemistry* **86**:417–438. DOI: https://doi.org/10.1146/annurev-biochem-061516-044709, PMID: 28301743

Bussotti G, Gouzelou E, Côrtes Boité M, Kherachi I, Harrat Z, Eddaikra N, Mottram JC, Antoniou M, Christodoulou V, Bali A, Guerfali FZ, Laouini D, Mukhtar M, Dumetz F, Dujardin JC, Smirlis D, Lechat P, Pescher P, El Hamouchi A, Lemrani M, et al. 2018. *Leishmania* genome dynamics during environmental adaptation reveal Strain-Specific differences in gene copy number variation, karyotype instability, and telomeric amplification. *mBio* **9**:e01399-18. DOI: https://doi.org/10.1128/mBio.01399-18, PMID: 30401775

Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, Andrau JC, van Helden J, Méchali M. 2015. The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Research* **25**:1873–1885. DOI: https://doi.org/10.1101/gr.192799.115, PMID: 26560631

Chen YH, Keegan S, Kahli M, Tonzi P, Fenyö D, Huang TT, Smith DJ. 2019. Transcription shapes DNA replication initiation and termination in human cells. *Nature Structural & Molecular Biology* **26**:67–77. DOI: https://doi.org/10.1038/s41594-018-0171-0, PMID: 30598550

Clayton C. 2019. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biology* **9**:190072. DOI: https://doi.org/10.1098/rsob.190072, PMID: 31164043

Comoglio F, Schlumpf T, Schmid V, Rohs R, Beisel C, Paro R. 2015. High-resolution profiling of *Drosophila* replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Reports* **11**:821–834. DOI: https://doi.org/10.1016/j.celrep.2015.03.070, PMID: 25921534

da Silva MS, Pavani RS, Damasceno JD, Marques CA, McCulloch R, Tosi LRO, Elias MC. 2017. Nuclear DNA replication in trypanosomatids: there are no easy methods for solving difficult problems. *Trends in Parasitology* **33**:858–874. DOI: https://doi.org/10.1016/j.pt.2017.08.002, PMID: 28844718

Damasceno JD, Nunes VS, Tosi LR. 2013. LmHus1 is required for the DNA damage response in *leishmania major* and forms a complex with an unusual Rad9 homologue. *Molecular Microbiology* **90**:1074–1087. DOI: https://doi.org/10.1111/mmi.12418, PMID: 24118609

Damasceno JD, Obonaga R, Santos EV, Scott A, McCulloch R, Tosi LR. 2016. Functional compartmentalization of Rad9 and Hus1 reveals diverse assembly of the 9-1-1 complex components during the DNA damage response in *leishmania*. *Molecular Microbiology* **101**:1054–1068. DOI: https://doi.org/10.1111/mmi.13441, PMID: 27301589

Damasceno JD, Obonaga R, Silva GLA, Reis-Cunha JL, Duncan SM, Bartholomeu DC, Mottram JC, McCulloch R, Tosi LRO. 2018. Conditional genome engineering reveals canonical and divergent roles for the Hus1 component of the 9-1-1 complex in the maintenance of the plastic genome of leishmania. *Nucleic Acids Research* **46**:11835–11846. DOI: https://doi.org/10.1093/nar/gky1017, PMID: 30380080

Deal RB, Henikoff JG, Henikoff S. 2010. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**:1161–1164. DOI: https://doi.org/10.1126/science.1186777, PMID: 20508129

Debatisse M, Le Tallec B, Letessier A, Dutrillaux B, Brison O. 2012. Common fragile sites: mechanisms of instability revisited. *Trends in Genetics* **28**:22–32. DOI: https://doi.org/10.1016/j.tig.2011.10.003, PMID: 22094264

Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, Cesaroni M, Mendoza-Maldonado R, Giacca M, Pelicci PG. 2013. Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Research* **23**:1–11. DOI: https://doi.org/10.1101/gr.142331.112, PMID: 23187890

Dhar MK, Sehgal S, Kaul S. 2012. Structure, replication efficiency and fragility of yeast ARS elements. *Research in Microbiology* **163**:243–253. DOI: https://doi.org/10.1016/j.resmic.2012.03.003, PMID: 22504206

Dumetz F, Imamura H, Sanders M, Seblova V, Myskova J, Pescher P, Vanaerschot M, Meehan CJ, Cuypers B, De Muylder G, Späth GF, Bussotti G, Vermeesch JR, Berriman M, Cotton JA, Volf P, Dujardin JC, Domagalska MA. 2017. Modulation of aneuploidy in *Leishmania donovani* during Adaptation to Different *In Vitro* and *In Vivo* Environments and Its Impact on Gene Expression. *mBio* **8**:e00599-17. DOI: https://doi.org/10.1128/mBio.00599-17, PMID: 28536289

Durkin SG, Glover TW. 2007. Chromosome fragile sites. *Annual Review of Genetics* **41**:169–192. DOI: https://doi.org/10.1146/annurev.genet.41.042007.165900, PMID: 17608616

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, et al. 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**:404–409. DOI: https://doi.org/10.1126/science.1112181, PMID: 16020724

Fouché N, Ozgür S, Roy D, Griffith JD. 2006. Replication fork regression in repetitive DNAs. *Nucleic Acids Research* **34**:6044–6050. DOI: https://doi.org/10.1093/nar/gkl757, PMID: 17071963

Garcia-Silva MR, Sollelis L, MacPherson CR, Stanojcic S, Kuk N, Crobu L, Bringaud F, Bastien P, Pagès M, Scherf A, Sterkers Y. 2017. Identification of the centromeres of *leishmania major*: revealing the hidden pieces. *EMBO Reports* **18**:1968–1977. DOI: https://doi.org/10.15252/embr.201744216, PMID: 28935715

Hahne F, Ivanek R. 2016. Visualizing genomic data using gviz and bioconductor. *Methods in Molecular Biology* **1418**:335–351. DOI: https://doi.org/10.1007/978-1-4939-3578-9_16, PMID: 27008022

Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, Thermes C. 2007. Human gene organization driven by the coordination of replication and transcription. *Genome Research* **17**:1278–1285. DOI: https://doi.org/10.1101/gr.6533407, PMID: 17675363

Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM, Warren WC, Berriman M, Sacks DL, Cotton JA, Grigg ME. 2017. Gene expression in *Leishmania* is regulated predominantly by gene dosage. *mBio* **8**:e01393-17. DOI: https://doi.org/10.1128/mBio.01393-17, PMID: 28900023

Ivanova D, Taylor T, Smith SL, Dimude JU, Upton AL, Mehrjouy MM, Skovgaard O, Sherratt DJ, Retkute R, Rudolph CJ. 2015. Shaping the landscape of the *Escherichia coli* chromosome: replication-transcription

encounters in cells with an ectopic replication origin. *Nucleic Acids Research* **43**:7865–7877. DOI: https://doi.org/10.1093/nar/gkv704, PMID: 26160884

Ivanova T, Maier M, Missarova A, Ziegler-Birling C, Dam M, Gomar-Alba M, Carey LB, Mendoza M. 2020. Budding yeast complete DNA synthesis after chromosome segregation begins. *Nature Communications* **11**. DOI: https://doi.org/10.1038/s41467-020-16100-3, PMID: 32385287

Jackson AP, Quail MA, Berriman M. 2008. Insights into the genome sequence of a free-living kinetoplastid: bodo saltans (Kinetoplastida: euglenozoa). *BMC Genomics* **9**:594. DOI: https://doi.org/10.1186/1471-2164-9-594, PMID: 19068121

Keeling PJ, Burki F. 2019. Progress towards the tree of eukaryotes. *Current Biology* **29**:R808–R817. DOI: https://doi.org/10.1016/j.cub.2019.07.031, PMID: 31430481

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. 2010. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. *PLOS Pathogens* **6**:e1001090. DOI: https://doi.org/10.1371/journal.ppat.1001090, PMID: 20838601

Laffitte MN, Leprohon P, Papadopoulou B, Ouellette M. 2016a. Plasticity of the *leishmania* genome leading to gene copy number variations and drug resistance. *F1000Research* **5**:2350. DOI: https://doi.org/10.12688/f1000research.9218.1, PMID: 27703673

Laffitte MC, Leprohon P, Hainse M, Légaré D, Masson JY, Ouellette M. 2016b. Chromosomal translocations in the parasite leishmania by a MRE11/RAD50-Independent Microhomology-Mediated end joining mechanism. *PLOS Genetics* **12**:e1006117. DOI: https://doi.org/10.1371/journal.pgen.1006117, PMID: 27314941

LeBowitz JH, Smith HQ, Rusche L, Beverley SM. 1993. Coupling of poly(A) site selection and trans-splicing in leishmania. *Genes & Development* **7**:996–1007. DOI: https://doi.org/10.1101/gad.7.6.996, PMID: 8504937

Leonard AC, Méchali M. 2013. DNA replication origins. *Cold Spring Harbor Perspectives in Biology* **5**:a010116. DOI: https://doi.org/10.1101/cshperspect.a010116, PMID: 23838439

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. DOI: https://doi.org/10.1093/bioinformatics/btp352, PMID: 19505943

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: https://doi.org/10.1093/bioinformatics/btp324, PMID: 19451168

Lombraña R, Almeida R, Revuelta I, Madeira S, Herranz G, Saiz N, Bastolla U, Gómez M. 2013. High-resolution analysis of DNA synthesis start sites and nucleosome architecture at efficient mammalian replication origins. *The EMBO Journal* **32**:2631–2644. DOI: https://doi.org/10.1038/emboj.2013.195, PMID: 23995398

Lombraña R, Álvarez A, Fernández-Justel JM, Almeida R, Poza-Carrión C, Gomes F, Calzada A, Requena JM, Gómez M. 2016. Transcriptionally driven DNA replication program of the human parasite leishmania major. *Cell Reports* **16**:1774–1786. DOI: https://doi.org/10.1016/j.celrep.2016.07.007, PMID: 27477279

MacAlpine HK, Gordân R, Powell SK, Hartemink AJ, MacAlpine DM. 2010. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Research* **20**:201–211. DOI: https://doi.org/10.1101/gr.097873.109, PMID: 19996087

Marques CA, Dickens NJ, Paape D, Campbell SJ, McCulloch R. 2015. Genome-wide mapping reveals single-origin chromosome replication in Leishmania, a eukaryotic microbe. *Genome Biology* **16**:230. DOI: https://doi.org/10.1186/s13059-015-0788-9, PMID: 26481451

Marques CA, Tiengwe C, Lemgruber L, Damasceno JD, Scott A, Paape D, Marcello L, McCulloch R. 2016. Diverged composition and regulation of the *Trypanosoma brucei* origin recognition complex that mediates DNA replication initiation. *Nucleic Acids Research* **44**:4763–4784. DOI: https://doi.org/10.1093/nar/gkw147, PMID: 26951375

Marques CA, McCulloch R. 2018. Conservation and variation in strategies for DNA replication of kinetoplastid nuclear genomes. *Current Genomics* **19**:98–109. DOI: https://doi.org/10.2174/1389202918666170815144627

Marsico G, Chambers VS, Sahakyan AB, McCauley P, Boutell JM, Antonio MD, Balasubramanian S. 2019. Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Research* **47**:3862–3874. DOI: https://doi.org/10.1093/nar/gkz179, PMID: 30892612

Martínez-Calvillo S, Yan S, Nguyen D, Fox M, Stuart K, Myler PJ. 2003. Transcription of leishmania major friedlin chromosome 1 initiates in both directions within a single region. *Molecular Cell* **11**:1291–1299. DOI: https://doi.org/10.1016/S1097-2765(03)00143-6, PMID: 12769852

Martínez-Calvillo S, Nguyen D, Stuart K, Myler PJ. 2004. Transcription initiation and termination on leishmania major chromosome 3. *Eukaryotic Cell* **3**:506–517. DOI: https://doi.org/10.1128/EC.3.2.506-517.2004, PMID: 15075279

McDonagh PD, Myler PJ, Stuart K. 2000. The unusual gene organization of leishmania major chromosome 1 may reflect novel transcription processes. *Nucleic Acids Research* **28**:2800–2803. DOI: https://doi.org/10.1093/nar/28.14.2800, PMID: 10908338

McIntosh D, Blow JJ. 2012. Dormant origins, the licensing checkpoint, and the response to replicative stresses. *Cold Spring Harbor Perspectives in Biology* **4**:a012955. DOI: https://doi.org/10.1101/cshperspect.a012955, PMID: 22904560

McMurray CT. 2010. Mechanisms of trinucleotide repeat instability during human development. *Nature Reviews Genetics* **11**:786–799. DOI: https://doi.org/10.1038/nrg2828, PMID: 20953213

Mesner LD, Valsakumar V, Karnani N, Dutta A, Hamlin JL, Bekiranov S. 2011. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome Research* **21**:377–389. DOI: https://doi.org/10.1101/gr.111328.110, PMID: 21173031

Minocherhomji S, Ying S, Bjerregaard VA, Bursomanno S, Aleliunaite A, Wu W, Mankouri HW, Shen H, Liu Y, Hickson ID. 2015. Replication stress activates DNA repair synthesis in mitosis. *Nature* **528**:286–290. DOI: https://doi.org/10.1038/nature16139, PMID: 26633632

Moreno A, Carrington JT, Albergante L, Al Mamun M, Haagensen EJ, Komseli ES, Gorgoulis VG, Newman TJ, Blow JJ. 2016. Unreplicated DNA remaining from unperturbed S phases passes through mitosis for resolution in daughter cells. *PNAS* **113**:E5757–E5764. DOI: https://doi.org/10.1073/pnas.1603252113, PMID: 27516545

Müller CA, Hawkins M, Retkute R, Malla S, Wilson R, Blythe MJ, Nakato R, Komata M, Shirahige K, de Moura AP, Nieduszynski CA. 2014. The dynamics of genome replication using deep sequencing. *Nucleic Acids Research* **42**:e3. DOI: https://doi.org/10.1093/nar/gkt878, PMID: 24089142

Müller LSM, Cosentino RO, Förstner KU, Guizetti J, Wedel C, Kaplan N, Janzen CJ, Arampatzi P, Vogel J, Steinbiss S, Otto TD, Saliba AE, Sebra RP, Siegel TN. 2018. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**:121–125. DOI: https://doi.org/10.1038/s41586-018-0619-8, PMID: 30333624

Müller CA, Nieduszynski CA. 2012. Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Research* **22**:1953–1962. DOI: https://doi.org/10.1101/gr.139477.112, PMID: 22767388

Özer Ö, Bhowmick R, Liu Y, Hickson ID. 2018. Human cancer cells utilize mitotic DNA synthesis to resist replication stress at telomeres regardless of their telomere maintenance mechanism. *Oncotarget* **9**:15836–15846. DOI: https://doi.org/10.18632/oncotarget.24745

Özer Ö, Hickson ID. 2018. Pathways for maintenance of telomeres and common fragile sites during DNA replication stress. *Open Biology* **8**:180018. DOI: https://doi.org/10.1098/rsob.180018, PMID: 29695617

Prieto Barja P, Pescher P, Bussotti G, Dumetz F, Imamura H, Kedra D, Domagalska M, Chaumeau V, Himmelbauer H, Pages M, Sterkers Y, Dujardin JC, Notredame C, Späth GF. 2017. Haplotype selection as an adaptive mechanism in the protozoan pathogen leishmania donovani. *Nature Ecology & Evolution* **1**:1961–1969. DOI: https://doi.org/10.1038/s41559-017-0361-x, PMID: 29109466

Prorok P, Artufel M, Aze A, Coulombe P, Peiffer I, Lacroix L, Guédin A, Mergny JL, Damaschke J, Schepers A, Cayrou C, Teulade-Fichou MP, Ballester B, Méchali M. 2019. Involvement of G-quadruplex regions in mammalian replication origin activity. *Nature Communications* **10**:3274. DOI: https://doi.org/10.1038/s41467-019-11104-0, PMID: 31332171

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**:W160–W165. DOI: https://doi.org/10.1093/nar/gkw257, PMID: 27079975

Reynolds D, Cliffe L, Förstner KU, Hon CC, Siegel TN, Sabatini R. 2014. Regulation of transcription termination by glucosylated hydroxymethyluracil, base J, in leishmania major and *Trypanosoma brucei*. *Nucleic Acids Research* **42**:9717–9729. DOI: https://doi.org/10.1093/nar/gku714, PMID: 25104019

Richard GF, Kerrest A, Dujon B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews* **72**:686–727. DOI: https://doi.org/10.1128/MMBR.00011-08, PMID: 19052325

Rocha-Granados MC, Klingbeil MM. 2016. Leishmania DNA replication timing: a stochastic event? *Trends in Parasitology* **32**:755–757. DOI: https://doi.org/10.1016/j.pt.2016.05.011, PMID: 27255527

Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin JC, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of leishmania. *Genome Research* **21**:2129–2142. DOI: https://doi.org/10.1101/gr.122945.111, PMID: 22038252

Song W, Dominska M, Greenwell PW, Petes TD. 2014. Genome-wide high-resolution mapping of chromosome fragile sites in *Saccharomyces cerevisiae*. *PNAS* **111**:E2210–E2218. DOI: https://doi.org/10.1073/pnas.1406847111, PMID: 24799712

Stanojcic S, Sollelis L, Kuk N, Crobu L, Balard Y, Schwob E, Bastien P, Pagès M, Sterkers Y. 2016. Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes leishmania and *Trypanosoma brucei* versus mammalian cells. *Scientific Reports* **6**:23142. DOI: https://doi.org/10.1038/srep23142, PMID: 26976742

Sterkers Y, Lachaud L, Crobu L, Bastien P, Pagès M. 2011. FISH analysis reveals aneuploidy and continual generation of chromosomal mosaicism in leishmania major. *Cellular Microbiology* **13**:274–283. DOI: https://doi.org/10.1111/j.1462-5822.2010.01534.x, PMID: 20964798

Theis JF, Irene C, Dershowitz A, Brost RL, Tobin ML, di Sanzo FM, Wang JY, Boone C, Newlon CS. 2010. The DNA damage response pathway contributes to the stability of chromosome III derivatives lacking efficient replicators. *PLOS Genetics* **6**:e1001227. DOI: https://doi.org/10.1371/journal.pgen.1001227, PMID: 21151954

Thomas S, Green A, Sturm NR, Campbell DA, Myler PJ. 2009. Histone acetylations mark origins of polycistronic transcription in leishmania major. *BMC Genomics* **10**:152. DOI: https://doi.org/10.1186/1471-2164-10-152, PMID: 19356248

Tiengwe C, Marcello L, Farr H, Dickens N, Kelly S, Swiderski M, Vaughan D, Gull K, Barry JD, Bell SD, McCulloch R. 2012a. Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Reports* **2**:185–197. DOI: https://doi.org/10.1016/j.celrep.2012.06.007, PMID: 22840408

Tiengwe C, Marcello L, Farr H, Gadelha C, Burchmore R, Barry JD, Bell SD, McCulloch R. 2012b. Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLOS ONE* **7**:e32674. DOI: https://doi.org/10.1371/journal.pone.0032674, PMID: 22412905

Tiengwe C, Marques CA, McCulloch R. 2014. Nuclear DNA replication initiation in kinetoplastid parasites: new insights into an ancient process. *Trends in Parasitology* **30**:27–36. DOI: https://doi.org/10.1016/j.pt.2013.10.009, PMID: 24287149

Torres-Guerrero E, Quintanilla-Cedillo MR, Ruiz-Esmenjaud J, Arenas R. 2017. Leishmaniasis: a review. *F1000Research* **6**:750. DOI: https://doi.org/10.12688/f1000research.11120.1, PMID: 28649370

Torres-Rosell J, De Piccoli G, Cordon-Preciado V, Farmer S, Jarmuz A, Machin F, Pasero P, Lisby M, Haber JE, Aragón L. 2007. Anaphase onset before complete DNA replication with intact checkpoint responses. *Science* **315**:1411–1415. DOI: https://doi.org/10.1126/science.1134025, PMID: 17347440

Ubeda JM, Légaré D, Raymond F, Ouameur AA, Boisvert S, Rigault P, Corbeil J, Tremblay MJ, Olivier M, Papadopoulou B, Ouellette M. 2008. Modulation of gene expression in drug resistant leishmania is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome Biology* **9**:R115. DOI: https://doi.org/10.1186/gb-2008-9-7-r115, PMID: 18638379

Ubeda JM, Raymond F, Mukherjee A, Plourde M, Gingras H, Roy G, Lapointe A, Leprohon P, Papadopoulou B, Corbeil J, Ouellette M. 2014. Genome-wide stochastic adaptive DNA amplification at direct and inverted DNA repeats in the parasite leishmania. *PLOS Biology* **12**:e1001868. DOI: https://doi.org/10.1371/journal.pbio.1001868, PMID: 24844805

van Luenen HG, Farris C, Jan S, Genest PA, Tripathi P, Velds A, Kerkhoven RM, Nieuwland M, Haydock A, Ramasamy G, Vainio S, Heidebrecht T, Perrakis A, Pagie L, van Steensel B, Myler PJ, Borst P. 2012. Glucosylated hydroxymethyluracil, DNA base J, prevents transcriptional readthrough in leishmania. *Cell* **150**:909–921. DOI: https://doi.org/10.1016/j.cell.2012.07.030, PMID: 22939620

Wedel C, Forstner KU, Derr R, Siegel TN. 2017. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in african trypanosomes. *The EMBO Journal* **36**:2581–2594.

Wickstead B, Ersfeld K, Gull K. 2003. Repetitive elements in genomes of parasitic protozoa. *Microbiology and Molecular Biology Reviews* **67**:360–375. DOI: https://doi.org/10.1128/MMBR.67.3.360-375.2003, PMID: 12966140

Xia X. 2018. Genomic Features: Content Sensors, Nucleotide Skew Plot, Strand Asymmetry, and DNA Methylation. In: *Bioinformatics and the Cell: Modern Computational Approaches in Genomics, Proteomics and Transcriptomics*. Springer International Publishing. p. 255–268. DOI: https://doi.org/10.1007/978-0-387-71337-3