



Marques, C. A. and McCulloch, R. (2018) Conservation and variation in strategies for DNA replication of kinetoplastid nuclear genomes. *Current Genomics*, 19(2), pp. 98-109.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/146253/>

Deposited on: 25 August 2017

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Mini Review

Conservation and variation in strategies for DNA replication of kinetoplastid nuclear genomes

Catarina A. Marques¹ and Richard McCulloch²

¹ Division of Biological Chemistry and Drug Discovery, School of Life Sciences, Dow Street, University of Dundee, Dundee, DD1 5EH, UK

² The Wellcome Centre for Molecular Parasitology, Institute of Infection, Immunity and Inflammation, Sir Graeme Davis Building, 120 University Place, University of Glasgow, Glasgow, G12 8TA, UK

Correspondence: c.marques@dundee.ac.uk and/or Richard.mcculloch@glasgow.ac.uk

Abstract

Understanding how the nuclear genome of kinetoplastid parasites is replicated received experimental stimulus from sequencing of the *Leishmania major*, *Trypanosoma brucei* and *Trypanosoma cruzi* genomes around 10 years ago. Gene annotations suggested key players in DNA replication initiation could not be found in these organisms, despite considerable conservation amongst characterised eukaryotes. Initial studies that indicated trypanosomatids might possess an archaeal-like origin recognition complex (ORC), composed of only a single factor termed ORC1/CDC6, have been supplanted by the more recent identification of an ORC in *T. brucei*. However, the constituent subunits of *T. brucei* ORC are highly diverged relative to other eukaryotic ORCs and the activity of the complex appears subject to novel, positive regulation. The availability of whole genome sequences has also allowed the deployment of genome-wide strategies to map DNA replication dynamics, to date in *T. brucei* and *Leishmania*. ORC1/CDC6 binding and function in *T. brucei* displays pronounced overlap with the unconventional organisation of gene expression in the genome. Moreover, mapping of sites of replication initiation suggests pronounced differences in replication dynamics in *Leishmania* relative to *T. brucei*. Here we discuss what implications these emerging data may have for parasite and eukaryotic biology of DNA replication.

Keywords: DNA replication, origin replication complex (ORC), origins of replication, *Trypanosoma brucei*, *Leishmania*.

Every cell cycle, the genome must be completely and accurately duplicated before being transmitted to daughter cells. Incomplete or inaccurate replication risks genome instability, a hallmark of cancer[1], and it is vital that the entire genome is normally strictly replicated just once during the S phase of the cell cycle. Hence, DNA replication is a complex, tightly regulated process[2-5] that has been extensively studied in model eukaryotes, in bacteria and, most recently, in archaea[6].

DNA replication is initiated at origins of replication, which are found as a single sequence-defined site in all bacterial genomes. In archaea, origins are also defined by conserved sequences, but their organisation differs from bacteria in that some archaeal genomes possess a few origins. The large genomes of eukaryotes differ yet further, in that each linear chromosome contains many (sometimes hundreds to thousands) of potential origins of replication, which are not all activated in every cell cycle and are rarely defined by sequence elements but, instead, are associated with poorly defined features like chromatin structure and status[7]. Despite these differences, origins of replication across the three domains of life are recognised by initiator factors belonging to the AAA+ superfamily of NTPases[8]: the single DnaA[9] and Orc1/Cdc6[10] factors in bacteria and archaea, respectively, and the six-subunit (Orc1-Orc6) origin recognition complex (ORC) in eukaryotes[11]. Binding of these factors demarcates the origins of replication and leads to the recruitment and activation of the replicative helicase. Subsequently, the remaining constituents of the replication machinery are in turn recruited, thus starting, and driving, DNA synthesis[3].

Much of the diversity of the eukaryotic domain resides in microbes, many of which display dramatic deviation in core biology from model, primarily opisthokont, organisms. Such diversity extends to nuclear biology, including variations in mitosis[12], gene expression[13], and genome organisation and stability[14]. Though most of our understanding of the machinery and execution of DNA replication in eukaryotes has focused on the opisthokonts, a number of recent studies have begun to explore nuclear replication in kinetoplastids, suggesting that here too there may be surprising variation [15,16].

The kinetoplastid origin recognition complex

Sequencing of the TriTryp genomes (*Trypanosoma brucei*, *Leishmania major* and *Trypanosoma cruzi*)[17-19] in 2005 provided a spur for the field of kinetoplastid biology[20], including nuclear DNA replication. Prior to genome sequencing, attempts to examine replication dynamics could not consider the whole genome (see below) and little work had examined the replication machinery. Surprisingly, sequence similarity searches were unable to identify various core DNA replication proteins, in particular those involved in the initiation steps of the process: while orthologues of most factors involved in replication fork assembly and DNA synthesis were readily identified, only one ORC-related factor (a putative orthologue of Orc1) could be found[19] (Figure 1, A and B). Like in other eukaryotes, this putative Orc1 subunit also shared homology with Cdc6 (an essential factor that interacts with ORC to allow loading of the replicative helicase onto origins of replication)[11]. The lack of identifiable orthologues of the other ORC subunits, Cdc6, and a

further helicase loader (Cdt1), suggested that initiation of DNA replication in these parasites might be mechanistically more similar to archaeal organisms than model eukaryotes[19]. In line with this hypothesis, and supported by experimental evidence, Orc1 in *T. cruzi* and *T. brucei* was re-named ORC1/CDC6[21] (Figure 1, A), reflecting the combined ORC and Cdc6 functions that reside in the single protein that binds each origin in archaea, while other studies analysed Orc1 in various *Leishmania* species[22,23]. A 2011 study[24], however, questioned whether the kinetoplastid initiator is just a single protein. While focusing on the functional analysis of components of the replicative helicase in *T. brucei* – the heterohexameric minichromosome maintenance complex (MCM2-7), GINS complex and CDC45 – the authors identified a second, highly divergent orthologue of Orc1[24] (Figure 1, A and B). Labelled as TbORC1B, this putative initiator was shown to interact with TbORC1/CDC6 and TbMCM3, but its role in DNA replication was not explored. Nonetheless, further evidence that *T. brucei* might not have a single protein initiator factor, but an ORC-like complex, was provided by a study the following year, in which three more TbORC1/CDC6-interacting factors were identified[25] (Figure 1, A and B): a highly divergent Orc4-like subunit (TbORC4), and two apparently kinetoplastid-specific factors, Tb7980 and Tb3120, with very limited homology with Orc proteins.

More recent experiments, thus far limited to *T. brucei*, provide strong evidence of an ORC-like complex. RNA interference (RNAi) shows that loss of TbORC1/CDC6, TbORC1B, TbORC4 or Tb3120 impedes DNA replication[21,25-27] and leads to similar growth and cell cycle defects[27]. Though loss of Tb7980 also results in the same proliferation defects[25], clear evidence of a role in DNA replication remains to be assessed. TbORC1/CDC6 depletion additionally leads to the expression of silent variant surface proteins (VSGs) in bloodstream form cells[26] and in procyclic cells[28], suggesting that, similar to other eukaryotic Orc1 subunits, and despite lacking a bromo adjacent homology (BAH) domain (Figure 1, B), TbORC1/CDC6 has a role in gene silencing. However, whether these activities are executed within the putative ORC is unknown. TbORC1/CDC6-binding sites have also been mapped genome-wide[28]. Within the core of the chromosomes TbORC1/CDC6 binds mainly at the boundaries of the multi-gene transcription units, where it overlaps with mapped origins of replication (see below for fuller explanation), supporting a role in replication initiation. In addition, ~60% of TbORC1/CDC6 binding sites locate to the chromosomes' subtelomeric regions[28]. Whether this dense binding in the subtelomeres is functionally related to binding directly to the telomeres[26], and whether this is related to TbORC1/CDC6's silencing function(s), is not clear. Finally, gel filtration analysis suggests TbORC1/CDC6 and TbORC4 are present, most likely together, in a high molecular complex (~1011 to 530 kDa) that also seems to include the helicase subunit TbMCM3[27]. As TbMCM3 has been shown to interact with TbORC1/CDC6 and TbORC1B[24], and its orthologue in yeast mediates the recruitment of the MCM2-7 helicase to the DNA-bound ORC[29], the accumulation of data indicate the presence of an ORC in *T. brucei*, albeit one that defied simple sequence-based identification. As each putative ORC component is syntenically conserved in *T. cruzi* and *Leishmania*, there is a strong likelihood that the diverged composition of *T. brucei* ORC is a feature of all kinetoplastids. What is less clear is why kinetoplastid ORC might be diverged, and whether this extends to changes in function or regulation.

Most of our knowledge of the processes and molecular machineries involved in the eukaryotic cell cycle has been inferred from studies in a limited group of model organisms, including *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Xenopus laevis*, mouse and human, and extrapolated across the eukaryotic tree. However, because these organisms belong to a single supergroup of the eukaryotic domain, termed the Opisthokonta, it is less clear how well details are conserved in the five other eukaryotic supergroups, and thus, the extent of diversity in replication processes and molecular machineries across eukaryotes domain is poorly explored[30,31]. From studies in model eukaryotes, Orcs1-5 each appear to be structurally similar, with each subunit containing a central or N-terminal AAA+ ATPase domain and at least one C-terminal winged-helix (WH) DNA-binding domain[6,11] (Figure 1, B, insert box). Within the AAA+ ATPase domain, conserved Walker A and B motifs (fundamental components of the ATP-binding site and essential for ATPase activity) and a signature arginine finger (involved in ATP hydrolysis of the adjacent AAA+ ATPase subunit in the protein complex) are needed to modulate ORC activity[32,33]. In contrast, Orc6 appears to have evolved independently, and thus does not possess any of these Orc-characteristic domains[11]. Recently, the ever-growing availability of sequenced genomes from a wide range of organisms, spanning all eukaryotic supergroups, has allowed sequence-based analysis of ORC-orthologues across the eukaryotic domain[30]. Reflecting the putative kinetoplastid ORC composition, not all six subunits of ORC were found across the analysed genomes, suggesting there is variability in the composition of ORC between eukaryotes. However, whether this reflects the existence of simpler ORCs, high divergence at the protein sequence level, less functional

conservation of some ORC subunits, or the existence unrelated factors performing the role of Orc-subunits in certain organisms is unknown, and requires experimental investigation.

Recently, the structure of *D. melanogaster* ORC was solved[34]. Prior to this elegant study, only archaeal ORC1/CDC6 protein structures, and a fragment of the human Orc6 subunit[35], were solved and available on the RCSB Protein Data Bank. This meant that protein structure modelling software could only model eukaryotic ORC subunits structure based on the archaeal ORC1/CDC6 structures. While this had the potential to model Orc4 and Orc5 sequences, as these subunits possess a conserved AAA+ ATPase domain[6,11], modelling of Orc2 and Orc3 was particularly problematic, as these subunits have highly divergent AAA+ folds containing non-canonical Walker A and B motifs that are, however, conserved amongst different organisms' Orc2 and Orc3 subunits[36,37]. Because Orc6 is the least conserved of the Orc subunits in both sequence and function[35,38-40], it was unlikely that other organisms' Orc6 subunits could be modelled based on the available human structure. The availability of structures of all *D. melanogaster* ORC subunits [34], and very recently, of the human counterparts[41], has therefore opened the door for the modelling of Orc2-Orc5 related sequences. From this approach *T. brucei* Tb7980 and Tb3120 emerge with potential structural similarity to *D. melanogaster* Orc5 and Orc2 subunits, respectively[27]. Nevertheless, whether these act as functional TbORC5 and TbORC2 orthologues is unknown, and requires further study. Very recently, we have further identified another TbORC1/CDC6-interacting factor, which is a hypothetical zinc-finger protein [42,43] that appears to possess feeble homology with *D. melanogaster* Orc3 (unpublished). Again, however, functional analysis is needed to verify such potential orthology. Such functional studies may be of value, since little work has explored if individual ORC subunits provide conserved, discrete functions across eukaryotes. To our knowledge, there is currently no hint for the presence of an Orc6 subunit in kinetoplastids. Nonetheless, if the conformation of kinetoplastid ORC follows that of *D. melanogaster*, where the subunits are arranged in a Orc1-Orc4-Orc5-Orc3-Orc2 ring, with Orc6 interacting with Orc3[34], then the greater conservation of Orc1 and Orc4 would suggest these components represent a functionally constrained 'core', with greater functional flexibility amongst the other subunits, an arrangement consistent with eukaryote-wide ORC sequence comparisons[30].

Structural analysis is needed to test the above prediction of kinetoplastid ORC architecture, but sequence and experimental analyses provide further clues about subunit activity. Reflecting its early identification[19], ORC1/CDC6 is the subunit showing the greatest conservation with other eukaryotic Orc proteins. Analyses of *L. major*, *T. brucei* and *T. cruzi* ORC1/CDC6 protein sequences[21,22] were able to identify a putative N-terminal AAA+ ATPase domain, with clearly identifiable Walker A, Walker B and arginine finger motifs, as well as a putative WH domain at the C-terminus[21,22] (Figure 1, B). Functionality of these domains has been confirmed in *T. brucei* and *T. cruzi* ORC1/CDC6 proteins[21], while a nuclear localisation signal (NLS) at the N-terminus has been experimentally confirmed in the *L. donovani* counterpart[23]. Nonetheless, despite their sequence homology with Orc1 subunits, the kinetoplastid ORC1/CDC6 proteins are atypical, as none possesses an identifiable BAH domain at their N-terminus[21,22], a feature thought to be universal across Orc1 subunits and shown to be involved in origin recognition and transcriptional silencing[44-49]. Despite being larger than TbORC1/CDC6, TbORC1B (and its *L. major* and *T. cruzi* counterparts) also lacks a BAH domain[24]. An N-terminal AAA+ ATPase domain is seen in TbORC1B, with conserved but widely spaced Walker A and B motifs[24] (Figure 1, B), but no arginine finger signature is found[24]. Whether this explains experimental data suggesting TbORC1B is devoid of ATPase activity[24] is unclear. Finally, sequence alignments with various eukaryotic Orc1 and Cdc6 proteins allows a weak prediction of a WH domain at the C-terminus of TbORC1B[43], but whether it binds to DNA has not been examined. Taken together, it remains unclear if TbORC1/CDC6 and TbORC1B are orthologues of Orc1 and Cdc6 and, since functional data suggest TbORC1B is highly unorthodox (see below), it remains premature to assign such orthology. TbORC4 also seems to possess an AAA+ ATPase domain; although this includes an Orc4 conserved arginine finger motif, both Walker A and B motifs appear degenerate[25] (Figure 1, B), suggesting that TbORC4 is, most likely, devoid of ATPase activity. Nevertheless, Orc4 subunits in other eukaryotes do not possess ATPase activity themselves, but their conserved arginine finger appears to be necessary for Orc1 ATPase hydrolysis, by supplying the arginine finger to Orc1's ATP-binding site[34,50-52]. Analogously, it is thus possible that TbORC4 might stimulate TbORC1/CDC6 ATPase activity. It has not been possible to assess whether TbORC4 possesses a WH domain[25], and its binding to DNA has not been tested, though *S. pombe* Orc4 clearly provides such a function through AT-hook motifs[53]. Sequence analysis of Tb7980 suggests the presence of an AAA+ ATPase domain with relatively conserved Walker A and Walker B motifs, but no arginine finger or WH signatures [15,25,27] (Figure 1, B). Like TbORC4, Tb7980 ATPase and DNA binding activities have not been assessed. Structural modelling of Tb3120 hints at homology with *D. melanogaster* Orc2 [27], but this is

mainly restricted to their C-termini (where the DNA-binding WH domain of *D. melanogaster* Orc2 is present)[27] (Figure 1, B). Further alignment of Tb3120 with multiple Orc2 protein sequences revealed that Tb3120 contains non-canonical Walker A and B motifs with characteristic signatures of Orc2 proteins. These are, however, separated by a large insertion (also observed in *L. major* and *T. cruzi* Tb3120 orthologues), suggesting that Tb3120 does not possess an intact AAA+ ATPase domain and is, therefore, unlikely to possess ATPase activity. In addition, Tb3120 and its orthologues in *L. major* and *T. cruzi* are considerably larger than model eukaryotes' Orc2 subunits, and appear to have evolved a N-terminal extension with no detectable features or predicted function[27]. Further experimental analysis of Tb3120 will be essential to test whether this protein is the kinetoplastid equivalent of Orc2, and to better understand the functional implications of the sequence features highlighted above.

TbORC1/CDC6[21], TbORC4, Tb3120 and Tb7980 all localise to the *T. brucei* nucleus throughout the cell cycle [27]. In striking contrast, TbORC1B localisation and expression is cell cycle dependent, with TbORC1B being detected only in the nucleus of cells in late G1 to late S or G2 phase [27]. Together with the rapid impairment of DNA replication and cell cycle progression upon depletion by RNAi[27], the available data suggest that TbORC1B might act as a positive regulator of DNA replication in *T. brucei*, rather than being a static member of the putative ORC-like complex. However, TbORC1B expression and localisation dynamics do not resemble any regulatory DNA replication factor described in other eukaryotes, including Cdc6. Indeed, TbORC1B's lack of ATPase activity renders it an unlikely candidate to provide Cdc6 function, which acts by ATPase-dependent remodelling of ORC[51]. Thus, TbORC1B may represent a truly kinetoplastid-specific adaptation and it remains possible that *T. brucei* ORC-MCM interactions are archaeal-like in lacking Cdc6/Cdt1 mediation[54], consistent with TbORC1/CDC6 having been shown to be able to complement *S. cerevisiae* *cdc6* temperature-sensitive mutants[21].

Whether or not the emerging data on the putative ORC in kinetoplastids can provide potential targets for drug development against these parasites remains to be seen [15,16,55], but unexpected levels of divergence in a core component of genome maintenance is emerging. It is intriguing to note that characterisation of the kinetoplastid kinetochore complex has also required protein interaction-based recovery of novel diverged components[56,57], with an emerging picture of greatest variation in the subunits that interact with the genome[58]. Perhaps these parallel sets of studies hint that the unusual organisation of the kinetoplastid genome has necessitated widespread innovation in the protein factories that act in chromosome biology.

DNA replication dynamics in *Trypanosoma brucei* and *Leishmania*

In order to be completely and accurately duplicated within the S phase of the cell cycle, the large linear eukaryotic genomes are replicated from many origins of replication distributed throughout their multiple chromosomes. Increased numbers of origins necessitated the evolution of DNA replication controls to ensure that the genome is duplicated exactly once during the cell cycle, and that under-replication and/or re-replication of parts of the genome is prevented. To this end, the process of eukaryotic DNA replication is divided into two non-overlapping phases: origin licensing and origin firing.

Origin licensing takes place from late mitosis to the end of G1, and consists of the loading of the replicative helicase (MCM2-7 double hexamers), in an inactive state, to every potential origin of replication in the genome, which are already demarcated by the binding of ORC[2-5]. During S phase, only a subset of origins is fired, not simultaneously but according to a DNA replication-timing programme[59], by the recruitment of additional replication complexes to the potential origins, establishing bidirectional replication forks at sites distributed across the genome[2-5]. By licensing origins before the entry in S phase, already replicated origins cannot be re-licensed during S or G2 phases of the cell cycle, thus avoiding re-replication and the risk of genomic instability[2-5]. However, this implies that before entering S phase, the number of licensed origins must be sufficient to allow the complete replication of the genome, including in the event of replication stress from internal and external sources[1]. Therefore, the licensing of more origins than are needed to fire during S phase provides dormant origins that can be activated as a failsafe to prevent under-replication of regions of the genome and, thus, ensure genomic integrity[60]. If not activated, the excess licensed origins are passively replicated by replication forks established at neighbouring origins. Origin usage in eukaryotes is, therefore, very flexible[4].

With the exception of *S. cerevisiae*[61], eukaryotic origins of replication are not defined by specific DNA sequences. Instead, ORC-binding to origins and their later activation during S phase seems to depend on the combination of an assortment of less defined markers[2,4,7,62,63], such as: sequence features, including AT-rich regions[53,64,65] and

CpG islands[66-71]; DNA topology, like G-quadruplexes[71-76] and negatively supercoiled DNA[77]; chromatin structure and status, such as nucleosome-free chromatin[78-80] and DNase-sensitive sites[76]; histone modifications[76,81-84]; transcription[85,86], with regions containing transcriptionally active genes being replicated earlier[87,88], and some origins found at promoters (including RNA Polymerase II-binding sites)[65,68,78]; and the positioning of origin DNA in the nuclear space[89-92], with early fired origins localising to the nuclear interior, and late activated ones locating to the nuclear periphery. Due to their elusive characteristics, identification and mapping of origins of replication in any eukaryotic genome has not been straightforward. However, the development of high throughput sequencing methodologies[93] has increased the availability of sequenced genomes and allowed the development of techniques to map and correlate processes genome-wide, including, amongst others, the chromatin landscape[94-97], gene expression[98-100], and DNA replication dynamics[63,101] – including origin localisation and usage. Recent studies in kinetoplastids provide an example of how these developments rapidly improved our understanding of DNA replication in these parasites. Prior to genome sequencing, origin mapping relied upon evaluating the ability of genome sequences to enhance episome stability[102] or maintain stability of fragmented chromosomes[103], or on analysis of small *T. brucei* minichromosomes[104]. The availability of near wholly compiled genomes has, to date, provided a genome-wide view of replication in *T. brucei* and two *Leishmania* species.

The *T. brucei*[17] and *L. major*[18] genomes (~50-66 Mbp, diploid) are organised in an unconventional architecture: unlike most other eukaryotes, the genomes of these closely related parasites are structured into ~200 multi-gene clusters, each of which is transcribed from its own single RNA Polymerase (Pol) II transcription start site[105-107]. Genome-wide multigenic transcription means most of the genome is traversed by RNA Pol II and gene expression controls are mainly post-transcriptional[108,109]. Moreover, the core genomes of these parasites are extensively syntenic[105], despite being organised in strikingly different numbers of chromosomes: 11 stably diploid megabase chromosomes in *T. brucei*[17], and 36 chromosomes of variable ploidy in *L. major*[18,110,111]. In *T. brucei* aneuploidy appears much less pervasive than in *Leishmania*, being limited to the subtelomeres of the megabase chromosomes and to mini- and intermediate chromosomes, all of which act as stores for *VSG* genes[112].

Genome-wide mapping of origins of replication was first performed in *T. brucei*[28] using deep sequencing marker frequency analysis (MFA-seq), also known as Sort-seq in yeast[113,114]. Briefly, starting from an unsynchronised population, cells in S and G2 phase are isolated based on their DNA content by fluorescence activated cell sorting (FACS), and their DNA is then purified and sequenced. The resulting reads are mapped to the reference genome, allowing the ratio between S (replicating) and G2 (non-replicating) phases to be calculated. Representing the S/G2 read depth ratios across the chromosomes (Figure 2) reveals ‘peaks’ that emanate from active origins, while “valleys” denote zones of replication termination[114], thus allowing assessment of origin location and strength genome-wide. In yeast, Sort-seq shows remarkably precise correlation with predicted sequence conserved origins[113] and with recent replication mapping by Okazaki fragment sequencing (OK-seq)[115]. MFA-seq in *T. brucei* insect stage form (procyclic forms) cells in early-mid S phase[28] revealed that these parasites comply with the general principles of the eukaryotic DNA replication dynamics model: multiple origins were mapped per chromosome; the number of origins per chromosome correlates with chromosome size; all origins overlap with a subset (~4.4%) of TbORC1/CDC6-binding sites; differences in MFA-peak height suggest origins are activated at different times during S phase, with the earliest activated origins co-localising with mapped centromeres (chromosomes 1 to 8[116]); and neither active origins nor TbORC1/CDC6-binding sites appear to be defined by specific sequences. Though MFA-seq cannot localise origins with great accuracy, all peaks centred on the so-called strand-switch regions (SSRs) that separate multi-gene transcription units and contain transcription start (divergent SSRs) or end (convergent SSRs) locations, or both (head-to-tail SSRs)[28]. In fact, evidence for DNA replication and transcription functionally intersecting could be seen after RNAi depletion of TbORC1/CDC6, which results in increased transcript levels at the SSRs, suggesting TbORC1/CDC6 might contribute to the outlining of transcription boundaries[28]. However, not all of the putative ORC-like components have been localised in the genome, the manner of ORC recruitment to DNA remains unclear, and MFA-seq could not map origin activity in the chromosome subtelomeres or in mini- and intermediate chromosomes.

Despite localisation of both origins and TbORC1/CDC6-binding sites to SSRs, no common sequence elements have to date been identified, suggesting *T. brucei* origin demarcation is dissimilar to yeast. Moreover, the number of detectable origins (42 MFA-seq peaks in the ~26 Mbp haploid genome) suggests an inter origin distance (IOD) of ~600 kbp, which is considerably greater than the predicted IOD of ~42 kbp in *S. cerevisiae* (based on ~280 Sort-seq peaks in a haploid genome of ~12 Mbp,)[114]. More recently, MFA-seq was used to map origins in early and late S phase *T.*

brucei cells, in two different strains (TREU927 and Lister 427), and in the two culturable life cycle stages of the parasite (procyclic and bloodstream forms)[43,117]. In all these settings, essentially the same MFA-seq peaks are observed in the core genome, suggesting that origin location and usage is relatively rigid, though what distinguishes an origin-active SSR from an origin-inactive SSR is unknown. However, one locus shows pronounced changes in replication usage dependent on transcription activity: the single telomeric bloodstream *VSG* expression site (BES) that is transcriptionally active in bloodstream form cells is replicated early, whereas the remaining ~15 silent BES replicate in late S phase; moreover, all *VSG* BES are late-replicating in procyclic cell forms, when transcription is suppressed at all the sites [117]. These data strengthen the link between *T. brucei* DNA replication and transcription, and suggest a potential exploitation of DNA replication to drive antigenic variation, a topic recently reviewed elsewhere[118]. The wide spacing of *T. brucei* origins and pronounced rigidity in replication dynamics across the chromosomes' core may result from the parasite's odd genome organisation in well-defined multi-gene transcription units, meaning there is little flexibility in sites of replication initiation, as origins sited within the transcription units could lead to catastrophic clashes between RNA Pol II and the replisome. Nevertheless, single molecule analysis of DNA replication in chromosome 1 suggests at least some SSRs might be activated after hydroxyurea-induced replicative stress, and some replication might initiate from undefined subtelomeric sites[119]. Whether there is a genome-wide flexibility in origin usage under replicative stress or other conditions requires further investigation.

Given that *T. brucei* and *Leishmania* share an unconventional genome architecture and possess high levels of gene synteny, it might be predicted that DNA replication dynamics are comparable in the two parasites. However, MFA-seq in *Leishmania* insect stage cells (promastigotes) confounds this expectation (Figure 2)[120]. MFA-seq has so far been performed in two species of *Leishmania*, *L. major* (old world) and *L. mexicana* (new world), whose genomes are, respectively, distributed into 36 and 34 chromosomes[121,122] of variable ploidy[110]. Strikingly, in both *Leishmania* species, and in both early and late S cells, only one origin could be identified by MFA-seq per chromosome[120]. If confirmed, origin singularity in *Leishmania* is unprecedented in eukaryotes studied to date, as this was thought to be exclusive to the smaller genomes of bacteria and some archaea. *L. mexicana* chromosomes 8 and 20 are each syntenic not to one, but to two *L. major* chromosomes (29 plus 8, and 36 plus 20, respectively)[110]. Despite this genome reorganisation, only one origin could be mapped to each *L. mexicana* chromosome, suggesting that a single origin per chromosome is retained in the face of pronounced chromosome rearrangements[120]. Additionally, all detected MFA-seq peaks have a similar height and width, suggesting that all origins are of similar strength/usage and that there is no replication timing programme in *Leishmania*[120]. Nonetheless, like in *T. brucei*, all detected origins overlap exclusively with SSRs and, remarkably, ~40% of the identified origins were conserved in location in the two genomes[120]. Though such conservation might indicate a commonality in kinetoplastid origins, no origin-defining sequence, motif or pattern could be identified[120]. Nonetheless, detailed analysis of SSRs with and without origin activity revealed that in *Leishmania*, but not in *T. brucei*, the distance between the two most proximal genes to the SSR was significantly larger in SSRs with origin activity than in non-origin active SSRs[120]. This difference even extended to the origin-active SSRs of *L. major* chromosomes 29 and 36 and their syntenic, but non-origin active SSRs in *L. mexicana* chromosomes 8 and 20, respectively[120]. These data support the idea that in each *Leishmania* chromosome a specific SSR is associated with pronounced origin activity, though what causes this effect (e.g. the accumulation of specific factors) has not been investigated. One possibility is that, as in *T. brucei*, the *Leishmania* origin-active SSRs overlap with centromeres, but to date centromeres in *Leishmania* have not been successfully mapped. Many of the *Leishmania* SSRs, which provide origins of equal strength, are conserved with SSRs that in *T. brucei* map as origins of variable strength [28,117], a change in function that also deserves further examination.

The MFA-seq prediction of a single origin in each *Leishmania* chromosome appears inadequate to explain complete genome replication. Estimations of a 3-4 hour S phase[123] and a ~2.5 kbp.min⁻¹ DNA replication rate[124] suggest that while a single origin might be enough to drive the complete replication of the parasite's smaller chromosomes (0.28–0.84 Mbp, ~66% of the genome), it is insufficient to fully replicate the larger ones (up to ~3.3 Mbp)[120]. How, then, *Leishmania* parasites completely replicate their genome is unclear. It is possible that less efficient or less frequently used origins might have escaped MFA-seq detection (detection threshold of ~25% of the activity of the mapped origins)[120]. However, because the detected origins in *Leishmania* and *T. brucei* localise exclusively to SSRs, one would predict that other origins would also localise to SSRs. Chromosome 32, of ~1.6 Mbp, for instance, has only four SSRs and, therefore, only four potential origins, but only one displays an MFA-seq peak. Origin activity at the other SSRs, if based on stochastic firing in the population, would therefore exceed the MFA-seq detection threshold. These considerations suggest that other events, perhaps not used in *T. brucei*, may promote complete *Leishmania*

genome replication (see below). Two more recent studies have investigated DNA replication in *Leishmania* using different approaches to MFA-seq and the conflicting results may provide clues to how genome replication occurs in these parasites[124,125].

One study, using the highly sensitive mapping methodology of short nascent leading strand purification coupled with sequencing (SNS-seq)[72] predicts ~5100 DNA replication initiation sites of replication across the *L. major* genome (Figure 2), with no preferential localisation at SSRs[125]. This prediction is >100 fold more than the origins mapped by MFA-seq and suggests one potential origin every other gene (~8300 total genes), meaning origins are found throughout the polycistronic transcription units. The huge number of SNS-seq predicted origins may suggest the technique's sensitivity identifies origins used by only a few cells in the population, perhaps indicating flexibility in origin usage across the cellular population, which contrasts with the rigidity in origin usage and timing in *T. brucei*[28,117]. SNS-seq has not to date been applied to *T. brucei*, but if similar numbers of origins of replication were detected, it would represent 14-fold greater origins than TbORC1/CDC6-binding sites (~360)[28]. DNA fibre stretching analysis, in two independent studies[124,125], provides evidence that MFA-seq and SNS-seq respectively underestimate and overestimate DNA replication initiation events in *Leishmania*. Both studies could detect molecules with >1 site of replication initiation, which can be extrapolated to IODs of ~60-300 kbp (with considerably greater variation than fibre analysis predicts in *T. brucei*)[119]. Such IODs are 10-30 fold greater than the potential IOD of ~6.5 kbp that SNS-seq predicts[125]. Moreover, further extrapolation from the fibre analysis predicts ~150-400 origins, equalling or exceeding the number of mapped SSRs. It is currently impossible to correlate the *Leishmania* fibre data with MFA-seq, SNS-seq or genomic landmarks as the labelling is unable to localise replication sites within or between chromosomes, or indeed exclude that episomes have been examined. In contrast, fibre analysis linked to fluorescence *in situ* hybridisation (FISH) in chromosome 1 of *T. brucei* provides good correspondence with MFA-seq[119].

Can the apparent discord in current *Leishmania* replication data be explained? The above considerations suggest replication dynamics differ dramatically between *T. brucei* and *Leishmania*, despite the common use of SSRs as origins (as revealed by MFA-seq; Figure 2). Though there are gaps in our understanding in both parasites (e.g. ORC1/CDC6-binding sites have not been mapped in *Leishmania*; SNS-seq has not been applied to *T. brucei*), it is attractive to speculate that differing replication strategies are found in the two parasites and that these might explain their differing genome organisations (larger, stable chromosome in *T. brucei* and smaller, frequently aneuploid chromosomes in *Leishmania*). At least two hypotheses for differing replication dynamics might be considered.

First, it is possible that MFA-seq hugely underestimates the number of origins in the *Leishmania* genome and the single peaks observed in each chromosome arise from dense clustering of multiple initiation sites, as revealed by SNS-seq [125,126]. This model suggests *Leishmania* adopts a replication strategy similar to metazoans [72,74,127], with domains of early and late firing clusters of origins. This may be plausible, as SNS-seq has been widely used in metazoans but shows limited overlap with other established origin-mapping methods, such as ORC chromatin immunoprecipitation (ChIP-seq) or OK-seq mapping[128], and has an apparent bias towards CpG islands and G-quadruplex motifs[63]. However, if this model is correct, it is not clear why only a single, early-replicating origin cluster (MFA-seq peak) is seen per chromosome in *Leishmania*, whereas multiple putative clusters (MFA-seq peaks) are seen in each *T. brucei* chromosome. In addition, the model does not readily explain why the origins cluster so precisely around an SSR in each chromosome. In *T. brucei*, we know that SSRs are bound by ORC1/CDC6 and therefore the MFA-seq peaks, like similar analyses in yeast [113,115], detect ORC-defined origins and not origin clusters [28]. In considering any such genome-wide mapping approaches it is important to take account of the very different genome sizes being examined in single-celled eukaryotes relative to metazoans [59]. Finally, if all SNS-seq peaks represent replication initiation at a conventional ORC-defined origin, then it is legitimate to ask how *Leishmania* (and perhaps *T. brucei*) copes with the likely pronounced impediment to efficient RNA Pol II progression through the multigene transcription units. Given all these questions, it is possible that DNA replication initiation in *Leishmania* does not rely on ORC, but on transcription-associated events, as postulated by Lombr a *et al.* (2016), who showed that 87-90% of SNS-seq peaks localise to gene *trans*-splicing regions, suggesting they correlate with genomic regions where transcription is predicted to decelerate or stall. In contrast, <4% of SNS-seq peaks map to convergent and head-to-tail SSRs, which correspond with ~81-85% of all transcription termination sites, and may be where ORC1/CDC6 localises (by analogy with *T. brucei*)[125].

A second hypothesis is that *Leishmania* uses a bi-modal DNA replication strategy. In this model, the MFA-seq peaks represent constitutive activation of a single ORC-bound origin, localised at an SSR, in each chromosome. These

initiation events are therefore comparable with the MFA-seq mapping in *T. brucei*, explaining origin location conservation [129]. The second arm of the bi-modal strategy, necessary for complete duplication of the genome, is stochastic, with DNA replication being randomly initiated at ORC-independent (thus “origin”-independent) sites throughout the chromosomes. What these putative origin-independent replication initiation events are is unclear, but two possibilities can be envisaged, both of which might explain the SNS-seq data. The first possibility was suggested by Lombrana *et al.* (2016), with the SNS-seq peaks representing sites where the MCM helicase accumulates and recruits the replisome without ORC [130]. An alternative, and more radical explanation, is that stochastic, ORC-independent initiation of DNA replication relies on DNA recombination. This suggestion is also compatible with the SNS-seq data, since, similar to mouse and human cells, ~74% of the SNS-seq peaks appeared to be significantly associated with G-quadruplex motifs [125], structures that can lead to DNA Pol slow down/stalling and, consequently, DNA damage [75]. In this model, ORC does not recruit the rest of the replisome, but instead recombination factors perform this role, such as during break-induced replication [131,132]. Recombination-directed initiation of DNA replication, while unprecedented on such a scale, has the attraction of explaining the pervasive genome plasticity of *Leishmania*, including the role of recombination in the formation of genome-wide episomes [133-135]. Furthermore, recombination is needed to support continued replication in at least some polyploid archaea in which origins have been deleted [136]. Testing all these models will require genome-wide mapping of origins of replication in *Leishmania* using techniques other than MFA-seq and SNS-seq, allied to ChIP-seq mapping of key replication and recombination factors. In addition, it will be illuminating to ask if putative ORC-independent events are found in some circumstances in the *T. brucei* genome, and in other kinetoplastids.

Conclusions

Our understanding of DNA replication in kinetoplastids remains limited, but is growing. Most of our knowledge on this vital cellular process has been unravelled in the last ten years. In this period, we have moved from an archaeal-like model of the DNA replication initiator factor to a eukaryotic ORC-like complex, though a highly divergent one in which some subunits still remain to be identified or may be absent, and some appear to display pronounced sequence divergence. Due to its divergence from model eukaryotic ORCs, further analysis of kinetoplastid ORC structure and the functionality of the most diverged subunits, such as Tb3120, may illuminate ORC evolution and activity, as well as their potential as targets for drug development.

Presently, the emerging data on DNA replication in *Leishmania* is confusing, but we envisage that the current discord will be reconciled in time. Nonetheless, it seems clear that the dynamics of origin usage, replication timing and potentially replication execution in *Leishmania* are drastically different from *T. brucei* and other characterised microbial eukaryotes. Comparison of these two related kinetoplastids, and analysis of further organisms in this eukaryotic tree grouping, may reveal fundamental features of the evolution of eukaryotic chromosome replication, while uncovering the processes that shape the biology of *Leishmania*.

Conflict of Interest

The authors declare no competing interests.

Acknowledgments

Our recent work that contributed to this article was supported by the Wellcome Trust [089172, 093589], the BBSRC [BB/K006495/1] and Fundação para a Ciência e Tecnologia (FCT, Portugal) [SFRH/BD/68784/2010]. The Wellcome Centre for Molecular Parasitology is supported by core funding from the Wellcome Trust [104111]. We thank all friends and colleagues with whom we have enjoyed discussing the science explored in this article.

Figure Legends

Figure 1. The Origin Recognition Complex (ORC) of *T. brucei*. **A)** Evolution of our understanding of the origin-binding machinery of *T. brucei* in the last ten years, from a single factor like in archaea, to a highly divergent ORC-like complex resembling the system in model eukaryotes. **B)** Schematic representation of *T. brucei* ORC-like complex subunits and identified domains.

Figure 2. Origins of replication in *T. brucei* and *Leishmania*. Origins of replication in *T. brucei* chromosome 8 (orange) and *L. major* chromosome 36 (blue) mapped in early S phase cells by MFA-seq. The top track represents the gene coding sequences, and the graph below illustrates the read depth ratio between early S phase and G2 populations (y-axis), with each dot representing the median S/G2 ratio (y-axis) in 2.5 kbp windows across the chromosome (x-axis); adapted from Marques *et al.*, 2015. In the *L. major* panel, the vertical blue lines represent origins of replication mapped by SNS-seq; adapted from Lombraña *et al.*, 2016.

References

1. Gaillard H, Garcia-Muse T, Aguilera A (2015) Replication stress and cancer. *Nat Rev Cancer* 15: 276-289.
2. Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M (2010) Eukaryotic chromosome DNA replication: where, when, and how? *Annu Rev Biochem* 79: 89-130.
3. O'Donnell M, Langston L, Stillman B (2013) Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol* 5.
4. Fragkos M, Ganier O, Coulombe P, Mechali M (2015) DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol* 16: 360-374.
5. Siddiqui K, On KF, Diffley JF (2013) Regulating DNA replication in eukarya. *Cold Spring Harb Perspect Biol* 5.
6. Costa A, Hood IV, Berger JM (2013) Mechanisms for initiating cellular DNA replication. *Annu Rev Biochem* 82: 25-54.
7. Leonard AC, Mechali M (2013) DNA replication origins. *Cold Spring Harb Perspect Biol* 5: a010116.
8. Kawakami H, Katayama T (2010) DnaA, ORC, and Cdc6: similarity beyond the domains of life and diversity. *Biochem Cell Biol* 88: 49-62.
9. Skarstad K, Katayama T (2013) Regulating DNA replication in bacteria. *Cold Spring Harb Perspect Biol* 5: a012922.
10. Kelman LM, Kelman Z (2014) Archaeal DNA replication. *Annu Rev Genet* 48: 71-97.
11. Duncker BP, Chesnokov IN, McConkey BJ (2009) The origin recognition complex protein family. *Genome Biol* 10: 214.
12. Sazer S, Lynch M, Needleman D (2014) Deciphering the evolutionary history of open and closed mitosis. *Curr Biol* 24: R1099-1103.
13. Kolev NG, Ullu E, Tschudi C (2014) The emerging role of RNA-binding proteins in the life cycle of *Trypanosoma brucei*. *Cell Microbiol* 16: 482-489.
14. Yao MC, Chao JL, Cheng CY (2014) Programmed Genome Rearrangements in Tetrahymena. *Microbiol Spectr* 2(6): MDNA3-0012-2014.
15. Tiengwe C, Marques CA, McCulloch R (2014) Nuclear DNA replication initiation in kinetoplastid parasites: new insights into an ancient process. *Trends Parasitol* 30: 27-36.
16. Calderano SG, de Melo Godoy PD, da Cunha JP, Elias MC (2011) Trypanosome prereplication machinery: a potential new target for an old problem. *Enzyme Res* 2011: 518258.
17. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renault H, et al. (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309: 416-422.
18. Ivens AC, Peacock CS, Wortley EA, Murphy L, Aggarwal G, et al. (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309: 436-442.
19. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409-415.
20. Matthews KR (2015) 25 years of African trypanosome research: From description to molecular dissection and new drug discovery. *Mol Biochem Parasitol* 200(1-2): 30-40.
21. Godoy PD, Nogueira-Junior LA, Paes LS, Cornejo A, Martins RM, et al. (2009) Trypanosome prereplication machinery contains a single functional *orc1/cdc6* protein, which is typical of archaea. *Eukaryot Cell* 8: 1592-1603.
22. Kumar D, Mukherji A, Saha S (2008) Expression and subcellular localization of ORC1 in *Leishmania major*. *Biochem Biophys Res Commun* 375: 74-79.
23. Kumar D, Kumar D, Saha S (2012) A highly basic sequence at the N-terminal region is essential for targeting the DNA replication protein ORC1 to the nucleus in *Leishmania donovani*. *Microbiology* 158: 1775-1782.
24. Dang HQ, Li Z (2011) The Cdc45.Mcm2-7.GINS protein complex in trypanosomes regulates DNA replication and interacts with two Orc1-like proteins in the origin recognition complex. *J Biol Chem* 286: 32424-32435.
25. Tiengwe C, Marcello L, Farr H, Gadelha C, Burchmore R, et al. (2012) Identification of ORC1/CDC6-interacting factors in *Trypanosoma brucei* reveals critical features of origin recognition complex architecture. *PLoS One* 7: e32674.
26. Benmerzouga I, Concepcion-Acevedo J, Kim HS, Vadoros AV, Cross GA, et al. (2013) *Trypanosoma brucei* Orc1 is essential for nuclear DNA replication and affects both VSG silencing and VSG switching. *Mol Microbiol* 87: 196-210.
27. Marques CA, Tiengwe C, Lemgruber L, Damasceno JD, Scott A, et al. (2016) Diverged composition and regulation of the *Trypanosoma brucei* origin recognition complex that mediates DNA replication initiation. *Nucleic Acids Res* 44(10): 4763-4784.

28. Tiengwe C, Marcello L, Farr H, Dickens N, Kelly S, et al. (2012) Genome-wide analysis reveals extensive functional interaction between DNA replication initiation and transcription in the genome of *Trypanosoma brucei*. *Cell Rep* 2: 185-197.
29. Frigola J, Remus D, Mehanna A, Diffley JF (2013) ATPase-dependent quality control of DNA replication origin licensing. *Nature* 495: 339-343.
30. Aves SJ, Liu Y, Richards TA (2012) Evolutionary diversification of eukaryotic DNA replication machinery. *Subcell Biochem* 62: 19-35.
31. Harashima H, Dissmeyer N, Schnittger A (2013) Cell cycle control across the eukaryotic kingdom. *Trends Cell Biol* 23: 345-356.
32. Matte A, Delbaere LTJ (2010) ATP-binding Motifs. eLS: John Wiley & Sons, Ltd.
33. Iyer LM, Leipe DD, Koonin EV, Aravind L (2004) Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 146: 11-31.
34. Bleichert F, Botchan MR, Berger JM (2015) Crystal structure of the eukaryotic origin recognition complex. *Nature* 519: 321-326.
35. Liu S, Balasov M, Wang H, Wu L, Chesnokov IN, et al. (2011) Structural analysis of human Orc6 protein reveals a homology with transcription factor TFIIB. *Proc Natl Acad Sci U S A* 108: 7373-7378.
36. Speck C, Chen Z, Li H, Stillman B (2005) ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nat Struct Mol Biol* 12: 965-971.
37. Clarey MG, Erzberger JP, Grob P, Leschziner AE, Berger JM, et al. (2006) Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nat Struct Mol Biol* 13: 684-690.
38. Chesnokov IN, Chesnokova ON, Botchan M (2003) A cytokinetic function of *Drosophila* ORC6 protein resides in a domain distinct from its replication activity. *Proc Natl Acad Sci U S A* 100: 9150-9155.
39. Prasanth SG, Prasanth KV, Stillman B (2002) Orc6 involved in DNA replication, chromosome segregation, and cytokinesis. *Science* 297: 1026-1031.
40. Takara TJ, Bell SP (2011) Multiple Cdt1 molecules act at each origin to load replication-competent Mcm2-7 helicases. *EMBO J* 30: 4885-4896.
41. Tocilj A, On KF, Yuan Z, Sun J, Elkayam E, et al. (2017) Structure of the active form of human origin recognition complex and its ATPase motor module. *Elife* 6.
42. Tiengwe C (2010) Analysis of the architecture and function of the nuclear DNA replication apparatus in *Trypanosoma brucei* [PhD]: University of Glasgow. 270 p.
43. Marques CA (2015) Initiation of nuclear DNA replication in *Trypanosoma brucei* and *Leishmania*. [PhD]: University of Glasgow. 402 p.
44. Hou Z, Bernstein DA, Fox CA, Keck JL (2005) Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. *Proc Natl Acad Sci U S A* 102: 8489-8494.
45. Ozaydin B, Rine J (2010) Expanded roles of the origin recognition complex in the architecture and function of silenced chromatin in *Saccharomyces cerevisiae*. *Mol Cell Biol* 30: 626-639.
46. Bell SP, Kobayashi R, Stillman B (1993) Yeast origin recognition complex functions in transcription silencing and DNA replication. *Science* 262: 1844-1849.
47. Bell SP, Mitchell J, Leber J, Kobayashi R, Stillman B (1995) The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. *Cell* 83: 563-568.
48. Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, et al. (2012) The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484: 115-119.
49. Noguchi K, Vassilev A, Ghosh S, Yates JL, DePamphilis ML (2006) The BAH domain facilitates the ability of human Orc1 protein to activate replication origins *in vivo*. *The EMBO Journal* 25: 5372-5382.
50. Davey MJ, Jeruzalmi D, Kuriyan J, O'Donnell M (2002) Motors and switches: AAA+ machines within the replisome. *Nat Rev Mol Cell Biol* 3: 826-835.
51. Sun J, Kawakami H, Zech J, Speck C, Stillman B, et al. (2012) Cdc6-induced conformational changes in ORC bound to origin DNA revealed by cryo-electron microscopy. *Structure* 20: 534-544.
52. Bleichert F, Balasov M, Chesnokov I, Nogales E, Botchan MR, et al. (2013) A Meier-Gorlin syndrome mutation in a conserved C-terminal helix of Orc6 impedes origin recognition complex formation. *Elife* 2: e00882.
53. Chuang RY, Kelly TJ (1999) The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proc Natl Acad Sci U S A* 96: 2656-2661.
54. Samson RY, Abeyrathne PD, Bell SD (2016) Mechanism of Archaeal MCM Helicase Recruitment to DNA Replication Origins. *Mol Cell* 62: 287-296.

55. Uzcanga G, Lara E, Gutierrez F, Beaty D, Beske T, et al. (2016) Nuclear DNA replication and repair in parasites of the genus *Leishmania*: Exploiting differences to develop innovative therapeutic approaches. *Crit Rev Microbiol*: 1-22.
56. Akiyoshi B, Gull K (2014) Discovery of unconventional kinetochores in kinetoplastids. *Cell* 156: 1247-1258.
57. Nerusheva OO, Akiyoshi B (2016) Divergent polo box domains underpin the unique kinetoplastid kinetochore. *Open Biol* 6(3): 150206.
58. D'Archivio S, Wickstead B (2016) Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J Cell Biol*.
59. Rhind N, Gilbert DM (2013) DNA replication timing. *Cold Spring Harb Perspect Biol* 5: a010132.
60. McIntosh D, Blow JJ (2012) Dormant origins, the licensing checkpoint, and the response to replicative stresses. *Cold Spring Harb Perspect Biol* 4(10): a012955.
61. Bell SP, Stillman B (1992) ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 357: 128-134.
62. Mechali M (2010) Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol* 11: 728-738.
63. Hyrien O (2015) Peaks cloaked in the mist: the landscape of mammalian replication origins. *J Cell Biol* 208: 147-160.
64. Segurado M, de Luis A, Antequera F (2003) Genome-wide distribution of DNA replication origins at A+T-rich islands in *Schizosaccharomyces pombe*. *EMBO Rep* 4: 1048-1053.
65. MacAlpine DM, Rodriguez HK, Bell SP (2004) Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev* 18: 3094-3105.
66. Delgado S, Gomez M, Bird A, Antequera F (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. *Embo j* 17: 2426-2435.
67. Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, et al. (2008) Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci U S A* 105: 15837-15842.
68. Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, et al. (2009) Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet* 5: e1000446.
69. Cayrou C, Coulombe P, Vigneron A, Stanojic S, Ganier O, et al. (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res* 21: 1438-1449.
70. Costas C, de la Paz Sanchez M, Stroud H, Yu Y, Oliveros JC, et al. (2011) Genome-wide mapping of *Arabidopsis thaliana* origins of DNA replication and their associated epigenetic marks. *Nat Struct Mol Biol* 18: 395-400.
71. Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, et al. (2012) New insights into replication origin characteristics in metazoans. *Cell Cycle* 11: 658-667.
72. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, et al. (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol* 19: 837-844.
73. Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, et al. (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *Embo j* 33: 732-746.
74. Cayrou C, Ballester B, Peiffer I, Fenouil R, Coulombe P, et al. (2015) The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome Res* 25: 1873-1885.
75. Valton AL, Prioleau MN (2016) G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet* 32: 697-706.
76. Comoglio F, Schlumpf T, Schmid V, Rohs R, Beisel C, et al. (2015) High-resolution profiling of *Drosophila* replication start sites reveals a DNA shape and chromatin signature of metazoan origins. *Cell Rep* 11: 821-834.
77. Remus D, Beall EL, Botchan MR (2004) DNA topology, not DNA sequence, is a critical determinant for *Drosophila* ORC-DNA binding. *Embo j* 23: 897-907.
78. MacAlpine HK, Gordan R, Powell SK, Hartemink AJ, MacAlpine DM (2010) *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res* 20: 201-211.
79. Soriano I, Morafraile EC, Vazquez E, Antequera F, Segurado M (2014) Different nucleosomal architectures at early and late replicating origins in *Saccharomyces cerevisiae*. *BMC Genomics* 15: 791.
80. Lubelsky Y, Sasaki T, Kuipers MA, Lucas I, Le Beau MM, et al. (2011) Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. *Nucleic Acids Res* 39: 3141-3155.

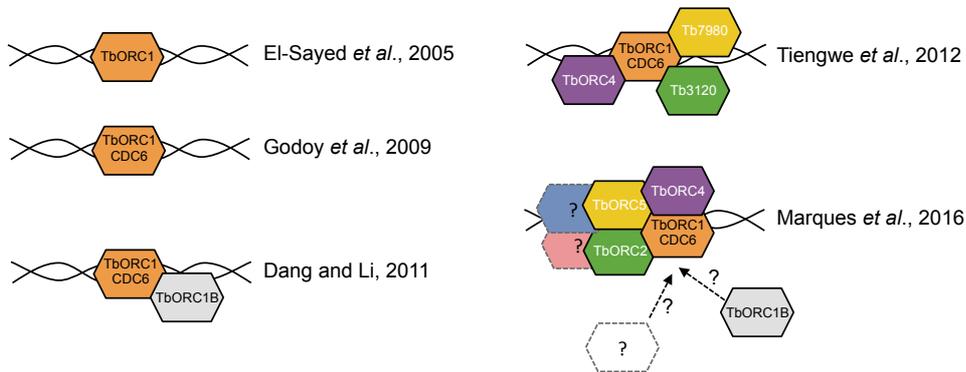
81. Unnikrishnan A, Gafken PR, Tsukiyama T (2010) Dynamic changes in histone acetylation regulate origins of DNA replication. *Nat Struct Mol Biol* 17: 430-437.
82. Crampton A, Chang F, Pappas DL, Jr., Frisch RL, Weinreich M (2008) An ARS element inhibits DNA replication through a SIR2-dependent mechanism. *Mol Cell* 30: 156-166.
83. Iizuka M, Matsui T, Takisawa H, Smith MM (2006) Regulation of replication licensing by acetyltransferase Hbo1. *Mol Cell Biol* 26: 1098-1108.
84. Miotto B, Struhl K (2010) HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol Cell* 37: 57-66.
85. Sequeira-Mendes J, Gomez M (2012) On the opportunistic nature of transcription and replication initiation in the metazoan genome. *Bioessays* 34: 119-125.
86. Helmrich A, Ballarino M, Nudler E, Tora L (2013) Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol* 20: 412-418.
87. Goldman MA, Holmquist GP, Gray MC, Caston LA, Nag A (1984) Replication timing of genes and middle repetitive sequences. *Science* 224: 686-692.
88. Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, et al. (2013) Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome Res* 23: 1-11.
89. Heun P, Laroche T, Raghuraman MK, Gasser SM (2001) The positioning and dynamics of origins of replication in the budding yeast nucleus. *J Cell Biol* 152: 385-400.
90. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, et al. (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A* 107: 139-144.
91. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, et al. (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453: 948-951.
92. Wilson RH, Coverley D (2013) Relationship between DNA replication and the nuclear matrix. *Genes Cells* 18: 17-31.
93. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17: 333-351.
94. Zentner GE, Henikoff S (2014) High-resolution digital profiling of the epigenome. *Nat Rev Genet* 15: 814-827.
95. Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12: 7-18.
96. Laird PW (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 11: 191-203.
97. Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 10: 669-680.
98. Ingolia NT (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat Rev Genet* 15: 205-213.
99. Konig J, Zarnack K, Luscombe NM, Ule J (2012) Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* 13: 77-83.
100. Oszolak F, Milos PM (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12: 87-98.
101. Gilbert DM (2010) Evaluating genome-scale approaches to eukaryotic DNA replication. *Nat Rev Genet* 11: 673-684.
102. Patnaik PK (1997) Studies with artificial extrachromosomal elements in trypanosomatids: could specificity in the initiation of DNA replication be linked to that in transcription? *Parasitol Today* 13: 468-471.
103. Tamar S, Papadopoulou B (2001) A telomere-mediated chromosome fragmentation approach to assess mitotic stability and ploidy alterations of *Leishmania* chromosomes. *J Biol Chem* 276: 11662-11673.
104. Weiden M, Osheim YN, Beyer AL, Van der Ploeg LH (1991) Chromosome structure: DNA nucleotide sequence elements of a subset of the minichromosomes of the protozoan *Trypanosoma brucei*. *Mol Cell Biol* 11: 3823-3834.
105. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, et al. (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309: 404-409.
106. Jackson AP (2014) Genome evolution in trypanosomatid parasites. *Parasitology* 142: S40-S56.
107. Ersfeld K (2011) Nuclear architecture, genome and chromatin organisation in *Trypanosoma brucei*. *Res Microbiol* 162: 626-636.
108. Michaeli S (2011) Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol* 6: 459-474.

109. De Gaudenzi JG, Noe G, Campo VA, Frasch AC, Cassola A (2011) Gene expression regulation in trypanosomatids. *Essays Biochem* 51: 31-46.
110. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, et al. (2011) Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* 21: 2129-2142.
111. Mannaert A, Downing T, Imamura H, Dujardin JC (2012) Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. *Trends Parasitol* 28: 370-376.
112. Glover L, Hutchinson S, Alford S, McCulloch R, Field MC, et al. (2013) Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in VSG expression control. *Cell Microbiol* 15: 1984-1993.
113. Muller CA, Nieduszynski CA (2012) Conservation of replication timing reveals global and local regulation of replication origin activity. *Genome Res* 22: 1953-1962.
114. Muller CA, Hawkins M, Retkute R, Malla S, Wilson R, et al. (2014) The dynamics of genome replication using deep sequencing. *Nucleic Acids Res* 42: e3.
115. McGuffee SR, Smith DJ, Whitehouse I (2013) Quantitative, genome-wide analysis of eukaryotic replication initiation and termination. *Mol Cell* 50: 123-135.
116. Obado SO, Bot C, Nilsson D, Andersson B, Kelly JM (2007) Repetitive DNA is associated with centromeric domains in *Trypanosoma brucei* but not *Trypanosoma cruzi*. *Genome Biol* 8: R37.
117. Devlin R, Marques CA, Paape D, Prorocic M, Zurita-Leal AC, et al. (2016) Mapping replication dynamics in *Trypanosoma brucei* reveals a link with telomere transcription and antigenic variation. *Elife* 5: e12765.
118. Devlin R, Marques CA, McCulloch R (2016) Does DNA replication direct locus-specific recombination during host immune evasion by antigenic variation in the African trypanosome? *Curr Genet*: doi:10.1007/s00294-00016-00662-00297.
119. Calderano SG, Drosopoulos WC, Quaresma MM, Marques CA, Kosiyatrakul S, et al. (2015) Single molecule analysis of *Trypanosoma brucei* DNA replication dynamics. *Nucleic Acids Res* 43(5): 2655-2665.
120. 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 Biol* 16: 230.
121. Wincker P, Ravel C, Blaineau C, Pages M, Jauffret Y, et al. (1996) The *Leishmania* genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. *Nucleic Acids Res* 24: 1688-1694.
122. Britto C, Ravel C, Bastien P, Blaineau C, Pages M, et al. (1998) Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. *Gene* 222: 107-117.
123. Wheeler RJ, Gluenz E, Gull K (2011) The cell cycle of *Leishmania*: morphogenetic events and their implications for parasite biology. *Mol Microbiol* 79: 647-662.
124. Stanojic S, Sollelis L, Kuk N, Crobu L, Balard Y, et al. (2016) Single-molecule analysis of DNA replication reveals novel features in the divergent eukaryotes *Leishmania* and *Trypanosoma brucei* versus mammalian cells. *Sci Rep* 6: 23142.
125. Lombrana R, Alvarez A, Fernandez-Justel JM, Almeida R, Poza-Carrion C, et al. (2016) Transcriptionally Driven DNA Replication Program of the Human Parasite *Leishmania major*. *Cell Rep* 16: 1774-1786.
126. Rocha-Granados MC, Klingbeil MM (2016) *Leishmania* DNA Replication Timing: A Stochastic Event? *Trends Parasitol* 32: 755-757.
127. Picard F, Cadoret JC, Audit B, Arneodo A, Alberti A, et al. (2014) The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet* 10: e1004282.
128. Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, et al. (2016) Replication landscape of the human genome. *Nat Commun* 7: 10208.
129. 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 Biol* 16: 230.
130. Hyrien O (2016) How MCM loading and spreading specify eukaryotic DNA replication initiation sites. *F1000Res* 5.
131. Lydeard JR, Lipkin-Moore Z, Sheu YJ, Stillman B, Burgers PM, et al. (2010) Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes Dev* 24: 1133-1144.

132. Anand RP, Lovett ST, Haber JE (2013) Break-induced DNA replication. *Cold Spring Harb Perspect Biol* 5: a010397.
133. Laffitte MC, Genois MM, Mukherjee A, Legare D, Masson JY, et al. (2014) Formation of linear amplicons with inverted duplications in *Leishmania* requires the MRE11 nuclease. *PLoS Genet* 10: e1004805.
134. Genois MM, Mukherjee A, Ubeda JM, Buisson R, Paquet E, et al. (2012) Interactions between BRCA2 and RAD51 for promoting homologous recombination in *Leishmania infantum*. *Nucleic Acids Res* 40: 6570-6584.
135. Laffitte MC, Leprohon P, Hainse M, Legare D, Masson JY, et al. (2016) Chromosomal Translocations in the Parasite *Leishmania* by a MRE11/RAD50-Independent Microhomology-Mediated End Joining Mechanism. *PLoS Genet* 12: e1006117.
136. Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T (2013) Accelerated growth in the absence of DNA replication origins. *Nature* 503: 544-547.

Figure 1

A.



B.

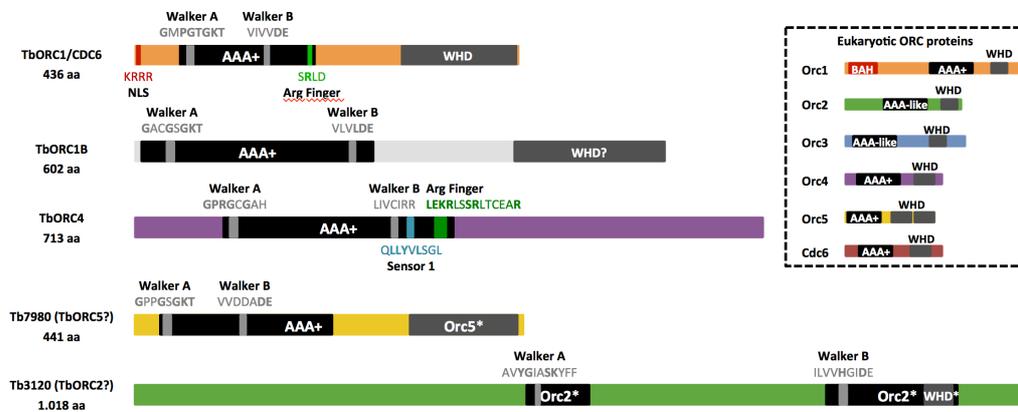
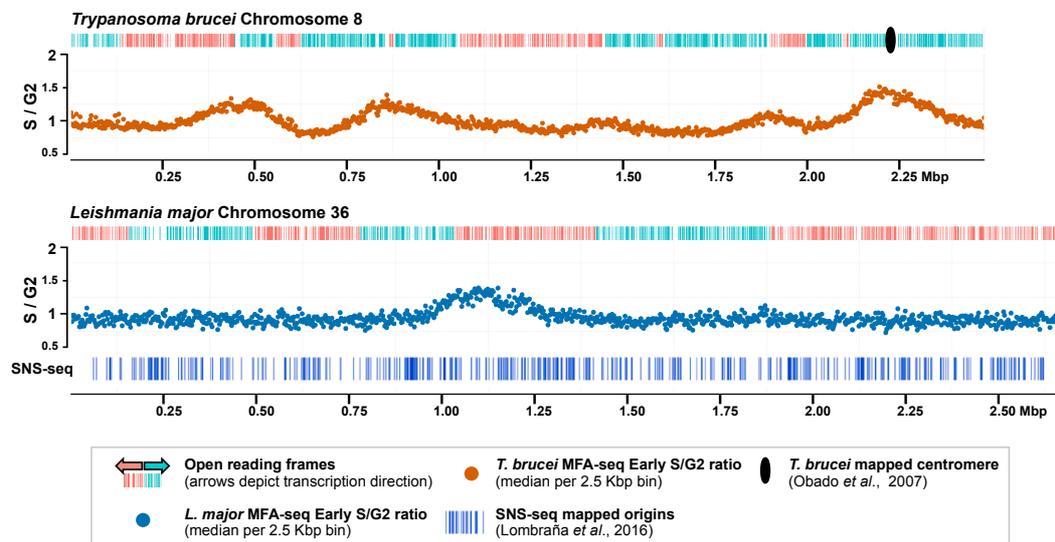
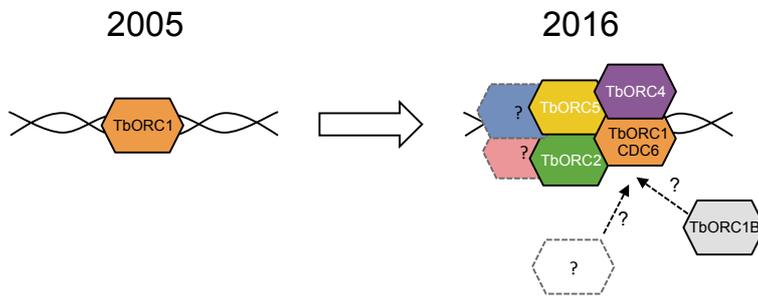


Figure 2



Graphical Abstract

Origin recognition complex (ORC) in kinetoplastids**Origins of replication in kinetoplastids***Trypanosoma brucei**Leishmania major*