Protist, Vol. 170, 141–152, April 2019
http://www.elsevier.de/protis
Published online date 13 February 2019

Protist

ORIGINAL PAPER

Mapping the Single Origin of Replication in the *Naegleria gruberi* Extrachromosomal DNA Element

John C. Mullican[1,2], **Nora M. Chapman, and Steven Tracy**

Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA

Submitted October 25, 2018; Accepted February 6, 2019
Monitoring Editor: C. Graham Clark

The genes encoding the ribosomal RNA (rRNA) subunits of the amoeba *Naegleria gruberi* are encoded in a relatively uncommon arrangement: on a circular extrachromosomal DNA element with each organism carrying about 4,000 copies of the element. As complete sequence analysis of the *N. gruberi* chromosomal DNA revealed no copy of the rRNA genes, these extrachromosomal elements must therefore replicate autonomously. We reported elsewhere the molecular cloning and the complete sequence analysis of the entire rRNA gene-containing element of *N. gruberi* (strain EG_B). Using neutral/neutral two-dimensional agarose electrophoresis, the region in the element enclosing the single replication origin using DNA from asynchronous and axenically propagated *N. gruberi* populations was localized within a 2.1 kbp fragment located approximately 2,300 bp from the 18*S* rRNA gene and 3,700 bp from the 28*S* rRNA gene. The results indicate that replication occurs from a single origin via a *theta*-type mode of replication rather than by a rolling circle mode. Further, G-quadruplex elements, often located near DNA replication origins, occur in and near this fragment in a repeated sequence.

Key words: Amoeba; *Naegleria gruberi*; ribosomal RNA; extrachromosomal DNA element; origin of replication; G-quadruplex.

[1]Corresponding author.
[2]Current address: Department of Biology, Washburn University, Topeka, Kansas 66621, USA.
e-mail john.mullican@washburn.edu (J.C. Mullican).

Abbreviations: TBE, Tris-borate-EDTA; ARS, Autonomous replication sequence; ORC, Origin Replication Complex; *ori*, origin; mtDNA, mitochondrial DNA; NRS, Non-Ribosomal Sequence; G4, G-quadruplex; pG4, Potential G-quadruplex; rDNA, ribosomal DNA; RDE, Ribosomal DNA-containing Element; rRNA, ribosomal RNA.

https://doi.org/10.1016/j.protis.2019.02.001
1434-4610/

Introduction

Free-living amoebae, such as the 47 characterized species of the *Naegleria* genus (De Jonckheere 2014), are ubiquitous in soil and fresh water worldwide (Denet et al. 2017; Rodriguez-Zaragoza 1994; Tyml et al. 2016). *Naegleria fowleri* is the sole known *Naegleria* species that is pathogenic for humans and is infamous due to its ability, following ingestion into the nasal cavity, to induce the nearly universally fatal disease, primary amoebic meningoencephalitis (Grace et al. 2015). The non-pathogenic *Naegleria gruberi*, on the other hand, represents a useful research model organism (Ahmed Khan et al. 2015; Opperdoes et al. 2011).

From bacteria to humans, all cells contain repetitive rRNA genes (rDNA) (Long and Dawid 1980). Examining the NEG-M strain of *N. gruberi*, Clark and Cross (1987) first demonstrated that this *Naegleria* species carries its ribosomal RNA (rRNA) genes on an extrachromosomal element; later, Maruyama and Nozaki (2007) sequenced the extrachromosomal element of this particular strain. Complete sequence analysis of the *N. gruberi* organism's nuclear DNA (Fritz-Laylin et al. 2010) confirmed earlier work based on Southern blot hybridization analysis (Clark and Cross 1987) that there is no copy of the ribosomal DNA in the chromosomal DNA. While this arrangement is uncommon, the placement of rDNA only on an extrachromosomal element has also been demonstrated for other members of the order *Schizopyrenida* (Clark and Cross 1988). *Entamoeba histolytica* and *E. invadens* also lodge the rDNA on an extrachromosomal element (Bhattacharya et al. 1989, 1998) as do some other eukaryotic microbial species, several of which possess circular rDNA (Torres-Machorro et al. 2010).

Earlier work estimated the number of the ribosomal DNA-containing extrachromosomal element (RDE) molecules at about 4,000 per cell based on visual analysis of slot-blot data (Clark and Cross 1987). Complete sequence analysis of the entire genome of *N. gruberi* (strain NEG-M) (Fritz-Laylin et al. 2010) revealed the mass of the nuclear DNA per organism to be 0.045 pg DNA (41×10^6 bp per genome, calculated at 660 Da/bp). Each organism (Clark 1990) also contains approximately 420 copies of mitochondrial DNA (mtDNA) which, at 50 kbp per copy (Fritz-Laylin et al. 2010), yields about 0.023 pg mtDNA per cell. Thus, at about 4,000 copies of the 1.4×10^4 bp RDE per cell (Clark and Cross 1987; Maruyama and Nozaki 2007) or about 0.061 pg per cell, the RDE population represents about 47% of the total DNA mass (nuclear chromosomal, RDE and mtDNA) of the *N. gruberi* organism. This is considerably higher than the earlier estimation of 17% (Clark and Cross 1987), a discrepancy that can be attributed to more precise measurements of both the genome and element sizes (Fritz-Laylin et al. 2010). As the mass of the RDE population's DNA represents a significant fraction of the total DNA in a *Naegleria* organism, it will be important to understand the mechanism by which this extrachromosomal element replicates.

With thousands of copies of the RDE per cell (Clark and Cross 1987) and no 'master' copy of rDNA or the RDE itself in the chromosomal DNA (Clark and Cross 1987; Fritz-Laylin et al. 2010), Clark and Cross (1987) inferred that the RDE must replicate autonomously. There have been no published studies to date designed to map the origin(s) of replication (*ori*) of the RDE in *Naegleria*. Studies of eukaryotic DNA replication *ori* have focused on chromosomes in the yeast, human and murine genomes (Parker et al. 2017). It is noteworthy that the primary structure requirements of eukaryotic DNA *ori* are not well understood, save for yeast; indeed, there is a better understanding of the proteins involved in the *ori* replication complex (ORC) than of the requirements of the DNA site(s) that is/are recognized [reviewed in (Leonard and Mechali 2013; O'Donnell et al. 2013)].

Nonetheless, the search for specific primary DNA structure correlates of DNA *ori* has identified certain genetic elements often associated with replication initiation. The discovery that GC enrichment of regions where origins are found in metazoan DNA has led to the recognition of a role for a structural component called the G-quadruplex (G4) (Prioleau 2017). The G4s derive from primary sequences containing four separated repeats of two or more G nucleotides. They can form structures, based on the non-canonical interactions [Hoogsteen base pairing; (Hoogsteen 1963)] of the G nucleotides, to form 2 or more planar structures. These are termed G tetrads [(Sen and Gilbert 1988); reviewed and depicted in (Maizels 2006)]. Such tetrads can stack upon each other to form the G-quadruplex (Supplementary Material Fig. S1). Recognition of these sequence motifs associated with DNA *ori* (Cayrou et al. 2012) has prompted mutational analyses, which demonstrate that while these components are not sufficient for an origin of DNA replication, they are necessary at least for some defined sites (Valton et al. 2014).

Autonomous replication sequences (ARS) with conserved origin sequence motifs have been identified in yeast species. And, just as for the G4s, these appear to be necessary for *ori* function (Peng et al. 2015). Interestingly, two sequences with homology to the conserved *ori* sequence motifs of *Saccharomyces cerevisiae* were noted in the *N. gruberi* ribosomal DNA extrachromosomal element (Mullican et al. 2018). The nonribosomal region of the RDE is largely composed of directly repeated DNA sequences ranging in size up to 315 nucleotides and include both interspersed and tandemly repeated sequences. Due to the high copy number of these RDEs in *Naegleria* cells as well as the considerable evolutionary divergence of these *Excavata* protists from metazoans and yeast, and the repeated observations that DNA primary structure in *ori* is quite variable, it is possible that the primary (as well as higher order) structure of the *ori* in the *Naegleria* RDE may differ from other eukaryotic origins studied (Leonard and Mechali 2013; Parker et al. 2017).

As the mechanism of the *Naegleria* RDE replication and regulation remains unexplored, we initiated studies toward defining the RDE *ori*. We hypothesized that each RDE possesses at least one *ori* and we sought to map the location of such a domain within a *Naegleria* RDE. Using our previously published cloned sequence of the *N. gruberi* RDE sequence (Mullican et al. 2018) along with DNA isolated from asynchronously growing axenic cultures, we employed neutral/neutral (N/N) two-dimensional (2D) agarose gel electrophoresis to map the RDE's sole origin of DNA replication.

Results

We used the complete molecular clone (pNgrubEGB) of the *N. gruberi* RDE and its sequence information (Mullican et al. 2018) to localize the origin of replication (*ori*) within the RDE. DNA eluted from BND-cellulose columns was highly enriched for DNA fragments containing regions of ssDNA and replication intermediates, which enabled us to assay DNA replication events in asynchronous axenically grown cell populations for the work. Although *N. gruberi* can be synchronized mitotically when given bacteria as a food source (Fulton and Guerrini 1969), they have not been synchronized in axenic culture. Thus, we used cultures growing logarithmically in culture.

Cloning and Sequence Analysis of pNgrubEGB

The RDE was ligated into the pGEM-7Zf(+) vector at the *Bam*HI site and subsequently restriction mapped (Fig. 1). Sequence analysis of this clone revealed the location of the 5.8*S*, 18*S* and 28*S* rRNA genes as well as numerous repeat sequences [(Mullican et al. 2018); annotated in GenBank Accession MG699123]. We found (Mullican et al. 2018), as had Maruyama and Nozaki (2007) earlier, that the 18*S* rDNA of the two different strains of *N. gruberi* lacked the type I intron reported to be in most *Naegleria* species' RDEs (Wikmark et al. 2006). The two sequences derived from two closely related *N. gruberi* strains are very similar with most differences contained within a region of 120 bp tandem repeats in the non-rDNA coding region. The RDEs of the 2 different amoebic strains differ at only 4 sites (with nucleotide numbering referring to the pNgrubEGB sequence positions; GenBank MG699123): a single nucleotide (nt) deletion (nt8581-8582), a single nucleotide variation (nt8703) and a 119 bp deletion (relative to the sequence for strain NEG-M; GenBank AB298288) of part of the 4th and 5th copy of the 120 bp tandem repeat (nt8748-8749) of the nonribosomal sequence (NRS). There is also a single nt deletion in the EG_B strain's RDE at nt13736-13737 upstream of the open reading frame closest to the 18*S* rDNA.

Seventy-five potential G-quadruplexes (pG4) of the form $G_{2\text{-}5}L1G_{2\text{-}5}L2G_{2\text{-}5}L3G_{2\text{-}5}$ as described by Dolinnaya and colleagues (Dolinnaya et al. 2016) exist in the RDE (43 on the rRNA coding strand and 32 on the antisense strand: Supplementary Table). By setting the minimum QGRS score (see Methods) to 20, we limited the pG4s to nineteen. These are located within three regions: the segment (nt188-538) upstream of 18*S* rRNA coding region (nt827-2846), in which nine pG4s are located on the strand that is antisense to the rRNA; four sense strand pG4s in the 28*S* coding region; and six pG4s (2 sense and 4 antisense) in the remainder of the NRS (Table 1, Fig. 1). The highest scoring pG4s (QGRS score 35) have 3 G-tetrads and are located on the antisense strand; these are nearly identical due to their presence in two copies of 79 bp direct repeat in the NRS sequence (nt9676-9754, nt11550-11628).

Mapping the Origin of Replication

To map the *ori*, overlapping restriction fragments of replicating DNA (Fig. 1) were analyzed by N/N two-dimensional agarose gel electrophoresis. Using this technique, one may discriminate between key structures occurring during plasmid replication, the primary sorts of structures of which are depicted in Figure 2. A replication fork (single Y structure) in a specific region will appear as an arc (Fig. 2A)

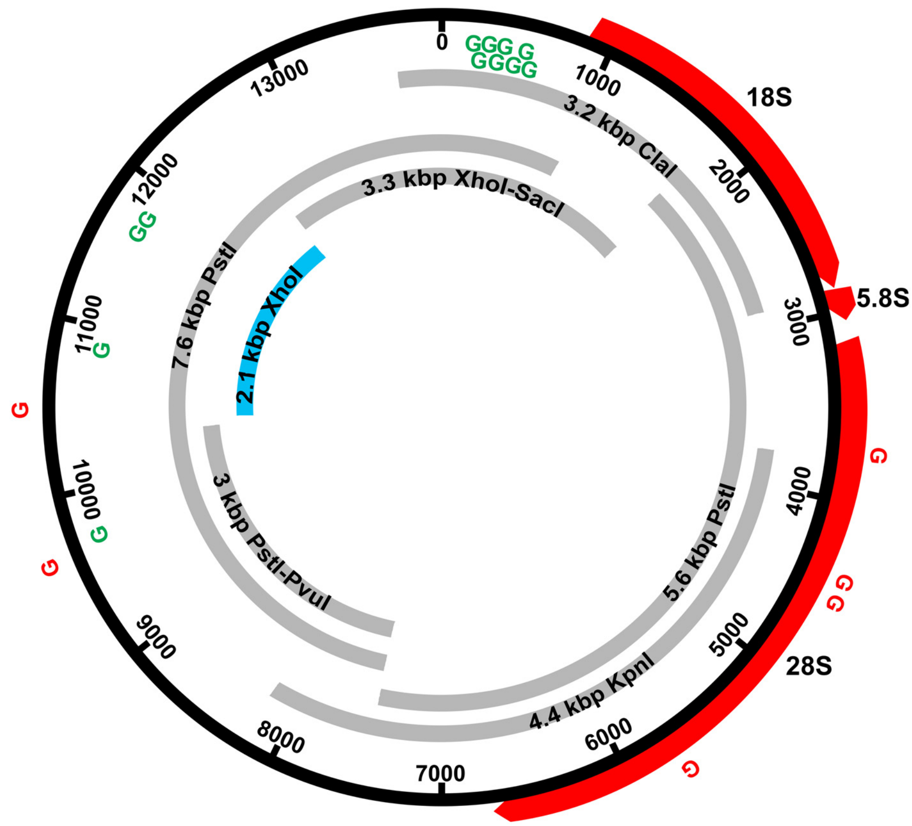

Figure 1. Circular restriction map of the native rDNA plasmid from *N. gruberi*, EG_B (GenBank Accession: MG699123). The 12 o'clock position is the unique *Bam*HI site in which the rDNA plasmid was cloned into the pGEM7Zf(+) vector. Locations of the rRNA genes oriented in a clockwise fashion are depicted on the map in red and overlapping restriction fragments (in grey) used in the N/N 2D agarose gel electrophoresis analyses. The *Xho*I fragment containing the *ori* is shown in blue. pG4s are indicated in approximate positions (red G indicates on rRNA coding strand; green G on antisense).

whereas a double fork (double Y structure) will appear as a straight line (Fig. 2B). A bubble structure, indicating the presence of a replication origin in that restriction fragment, will appear as an arcing 'flare' above the simple Y structures (Fig. 2C). Figure 2D depicts the expected pattern of both bubble and single Y structures when the origin is asymmetrically located within the restriction fragment. Other images reflecting variations on these structures have been described (Robinson et al. 2004).

Analyses of the 3.2 kbp *Cla*I fragment (nt13713-2913) and the 4.4 kbp *Kpn*I fragment (nt3771-8218) generated across the rRNA coding region (Fig. 1) showed predominantly simple Y structures (Fig. 3A, C). In the 5.6 kbp *Pst*1 fragment (nt1789-7436), predominately double Y structures were seen (Fig. 3B, Supplementary Material Fig. S2). Such a structure suggests a possible termination site in this fragment where opposing replication forks meet (Little et al. 1993). It is also possible that termination sequences exist downstream of the 28*S* rRNA gene which localize to this region (Bartsch et al. 1987; Grummt et al. 1985) but studies of sequence conservation within the *Naegleria* genus and experimental proof will be required to positively identify them. 2D-gel analysis of the 4.4 kbp *Kpn*I fragment, which also encompasses the downstream region of the 28*S* rRNA gene, does not produce any discernible double Y structures (Fig. 3C).

Evidence suggesting bubble structures was observed when using the large 7.6 kbp *Pst*1 fragment (nt7437-1000) derived from the NRS of pNgrubEGB, suggesting the presence of an *ori* within this fragment (Fig. 4A). The intense signal below the bubble arc in Figure 4A may indicate either a pause in replication or the presence of recombination intermediates. Alternatively, as this fragment is large, the signal may be a distortion of the Y arc during electrophoresis (Hyrien

Table 1. Potential G4s with G-scores (QGRS) 20 or greater in pNgrubEGB. Numbering according to GenBank MG699123 on the rDNA coding strand. RC indicates nucleotide positions on reverse complement strand. Underscored text indicates presence within the 2.1 kbp *Xho*I fragment of the extrachromosomal element. Asterisk (*) on the G-score indicates the presence within a copy of the 79 bp direct repeat.

Position	Length	QGRS	G-Score
RC of 188-202	15	GGTTTGGTTGGTTGG	20
RC of 213-238	26	GGCTGGCTGGTGGGCTGGTTGGCTGG	21
RC of 241-264	25	GGGCCAGAGGGCGCGTGGCTGGTGG	20
RC of 315-334	26	GGCTGGCTGGTGGGCTGGTTGGCTGG	21
RC of 346-363	18	GGTCAGGGCTGGCAGCGG	20
RC of 420-433	14	GGCCGGACGGTTGG	21
RC of 489-507	19	GGCTGCGGTTTGGCTGTGG	20
RC of 525-538	14	GGTAGGCTGGTAGG	21
18S rDNA (827-2846)			
ITS1 (2847-2879)			
5.8S rDNA (2880-3054)			
ITS2 (3055-3168)			
28S rDNA (3169-6681)			
3755-3773	19	GGCTTTGGTGTGGTACCGG	20
4410-4431	22	GGACGGTGGCCATGGAAGTCGG	20
4543-4555	13	GGGGGTAGGCAGG	20
5658-5684	27	GGGAACGGGCTTGGGTTGGTTAGCGGG	36
Nonribosomal (6682-14007, 1-826)			
9637-9653	17	GGGTCGGATCGGGTCGG	21
RC of 9719-9746	28	GGGTCTCCCAGGGCCCTACGTGGGGGGG	35*
<u>10465-10491</u>	27	GGTCAAAACGGCTTCTAGGACTCCTGG	20
<u>RC of 10841-10869</u>	29	GGAAGTAAGGGCATTTTTGGGAAATACGG	21
<u>RC of 11593-11620</u>	28	GGGTCTCCCAGGGCCCTACCTGGGGGGG	35*
<u>RC of 11728-11740</u>	13	GGTCGGGGGGTGG	20

Note: The bold-faced text identifies the significant portions of the rDNA plasmid, noting the sequence boundaries in parentheses. The underscored text indicates the presence of pG4s in the XhoI fragment of the Nonribosomal sequence.

and Mechali 1992). Analysis of the 3 kbp fragment generated by double digestion with *Pst*I and *Pvu*I (nt7436-10357) yielded simple Y structures (Fig. 4B). A fragment generated by cleavage with both *Xho*I and *Sac*I (3.3 kbp; nt12552-1859) also yielded predominately simple Y structures (Fig. 4C). However, another fragment of 2.1 kbp, located in the middle of the 7.6 kbp *Pst*I fragment (Fig. 1), derived by cleaving pNgrubEGB with *Xho*I (nt10423-12553), showed arcs representative of both bubbles and simple Y structures (Fig. 4D, Supplementary Material Fig. S3), which are indicative of an *ori* located asymmetrically within this sequence.

Repeat Sequences in the Element

Data from the sequence analysis of pNgrubEGB (Mullican et al. 2018) were used to analyze the presence of direct and indirect repeat sequences in the region (*Xho*I nt10423-12553) which contain the *ori*. This region is separated from the beginning of the 18*S* rRNA-encoding DNA by 2,280 bp and from the end of the 28*S* rRNA encoding DNA by 3,742 bp. Although two conserved origin sequence motif elements like those in *S. cerevisae* (Deshpande and Newlon 1992) have been identified in the *N. gruberi* NRS (Mullican et al. 2018), neither occurs within this *Xho*I restriction fragment. There are two large direct repeat sequences (a 245 bp repeat at nt10790-11034 and a 79 bp repeat at nt11550-11628) and one short, inverted repeat sequence (nt11855-11876 with nt12579-12600) which have copies within the *Xho*I restriction fragment. All three of these have only one copy within the *Xho*I fragment. Only one short (12 bp) direct repeat sequence has all copies of the repeat sequence within the *Xho*I fragment (nt11963-11975; nt12081-12092). Four pG4 sequences are located within the *Xho*1 fragment that have scores greater than or equal to 20. One of these is on the rRNA coding strand and three are on the antisense strand (Table 1). As noted above, the two highest scoring

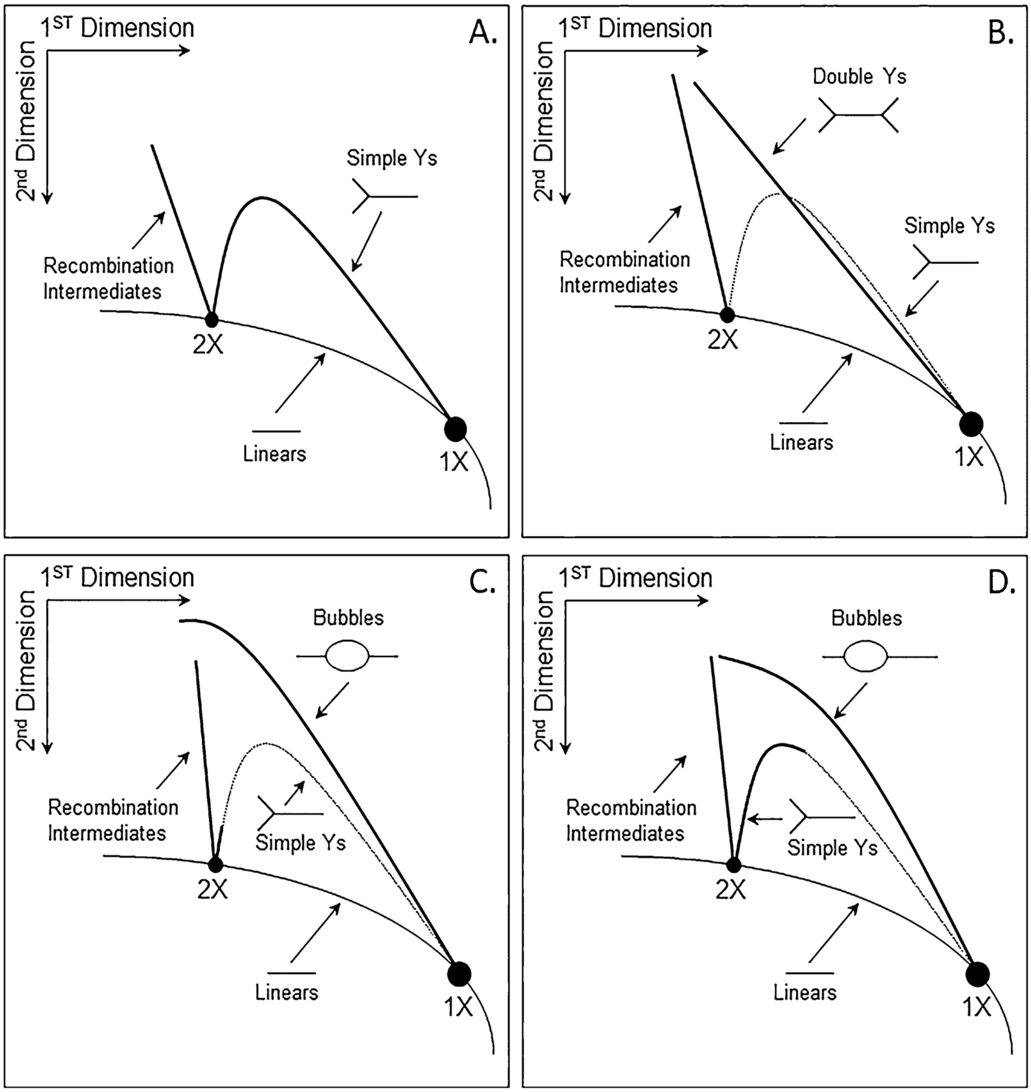

Figure 2. Representation of DNA replication structures derived from N/N two-dimensional agarose electrophoresis. Replicating DNA when subjected to restriction enzyme digestion will produce a variety of structures ranging from 1X (un-replicated) to 2X (nearly fully replicated). **A.** Simple Y structures are produced when an origin of replication lies outside the restriction fragment. **B.** Double Y structures are produced when two replication forks are advancing toward each other within the fragment. **C.** A symmetrically located origin within a fragment produces predominantly bubble structures. **D.** An asymmetrically located origin produces both bubbles and simple Y structures.

pG4s are in a 79 bp direct repeat sequence which has one copy in the *Xho*I fragment.

Discussion

In order to begin studies of the extrachromosomal rRNA-encoding DNA element's (RDE) mechanism of replication, we used a complete molecular clone of the *N. gruberi* (strain EG_B) RDE and its sequence information (Mullican et al. 2018) to localize a single origin of replication (*ori*) within the RDE. While the data support the presence of a single *ori*, the possibility that there may be more than one *ori* in this region cannot be formally ruled out at this time. For example, two loci located closely on either side of the fragment midpoint might be indistinguishable from a single *ori*. However, we consider such a possibility highly unlikely and biologically unnecessary. This work confirmed and significantly extended the strong inference originally voiced by Clark and Cross (1987) that, due to the high copy

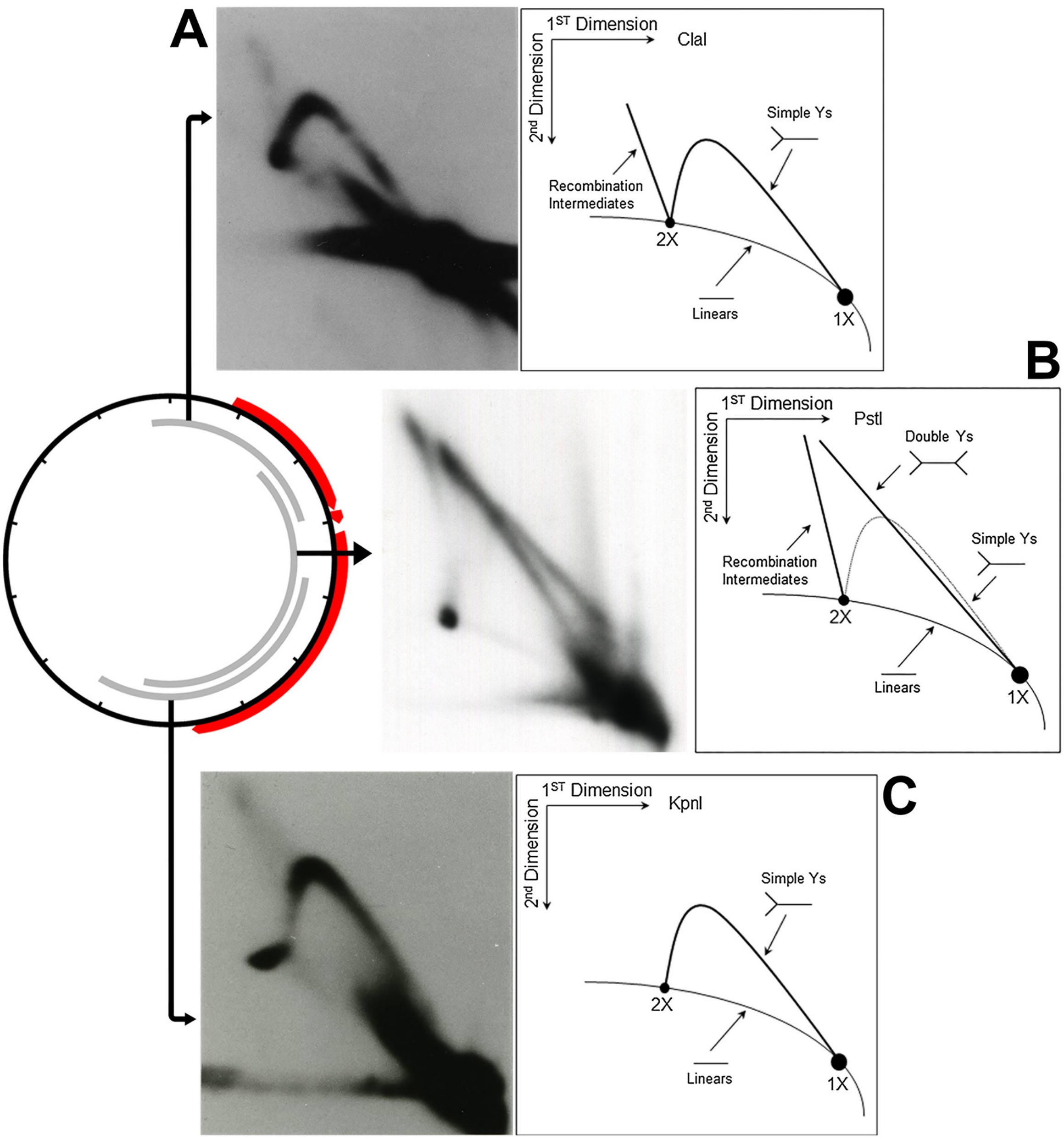

Figure 3. N/N 2-D agarose gel analysis of the region of the rDNA plasmid containing mostly rDNA (nt13712-8217) using three overlapping restriction fragments. **A**. 3.2 kbp *Cla*I fragment, **B**. 5.6 kbp *Pst*I fragment, and **C**. 4.4 kbp *Kpn*I fragment.

number of the RDE per cell, the RDE population must replicate separately from the *Naegleria* chromosomal DNA. The finding of a single *ori* in this relatively small RDE should facilitate future studies of its primary and higher order structures as well as the proteins engaged in replication. Additionally, the numerous species of *Naegleria* provide a group of related, yet differentiable, RDE *ori* sequences with which to pursue such studies.

The sequence encoding the *ori* was localized to within the 2.1 kbp *Xho*I restriction fragment (Fig. 1). Based on the pattern of both bubble and simple Y structures in this sequence (Fig. 4D, Supplementary Material Fig. S3), the data indicate that there is only a single *ori* and it is asymmetrically located within this sequence (Brewer and Fangman 1987). As bubble structures are located at a single site, the RDE utilizes *theta* replication rather than a rolling

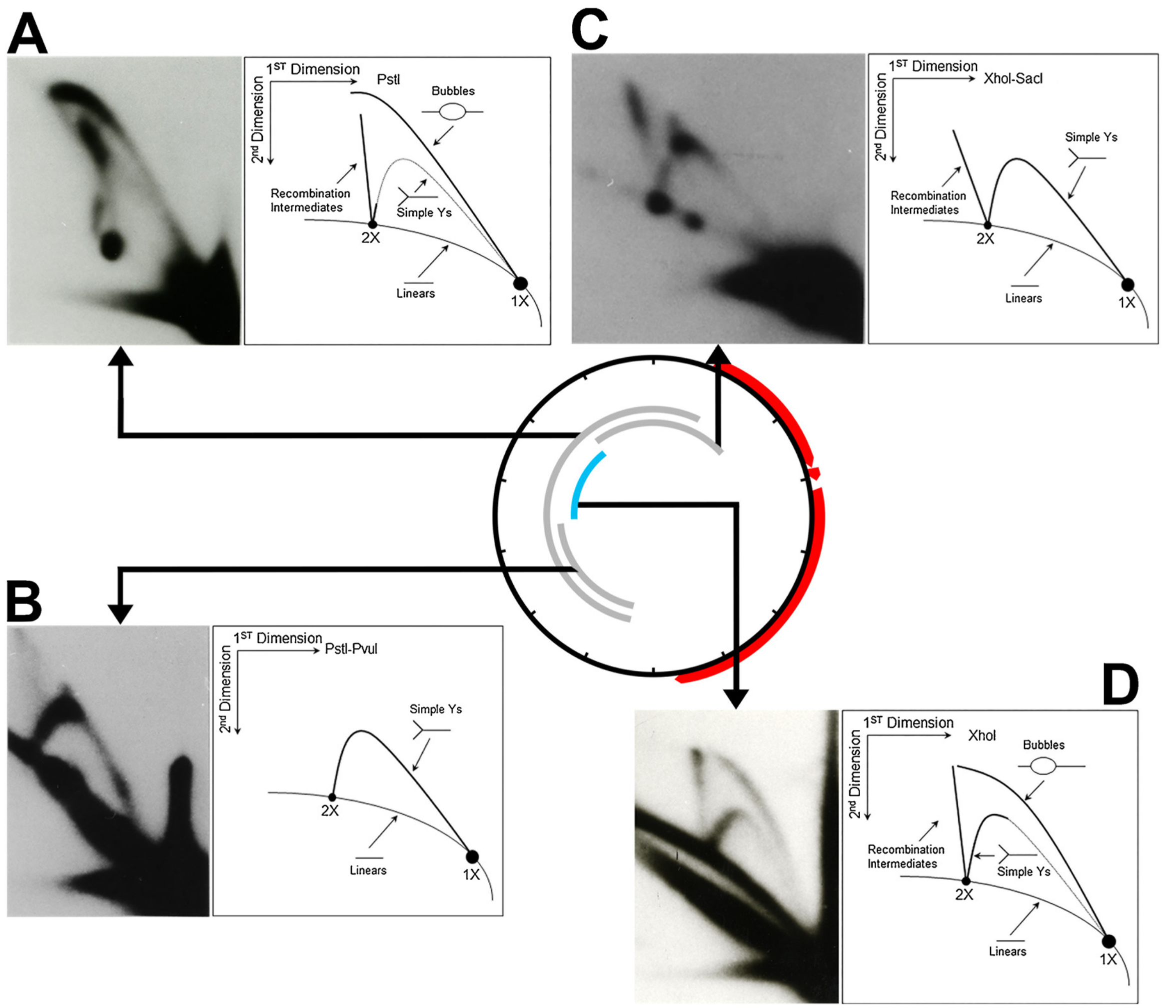

Figure 4. N/N 2-D agarose gel analysis of the remainder of the rDNA plasmid containing mostly non-ribosomal sequence (NRS; nt7435-1788) using four overlapping restriction fragments. **A**. 7.6 kbp *Pst*I fragment, **B**. 3 kbp *Pst*I-*Pvu*I fragment, **C**. 3.3 kbp *Xho*I-*Sac*I fragment, and **D**. 2.1 kbp *Xho*I fragment (in blue).

circle mechanism (Cairns 1963). The replication of the extrachromosomal (but linear) palindromic rDNA elements in *Physarum*, *Dictyostelium* as well as the macronucleus of *Tetrahymena*, all initiate near the center of these molecules within the non-transcribed spacer region located upstream of each rRNA cistron (Cech and Brehm 1981; Cockburn et al. 1978; Ferris 1985). For these RDE, semiconservative replication proceeds bidirectionally through each cistron to the ends of the linear molecule which contain telomere-like repeats involved in the replication and maintenance of these elements (Cech and Brehm 1981; Cockburn et al. 1978; Ferris 1985). Electron microscopy of this type of replication demonstrates the appearance of replication *ori* in bubbles or "eye-shaped" structures, similar to that observed here for the *N. gruberi* RDE, and so this type of replication has been termed *theta*-type (θ) replication.

Both G tetraplex and repeat sequence motifs have been associated with DNA origins of replication. Examination of the 2.1 kbp *Xho*I sequence shows no repeated sequence unique to this region other than two 12 bp direct repeats located in the third of the sequence closest to the 18*S* rDNA (nt11964-11975; nt12081-12092). Further, the ARS element of *S. cerevisiae* (Newlon and Theis 1993) which had been noted in the *N. gruberi* RDE (Mullican et al. 2018), is outside of the *Xho*I sequence. Although G-quadruplexes have been associated with sequences containing human DNA replication *ori* (Besnard et al. 2012), there is no sequence with a high density of high scoring pG4s in the *N. gruberi Xho*I fragment. Only two high scoring pG4s with 3 G-tetrads are present in the RDE and they exist in a 79 bp repeated sequence discussed above (although one of these repeats is present in the *Xho*I fragment). While there is no significant primary structure conservation in the NRS of the RDEs of *N. fowleri* (unpublished) and *N. lovaniensis* [GenBank CM10402; (Liechti et al. 2018)] with *N. gruberi*, it is worth noting that each of these contain pG4 elements scoring higher than 20 [QGRS Mapper; (Kikin et al. 2006)] within a repeated sequence in the non-ribosomal region of the RDE. Scoring (from 0 to 105) in the QGRS Mapper is based on studies of known G-quadruplex structures and allots higher scores to those with shorter loops, to those with loops of more equal size and to those with more G-tetrads (Kikin et al. 2006). These two pG4s (each containing 3 G-tetrads) in *N. gruberi* are separated by 1,874 bp and there are similar pairs of strong pG4s separated by approximately 2,000 bp in both *N. fowleri* and *N. lovaniensis*. While one copy of these pG4s is outside the *Xho*I fragment, it is possible that the two pG4s may bracket the site of the *ori* mapped in the *XhoI* fragment of *N. gruberi*.

Positioning of these pG4s is interesting because they have been hypothesized to be one aspect of the site requirements for mammalian *ori* (Prioleau and MacAlpine 2016). We note, however, that G4 elements have not been proven to be a requisite feature of all *ori;* further, known sites of *ori* in non-metazoan eukaryotes have not been shown to depend upon them (Prioleau 2017). Future research should focus on these and other structures in this region of the NRS in the RDEs of other *Naegleria* species as it seems reasonable that these closely related organisms (all of which have circular RDE) will have some degree of conservation of essential elements and structure of the DNA replication *ori.* Interestingly, our comparison (data not shown) of the sequence of the *Xho*I fragment from *N. gruberi* to the same regions in the RDEs of *N. fowleri* and *N. lovaniensis* has shown no apparent homology, consistent with the findings in other eukaryotes that there appears to be little DNA primary structure conservation in *ori* (Leonard and Mechali 2013). Gel mobility shift and DNase footprinting assays may be able to identify the core *ori* sequence.

Not surprisingly, *N. gruberi* cysts also contain the rDNA-containing element (J. Mullican, unpublished observations). Ribosomal RNA transcription is shut down in cysts of *Acanthamoeba* (Stevens and Pachler 1973), possibly due to an uncharacterized modification of the RNA polymerase (Paule et al. 1984). Our preliminary results using the 2D gel analysis approach with *N. gruberi* RDE DNA isolated from cysts showed the possibility of replication intermediates (J. Mullican, unpublished observations). While it is possible that a low order of RDE replication continues within cysts, these observations suggest an intellectually more pleasing alternative explanation: the RDE replication may be stalled in this cell stage. It is intriguing to consider that *Naegleria* might partially replicate rDNA at the start of, or early in, encystment and then arrest it upstream of the RNA pol I promoter such that upon excystment, this synchronization would provide large numbers of templates to 'jump start' rRNA synthesis. The cluster of pG4 upstream of the 18*S* rDNA could be stabilized by G4 binding proteins to stall replication forks (Hall et al. 2017). DNA enriched by BND-cellulose binding should permit a more thorough examination of this possibility.

The structures noted in the 2D gel analysis of the 5.6 kbp *Pst*I digest (Fig. 3B, Supplementary Material Fig. S2) suggests that there may be a site of convergence of the two replication forks and possibly a replication fork pause as has been noted in many eukaryotes at the site of transcriptional termination of the rRNA (Rothstein et al. 2000). While there is no repeated sequence characteristic of RNA polymerase termination sites which has been suggested to play a role in replication fork pausing as has been noted for other eukaryotes (Akamatsu and Kobayashi 2015), the DNA primary structure used for this process in these *Excavata* eukaryotes might be considerably divergent from species of the other eukaryotic clades.

In summary, this report demonstrates that a single *ori* exists within this eukaryotic circular extrachromosomal element and defines the region within which the site of the *ori* must exist. Finer mapping and comparison of this element with those in other *Naegleria* species will be able to define the essential sequences and/or higher order structures controlling initiation of the extrachromosomal DNA replication. Precise definition of the *ori* site's primary structure and of any relationship of repeated sequences or pG4 elements will rest upon the ability to generate an experimental system for analyzing initiation of replication from such sequences and mutated versions of them.

Methods

Propagation of amoebae: *Naegleria gruberi*, strain EG_B (Schuster and Svihla 1968), amoebae were propagated axenically in stationary tissue culture flasks at 30 °C in a medium consisting of equal parts Nelson's medium (Weik and John 1977a) and Balamuth's medium (Weik and John 1977b) in

Page's ameba saline (Page 1967) supplemented with 4% v/v untreated calf serum.

Isolation of *N. gruberi* DNA: We extracted nucleic acids from saline-washed amoebae by suspending and gently shaking pelleted cells in a rapidly lysing and highly chaotropic buffer (8 M NaSCN, 50 mM Tris–HCl, pH 7.5, 50 mM EDTA [diluted from a 500 mM stock at pH 8.0], 5 mM EGTA, and 142 mM *beta*-mercaptoethanol). The lysate was made 1% v/v in sodium sarkosyl from a 20% w/v stock solution and mixed gently. This was diluted 8-fold with 10 m*M* Tris-HCl, pH 8.3, 10 m*M* EDTA, to which proteinase K powder (Boehringer Mannheim) was subsequently added to a concentration of 200 μg/mL. This mixture was incubated at 50 °C for 2 hours, then overnight at 37 °C. Cell debris was removed by centrifugation (2,500 × *g* at room temperature for 15 minutes) after which purified oyster glycogen (Tracy 1981) was added to 25 μg/mL as carrier. The mixture was then extracted with 1:1 phenol:$CHCl_3$ until the interface was clear, then with $CHCl_3$ to remove residual phenol. One-third volume of 7.5 *M* ammonium acetate was added and the nucleic acids precipitated with 3 volumes of 95% ethanol. Following washing of the pellet with 70% ethanol and resuspension of the nucleic acid pellet in 10 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA, DNase-free RNase was added to 100 μg/mL and incubated for 1 hour at 37 °C. The solution was made 50 m*M* in NaCl and extracted as before. Following ammonium acetate addition as before, DNA was precipitated in ethanol. The DNA pellet was washed with 70% ethanol, dried and resuspended in sterile water, aliquoted and stored at −75 °C until use. DNA concentration was calculated following spectrophotometric measurement using 20 OD (A_{260}) per 1 mg of double-stranded DNA. The A_{260}/A_{280} ratios of DNA were routinely greater than 1.85 and below 2.0.

Cloning and sequence analysis of the *N. gruberi* extrachromosomal element: The molecular cloning of the extrachromosomal element has been described elsewhere (Mullican et al. 2018). Briefly, total *N. gruberi* DNA was digested with *Bst*I (a *Bam*HI isoschizomer) to linearize circular and supercoiled forms of the RDE. Linearized RDE DNA was isolated from agarose gels and ligated into dephosphorylated *Bst*I-digested pGEM-7Zf(+) (Promega Corp., Madison, WI). Colonies were screened for insert-containing DNA using the nick-translated ^{32}P-labeled plasmids pNgex27 (containing the entire ribosomal DNA element of *N. gruberi*, NEG-M) and pNgexNTS (containing the 6.6 kbp *Bst*I/*Pst*I non-ribosomal DNA sequence from pNgex27) (Clark and Cross 1987). Three separate clones containing an insert of 14 kbp were restriction endonuclease mapped and compared. One clone was used for complete sequence analysis (termed pNgrubEGB). Various restriction fragments were subcloned for ease of sequence determination and for use as hybridization probes as needed during origin mapping. The complete sequence of the molecularly cloned *N. gruberi*, EG_B RDE (pNgrubEGB) was derived and has been described and annotated elsewhere [(Mullican et al. 2018); GenBank Accession MG699123]. pNgrubEGB plasmid DNA is available upon request for non-commercial, research purposes.

N/N two-dimensional agarose gel analysis: We used the technique described by (Brewer and Fangman 1987). Briefly, this involves the electrophoretic separation of restriction endonuclease cleaved DNA in two dimensions through neutral agarose gels. The first dimension was run in either 0.4% (if the DNA fragment was larger than 3 kbp) or 0.5-0.7% agarose in 1X TBE buffer for smaller fragments. The second dimension was run in 1.0-1.2% agarose in 1X TBE buffer; smaller fragments could be run in 1.5% agarose. To suppress background, we enriched for replication intermediates using benzoylated naphthoylated DEAE-cellulose (BND-cellulose; Sigma-Aldrich, St. Louis MO) which efficiently binds ssDNA, RNA and dsDNA that contains single-stranded (ss) regions when the column is eluted in 1 *M* NaCl; dsDNA lacking ss regions bind weakly. DNA with single-stranded character was eluted from the column in 1 *M* NaCl, 1.8-2.0% caffeine. BND-cellulose chromatography was employed following methodology published previously (Dijkwel et al. 1991). DNA was digested with the appropriate restriction endonuclease, then loaded into gels and electrophoresed in the first dimension. Following staining with ethidium bromide and brief UV illumination, gel slices of interest were embedded within the second-dimension gel when it was poured. Second dimension gels were then Southern blotted, probed with (gamma)-^{32}P-dCTP-labelled restriction fragments and exposed to X-ray film.

Mapping of potential G-quadruplexes: Potential G-quadruplex (pG4) sequences on both strands of the extrachromosomal DNA were identified using the Quadruplex forming G-Rich Sequences (QGRS) Mapper (Kikin et al. 2006). The QGRS Mapper search was set to maximum sequence length of 30, minimum G-tetrads at 2 and loop size of 1-30. The QGRS Mapper generates a score for each pG4 based on shorter loops being more common; approximately equal loop size and the stability of the structure increases with the greater the number of G-tetrads in the structure (Kikin et al. 2006). To find the most likely pG4, the minimum score was set to 20.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

We wish to thank David John for kindly providing the *N. gruberi*, EG_B strain and C. Graham Clark for kindly providing the pNgex27 and pNgexNTS plasmids. We also thank Bonny Brewer for helpful discussions regarding 2D-gel setup and analyses.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.protis.2019.02.001.

References

Ahmed Khan N, Baqir H, Siddiqui R (2015) The immortal amoeba: a useful model to study cellular differentiation processes? Pathog Glob Health **109**:305–306

Akamatsu Y, Kobayashi T (2015) The human RNA polymerase I transcription terminator complex acts as a replication fork barrier that coordinates the progress of replication with rRNA transcription activity. Mol Cell Biol **35**:1871–1881

Bartsch I, Schoneberg C, Grummt I (1987) Evolutionary changes of sequences and factors that direct transcription termination of human and mouse ribsomal genes. Mol Cell Biol **7**:2521–2529

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. Nat Struct Mol Biol **19**:837–844

Bhattacharya S, Som I, Bhattacharya A (1998) The ribosomal DNA plasmids of *Entamoeba*. Parasitol Today **14**:181–185

Bhattacharya S, Bhattacharya A, Diamond LS, Soldo AT (1989) Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA. J Protozool **36**:455–458

Brewer BJ, Fangman WL (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. Cell **51**:463–471

Cairns J (1963) The bacterial chromosome and its manner of replication as seen by autoradiography. J Mol Biol **6**: 208–213

Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, Mechali M (2012) New insights into replication origin characteristics in metazoans. Cell Cycle **11**:658–667

Cech TR, Brehm SL (1981) Replication of the extrachromosomal ribosomal RNA genes of *Tetrahymena thermophilia*. Nucleic Acids Res **9**:3531–3543

Clark CG (1990) Genome structure and evolution of *Naegleria* and its relatives. J Protozool **37**:2s–6s

Clark CG, Cross GA (1987) rRNA genes of *Naegleria gruberi* are carried exclusively on a 14-kilobase-pair plasmid. Mol Cell Biol **7**:3027–3031

Clark CG, Cross GA (1988) Circular ribosomal RNA genes are a general feature of schizopyrenid amoebae. J Protozool **35**:326–329

Cockburn AF, Taylor WC, Firtel RA (1978) *Dictyostelium* rDNA consists of non-chromosomal palindromic dimers containing 5S and 36S coding regions. Chromosoma **70**: 19–29

De Jonckheere JF (2014) What do we know by now about the genus *Naegleria*? Exp Parasitol **145(Suppl)**:S2–S9

Denet E, Coupat-Goutaland B, Nazaret S, Pelandakis M, Favre-Bonte S (2017) Diversity of free-living amoebae in soils and their associated human opportunistic bacteria. Parasitol Res **116**:3151–3162

Deshpande AM, Newlon CS (1992) The ARS consensus sequence is required for chromosomal origin function in *Saccharomyces cerevisiae*. Mol Cell Biol **12**:4305–4313

Dijkwel PA, Vaughn JP, Hamlin JL (1991) Mapping of replication initiation sites in mammalian genomes by two-dimensional gel analysis: stabilization and enrichment of replication intermediates by isolation on the nuclear matrix. Mol Cell Biol **11**:3850–3859

Dolinnaya NG, Ogloblina AM, Yakubovskaya MG (2016) Structure, properties, and biological relevance of the DNA and RNA G-Quadruplexes: overview 50 years after their discovery. Biochemistry (Mosc) **81**:1602–1649

Ferris PJ (1985) Nucleotide sequence of the central non-transcribed spacer region of *Physarum polycephalum* rDNA. Gene **39**:203–211

Fritz-Laylin LK, Prochnik SE, Ginger ML, Dacks JB, Carpenter ML, Field MC, Kuo A, Paredez A, Chapman J, Pham J, Shu S, Neupane R, Cipriano M, Mancuso J, Tu H, Salamov A, Lindquist E, Shapiro H, Lucas S, Grigoriev IV, Cande WZ, Fulton C, Rokhsar DS, Dawson SC (2010) The genome of *Naegleria gruberi* illuminates early eukaryotic versatility. Cell **140**:631–642

Fulton C, Guerrini AM (1969) Mitotic synchrony in *Naegleria amebae*. Exp Cell Res **56**:194–200

Grace E, Asbill S, Virga K (2015) *Naegleria fowleri*: pathogenesis, diagnosis, and treatment options. Antimicrob Agents Chemother **59**:6677–6681

Grummt I, Maier U, Ohrlein A, Hassouna N, Bachellerie JP (1985) Transcription of mouse rDNA terminates downstream of the 3′ end of 28S RNA and involves interaction of factors with repeated sequences in the 3′ spacer. Cell **43**: 801–810

Hall AC, Ostrowski LA, Pietrobon V, Mekhail K (2017) Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. Nucleus **8**:162–181

Hoogsteen K (1963) The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. Acta Crystallogr **16**:907–916

Hyrien O, Mechali M (1992) Plasmid replication in *Xenopus* eggs and egg extracts: a 2D gel electrophoretic analysis. Nucleic Acids Res **20**:1463–1469

Kikin O, D'Antonio L, Bagga PS (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res **34**:W676–W682

Leonard AC, Mechali M (2013) DNA replication origins. Cold Spring Harb Perspect Biol **5**:a010116

Liechti N, Schurch N, Bruggmann R, Wittwer M (2018) The genome of *Naegleria lovaniensis*, the basis for a comparative approach to unravel pathogenicity factors of the human pathogenic amoeba *N. fowleri*. BMC Genomics **19**:654

Little RD, Platt TH, Schildkraut CL (1993) Initiation and termination of DNA replication in human rRNA genes. Mol Cell Biol **13**:6600–6613

Long EO, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem **49**:727–764

Maizels N (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat Struct Mol Biol **13**:1055–1059

Maruyama S, Nozaki H (2007) Sequence and intranuclear location of the extrachromosomal rDNA plasmid of the amoebo-flagellate *Naegleria gruberi*. J Eukaryot Microbiol **54**:333–337

Mullican JC, Chapman NM, Tracy S (2018) Complete genome sequence of the circular extrachromosomal element of *Naegleria gruberi* strain EGB ribosomal DNA. Genome Announc **6**:e00020–00018

Newlon CS, Theis JF (1993) The structure and function of yeast ARS elements. Current Opin Genet Dev **3**:752–758

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**:a010108

Opperdoes FR, De Jonckheere JF, Tielens AG (2011) *Naegleria gruberi* metabolism. Int J Parasitol **41**:915–924

Page FC (1967) Taxonomic criteria for limax amoebae, with descriptions of 3 new species of *Hartmannella* and 3 of *Vahlkampfia*. J Protozool **14**:499–521

Parker MW, Botchan MR, Berger JM (2017) Mechanisms and regulation of DNA replication initiation in eukaryotes. Crit Rev Biochem Mol Biol **52**:107–144

Paule MR, Iida CT, Perna PJ, Harris GH, Brown Shimer SL, Kownin P (1984) Faithful initiation of ribosomal RNA transcription from cloned DNA by purified RNA polymerase I. Biochemistry **23**:4167–4172

Peng C, Luo H, Zhang X, Gao F (2015) Recent advances in the genome-wide study of DNA replication origins in yeast. Front Microbiol **6**:117

Prioleau MN (2017) G-Quadruplexes and DNA replication origins. Adv Exp Med Biol **1042**:273–286

Prioleau MN, MacAlpine DM (2016) DNA replication origins—where do we begin? Genes Dev **30**:1683–1697

Robinson NP, Dionne I, Lundgren M, Marsh VL, Bernander R, Bell SD (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell **116**:25–38

Rodriguez-Zaragoza S (1994) Ecology of free-living amoebae. Crit Rev Microbiol **20**:225–241

Rothstein R, Michel B, Gangloff S (2000) Replication fork pausing and recombination or "gimme a break". Genes Dev **14**:1–10

Schuster FL, Svihla G (1968) Ribonucleoprotein-containing vesicles in cysts of *Naegleria gruberi*. J Protozool **15**:752–758

Sen D, Gilbert W (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature **334**:364–366

Stevens AR, Pachler PF (1973) RNA synthesis and turnover during density-inhibited growth and encystment of *Acanthamoeba castellanii*. J Cell Biol **57**:525–537

Torres-Machorro AL, Hernandez R, Cevallos AM, Lopez-Villasenor I (2010) Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? FEMS Microbiol Rev **34**:59–86

Tracy S (1981) Improved rapid methodology for the isolation of nucleic acids from agarose gels. Prep Biochem **11**:251–268

Tyml T, Skulinova K, Kavan J, Ditrich O, Kostka M, Dykova I (2016) Heterolobosean amoebae from Arctic and Antarctic extremes: 18 novel strains of *Allovahlkampfia*, *Vahlkampfia* and *Naegleria*. Europ J Protistol **56**:119–133

Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintome C, Riou JF, Prioleau MN (2014) G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J **33**:732–746

Weik RR, John DT (1977a) Agitated mass cultivation of *Naegleria fowleri*. J Parasitol **63**:868–871

Weik RR, John DT (1977b) Cell size, macromolecular composition, and O2 consumption during agitated cultivation of *Naegleria gruberi*. J Protozool **24**:196–200

Wikmark OG, Einvik C, De Jonckheere JF, Johansen SD (2006) Short-term sequence evolution and vertical inheritance of the *Naegleria* twin-ribozyme group I intron. BMC Evol Biol **6**:39