UNIVERSITY OF CALIFORNIA
RIVERSIDE

Replication and Recombination of Nematode Mitochondrial DNA

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics and Bioinformatics

by

Samantha Cleopatra Lewis

December 2013

Dissertation Committee:
Dr. Bradley C. Hyman, Chairperson
Dr. Dmitri Maslov
Dr. Thomas Girke

The Dissertation of Samantha Cleopatra Lewis is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

Above all I must thank my advisor, Dr. Bradley C. Hyman. Without his patient guidance, none of this work would have been possible.

Next, I thank my collaborators, Dr. Howard T. Jacobs of the Finnish Centre of Excellence in Mitochondrial Disease and Ageing Research, at the University of Tampere in Finland, and Dr. Jack Griffith, of the Lineberger Cancer Center, at the University of North Carolina, Chapel Hill.

Finally, I acknowledge Tyler WH Backman for his constant support and encouragement throughout my graduate school years.

ABSTRACT OF THE DISSERTATION

Replication and Recombination of Nematode Mitochondrial DNA

by

Samantha Cleopatra Lewis

Doctor of Philosophy, Graduate Program in Genetics, Genomics and Bioinformatics
University of California, Riverside, December 2013
Dr. Bradley C. Hyman, Chairperson

Eukaryotic cells are divided into many compartments, enabling spatial and temporal regulation in higher cells. Mitochondria are unique among these compartments in animal cells, as they harbor a DNA genome separate from the nuclear genome, the mitochondrial DNA (mtDNA). Defects in the replication, transcription, translation or assembly of mtDNA encoded subunits cause severe diseases in humans. Deleterious mtDNA mutations may be introduced by several routes, particularly replication error. A better understanding of the mechanism of mtDNA replication is required to generate hypotheses as to how to remedy the introduction of mtDNA mutations as a possible avenue to reduce the incidence of mitochondrial myopathies. To further understand the mechanism of mtDNA replication as well as the spectrum of mtDNA mutations present across eukaryotic species, I initiated computational and biochemical studies addressing mtDNA replication processes using the genetic nematode model, *Caenorhabditis*

elegans. I used a computational approach to analyze transition and transversion mutations which alter the distribution of nucleotides across mtDNA strands. The development of custom scripts allowed an expansion of this work to include all metazoan mitochondrial genomes publicly available. I identified lineage-specific variation in mtDNA mutation patterning that suggested differences in the mechanism of mtDNA replication between animal phyla. Furthermore, I examined mtDNA topoisomers isolated from the nematode *C. elegans*. A detailed analysis by two-dimensional agarose gel electrophoresis and transmission electron microscopy revealed that the mtDNA of *C. elegans* is not replicated by these mechanisms, but rather rolling circle replication, a fourth mechanism novel among animals. My results indicate that generalizations regarding mtDNA replication modes cannot be derived by studying model organisms alone. Multiple mtDNA replication mechanisms have evolved to protect replicating mtDNA from a vulnerable, single-stranded state which is thought to be highly mutable.

Table of Contents

List of Tables

List of Figures

Chapter 1. General Introduction

Eukaryotic cells are divided into many compartments, enabling spatial and temporal separation of biochemical processes, an architecture that contributes to the exquisite regulatory abilities of higher cells. Mitochondria, double membrane-bound organelles, are one such compartment (Lang et al. 1999). While colloquially known as the “powerhouses of the cell” due to their central role in the production of adenosine tri-phosphate by oxidative phosphorylation, mitochondria are in fact critical to a multitude of cellular processes including iron-sulfur cluster production, calcium buffering, inter-organelle signaling and cell cycle regulation (Nunnari and Suomalainen 2012). Mitochondria are unique among organelles of animal cells in that they harbor a DNA genome separate from the nuclear genome, termed the mitochondrial DNA (mtDNA)(Falkenberg et al. 2007).

MtDNA encodes 12-13 protein subunits essential to the assembly and operation of the respiratory complexes that reside in the inner mitochondrial membrane, as well as the transfer RNAs and ribosomal RNAs necessary for their translation (Lang et al. 1999). As such, defects in the replication, transcription, translation or assembly of mtDNA encoded subunits cause severe diseases in humans. MtDNA defects that reduce respiratory capacity, such as the family of maladies known as mitochondrial myopathies, disproportionately impact tissues with high energy demand (Falkenberg et al. 2007). These include skeletal muscle, brain and heart (Le Chen and Knowlton 2011).

Deleterious mtDNA mutations may be introduced by several routes, such as inefficient DNA repair, oxidative damage, recombination events, or replication error (Bratic and Trifunovic 2010). Early studies suggested that spontaneous deamination of nucleotides as a result of interaction with reactive oxygen species present in the mitochondrial matrix were a primary cause of mtDNA mutation (Rocha 2004; Hassanin et al. 2005). However, mutation accumulation studies which directly measured the number of spontaneous mutations per generation in a variety of model organisms have revealed that spontaneous point mutations in mtDNA occur at a low rate; too low, in fact, to explain the spectrum of mutations observed in natural populations (Montooth and Rand 2008). In recent years, mitochondrial disease genes have been identified, and many are components of the mammalian mtDNA replisome, including the dedicated mtDNA polymerase POLG (Di Re et al. 2009). These discoveries have led to a new consensus in the field that mtDNA replication errors are the primary cause of disease-causing mutations in humans (Nunnari and Suomalainen 2012).

A better understanding of the mechanism of mtDNA replication is required to generate hypotheses as to how to remedy the introduction of mtDNA mutations as a possible avenue to reduce the incidence of mitochondrial myopathies. Unfortunately the mechanism of mtDNA replication, and thus mtDNA mutation, remains poorly understood (Pohjoismäki and Goffart 2011).

Neutral mutations in mtDNA accumulate over time, and may provide a measure of the types and frequency of mtDNA mutation in various eukaryotic lineages(Castellana et al. 2011). Additionally, direct biochemical studies of mtDNA replication are needed to ascertain the key features of the mtDNA replication process, and how they are perturbed to produce disease states (Montooth and Rand 2008; Pohjoismäki and Goffart 2011).

To further understand the mechanism of mtDNA replication as well as the spectrum of mtDNA mutations present across eukaryotic species, I initiated computational and biochemical studies addressing mtDNA replication processes using the model genetic nematode model, *Caenorhabditis elegans*. I analyzed transition and transversion mutations which alter the distribution of nucleotides across mtDNA strands by a computational approach (Chapter 2). The development of custom scripts allowed a complete survey of the over 2400 sequenced mitochondrial genomes archived in the National Center for Biotechnology Information GenBank database. This high-throughput approach enabled the identification of lineage-specific differences in mtDNA mutation patterning that suggested differences in the mechanism of mtDNA replication between animal phyla.

Using a biochemical approach (Chapter 3), I examined mtDNA topoisomers isolated from live *C. elegans* and compared their structure and biochemical properties to those predicted by the three mechanisms of mtDNA

replication previously proposed in the literature. Astonishingly, a detailed analysis by two-dimensional agarose gel electrophoresis and transmission electron microscopy (Chapter 4) revealed that the mtDNA of *C. elegans* does is not replicated by any of these mechanisms, but rather rolling circle replication, a fourth mechanism novel among animals.

My results from Chapters 1-4 in aggregate indicate that generalizations regarding mtDNA replication modes cannot be derived by studying a limited number of species, often defined as model organisms. Moreover, it appears that several mtDNA mechanisms have evolved to protect replicating mtDNA from a vulnerable, single-stranded state which is thought to be highly mutable. Understanding the breadth of mtDNA replication mechanism and the protein machinery responsible for mtDNA replication mechanism may provide useful therapeutic targets for treatment of mitochondrial myopathies.

References

Bratic I, Trifunovic A. 2010. Mitochondrial energy metabolism and ageing. Biochim. Biophys. Acta 1797:961–967.

Castellana S, Vicario S, Saccone C. 2011. Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. Genome Biology and Evolution.

Di Re M, Sembongi H, He J, et al. 2009. The accessory subunit of mitochondrial DNA polymerase gamma determines the DNA content of mitochondrial nucleoids in human cultured cells. Nucleic Acids Research 37:5701–5713.

Falkenberg M, Larsson N-G, Gustafsson CM. 2007. DNA Replication and Transcription in Mammalian Mitochondria. Annu. Rev. Biochem. 76:679–699.

Hassanin A, Léger N, Deutsch J. 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. Systematic Biology 54:277–298.

Lang BF, Gray MW, Burger G. 1999. Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet.

Le Chen, Knowlton AA. 2011. Mitochondrial Dynamics in Heart Failure. Congestive Heart Failure 17:257–261.

Montooth KL, Rand DM. 2008. The spectrum of mitochondrial mutation differs across species. PLoS Biol 6:e213.

Nunnari J, Suomalainen A. 2012. Mitochondria: in sickness and in health. Cell 148:1145–1159.

Pohjoismäki JLO, Goffart S. 2011. Of circles, forks and humanity: Topological organisation and replication of mammalian mitochondrial DNA. Bioessays 33:290–299.

Rocha EPC. 2004. The replication-related organization of bacterial genomes. Microbiology 150:1609–1627.

Chapter 2. Comparative analysis of nucleotide compositional asymmetry within plant, fungal and animal mitochondrial genomes reveals lineage-specific patterns of substitution

Introduction

The architecture and evolution of mitochondrial DNA (mtDNA) is influenced by multiple factors, including DNA replication, transcription, and as recent works have demonstrated, recombination (Gissi et al. 2008; Bernt et al. 2012). These processes coordinately determine nucleotide composition within the mitogenome as well as the distribution of nucleotides between the two mtDNA strands, though the molecular mechanism(s) by which mutational gradients are produced remain controversial (Rocha et al. 2006; Seligmann 2012).

Strand asymmetry, defined as the violation of Chargaff's second parity rule (Nikolaou and Almirantis 2006), is a ubiquitous mtDNA feature of particular interest, as divergent mutational constraints significantly impact the utility of mtDNA sequences for phylogenetics (Hassanin 2006). Perna and Kocher (1995) formalized the quantification of strand asymmetries in a value known as *skew*, whereby GC and AT asymmetries are defined as follows:

$S_{GC} = (G-C)/(G+C)$ $S_{AT} = (A-T)/(A+T)$

Numerous studies have documented negative S_{GC} and positive S_{AT} values among mtDNAs; these studies often focus on mammals and arthropods, two well sampled metazoan lineages (Hassanin et al. 2005; Fonseca et al. 2008; Wei, Shi, Chen, et al. 2010; Wei, Shi, Sharkey, et al. 2010). For those taxa, "reversed"

strand asymmetries (positive S_{GC} and negative S_{AT} values) associate with the presence of codon usage biases and divergent nucleotide composition at informative fourfold-degenerate sites, and contribute to long-branch attraction artifacts in phylogenetic reconstruction (Fonseca et al. 2008). Thus, the characterization of strand asymmetry patterns among a wider sampling of eukaryotes, and subsequent identification of species and/or taxa harboring atypical strand asymmetries, is required to facilitate robust phylogenetic inferences. Yet, studies so far have failed to take advantage of the thousands of mtDNA sequences publicly available, which more fully represent the expanse of the Eukaryota, in systematic analyses of compositional asymmetries.

Multiple forces, including purifying selection, random genetic drift, and DNA repair mechanisms may influence nucleotide composition in mitochondrial genomes (Rand 2001; Montooth and Rand 2008; Stewart et al. 2008) including the occurance of strand asymmetry . As such, the spectrum of mutations observed in eukaryotic mtDNA is in part defined by the balance of selection and mutation forces. That both the strength of purifying selection and the rate of spontaneous mtDNA mutations are known to vary among independent eukaryotic lineages (Denver et al. 2000; Haag-Liautard et al. 2008; Montooth and Rand 2008), nucleotide composition and strand asymmetry could be expected to vary in a lineage specific fashion.

Beyond mtDNA, the presence and patterns of strand asymmetries have been investigated in chloroplast DNAs and in the nuclear genomes of animals and fungi (Nikolaou and Almirantis 2006; Huvet et al. 2007; Agier and Fischer 2012). However, compositional skews are best characterized within the context of prokaryotic genomes (Rocha 2004; Necşulea and Lobry 2007; Touchon and Rocha 2008). Like many mtDNAs, bacterial genomes are typically biased towards G and T on the leading strand of replication (Necşulea and Lobry 2007; Gissi et al. 2008; Bernt et al. 2012); the sign of skew values for coding genes is highly correlated with presence on the leading (+) or the lagging (-) strand of replicating mtDNA (Rocha 2004). Furthermore, bacterial genome skews are highly influenced by proximity to a replication origin and often undergo inversion of the direction of skew at replication initiation sites (Rocha et al. 2006; Necşulea and Lobry 2007; Sernova and Gelfand 2008; Seligmann 2012). The prokaryotic skew literature has provided a framework for mtDNA studies, and it is often assumed that mutational asymmetries in mtDNA will recapitulate the patterns observed in bacterial genomic DNA (Lin et al. 2012), though mtDNA replication in many taxa remains poorly understood.

As non-mammalian, non-arthropod mtDNAs are poorly sampled within previous skew studies, it is not clear whether the presence of a gene on either strand, or proximity to an origin, impacts mtDNA compositional asymmetries in most eukaryotes. Mutational asymmetries documented within the mammals or

arthropods may not reflect the status of metazoans or, indeed, eukaryotes as a whole.

To address these questions and place previous studies in a larger context, we applied computational approaches to analyze S_{GC} and S_{AT} among 2,428 mitogenomes available from NCBI Genbank representing the plant, fungal and animal kingdoms. We find that robust strand asymmetries are a signature of metazoan, but not plant or fungal mtDNAs. Mitochondrial genome S_{GC} and S_{AT} values are not normally distributed, and limits on the natural extent of nucleotide distribution biases occur independently of gene transcriptional polarity across the two DNA strands or proximity to non-coding regions presumed to harbor replication origins. Furthermore, we document reversed mutational constraints for over 300 animal mtDNAs, including the vast majority of nematode and platyhelminth species and we describe novel patterns of mutation in those lineages inconsistent with any previously known mtDNA replication mechanism. Finally, an in-depth analysis of compositional asymmetries at phylogenetically informative sites within nematode mtDNAs is presented; this phylum is strikingly neglected in the literature, and we discuss implications for nematode mtDNA evolution and maintenance.

Results

Comparison of mtDNA Nucleotide Compositional Asymmetry among the Animal, Plant and Fungal Kingdoms.

Both GC and AT strand asymmetries are significantly stronger in animal mitogenomes relative to plant or fungal mitogenomes. Compositional asymmetries were present within all mtDNAs sampled (Figure 1a); skew values for all mtDNAs studied here are available in Supplemental Table 1. As shown in Figures 1b and c, the degree of deviation from intra-strand parity varied among the eukaryotic kingdoms, as median AT and GC skews of Metazoan mtDNAs were two orders of magnitude stronger than values for the Viridiplantae (we include both land plants and green algae under this term), and ten times stronger than values for Fungi (Table 1, Supplemental Table 2). Plant and fungal mtDNA skew datasets each exhibited a tight range with infrequent outlier bias values; in contrast, animals exhibited a broad range of S_{AT} and S_{GC} values. Just over 87% of animal mtDNAs were characterized by a compositional bias favoring A and C nucleotides on the major strand, broadly consistent with previous studies on taxon-specific subsets of the metazoa (Hassanin 2006; Fonseca et al. 2008). The significantly lower AT and GC skew values for plants and fungi is consistent with the low spontaneous mutation rates reported for mtDNA within those lineages (Perna and Kocher 1995; Christensen 2013). Interestingly, chloroplast genomes

also tend to lack significant strand asymmetries (Nikolaou and Almirantis 2006), highlighting the presence of compositional asymmetries in the animal mtDNAs as specific to the lineage, and not the organelle itself.

We found that plant and fungal mtDNAs are characterized by the opposite biases relative to the animals, as T and G nucleotides are in excess on the primary strand (Table 1). This comparison suggests a fundamental difference in the processes leading to the generation and/or accumulation of mtDNA mutations that shape nucleotide composition between these lineages. It is well established that plant mtDNAs are larger in size than mtDNAs of animals or fungi; although plant mtDNAs contain a greater number of coding genes relative to animal mtDNAs, the major difference in mtDNA sizes is primarily due to the presence of lengthy non-coding, possibly intronic sequence (Bendich 1993; Backert et al. 1997). Non-coding mtDNA is typically absent from animal mtDNAs (Fonseca et al. 2008; Gissi et al. 2008), raising the possibility that repetitive non-coding sequences contribute to mutational asymmetry.

While recombination and gene conversion are ostensibly rare within animal mitochondria, the frequent occurrence of these processes within both plants and fungal mitochondria is extensively documented, impacting both the functional organization and maintenance of these mitogenomes (Bendich 1993; Backert et al. 1997; Necşulea and Lobry 2007; Touchon and Rocha 2008).

Thus, sequence homogenization by recurrent gene conversion and repair is one possible explanation for the comparative weakness of compositional asymmetries within plant and fungal mtDNAs relative to the metazoa. This hypothesis would suggest that mtDNA recombination contributes to the even distribution of nucleotides across the two DNA strands, and that recombination may be active in the mitochondria of animal species that lack significant strand asymmetries.

Characterization of Strand Asymmetry within Eumetazoan Mitochondrial Genomes.

Non-normal distribution of strand asymmetry values suggests natural limits on mutation processes driving compositional biases. We next investigated the range and trends in GC and AT compositional biases within metazoa. Inspection of the distribution of S_{GC} and S_{AT} among metazoan mtDNAs revealed that neither GC nor AT biases are normally distributed but instead are skewed in the statistical sense (Figure 2a). The distribution of GC bias values was right-tailed (positive), while AT bias values were left-tailed (negative). The tailing of either dataset cannot be attributed to a few extreme outliers, as the range is continuous. Among metazoans, GC and AT skew values were highly correlated by both Pearson and Spearman measures (Table 1), consistent with previous reports with more limited taxonomic sampling (Fonseca et al. 2008;

Wei, Shi, Chen, et al. 2010). As we have described, the majority of mitogenomes were biased in favor of A and C nucleotides. However, 13% of mitochondrial genomes were characterized by a preference for A and G nucleotides on the primary strand, including species representing Nematoda, Platyhelminthes, and Mollusca (discussed in detail below), taxa for which mtDNA nucleotide composition had not been previously evaluated in depth. Mitogenomes with biases towards A and G nucleotides on the primary strand have been defined within the literature as having undergone “reversal of mutational constraints” (Hassanin et al. 2005). We continue to employ this term here to enable comparisons with previous studies.

Regardless of the direction of compositional asymmetry, we next addressed the relative strength of skews across the range of animal mtDNAs. To facilitate comparison between taxa, we determined the frequency of mtDNAs whose GC or AT skew values differed statistically from median values. Because strand asymmetry data for the metazoa is not normally distributed, we use inter-quartile range as an alternative to standard deviation (reported in Supplemental File 1). Approximately 30% of GC skew values fell outside the inter-quartile range; these mitogenomes typically contained weaker biases as compared to the median (Figure 2a, top panel). Similarly, 25% of AT skew values fell outside the inter-quartile range (Figure 2a, bottom panel). These results indicate that while a maximum cap is imposed on both S_{GC} and S_{AT} within animal mtDNAs, there is no

apparent minimum to either value. We suggest that this finding is consistent with presence of a 'ceiling' on nucleotide asymmetry beyond which deleterious effects can occur. The statistical skewness of the S_{GC} and S_{AT} distributions plotted in Figure 2a would then suggest no deleterious consequences for mtDNAs lacking strong compositional biases. This interpretation challenges the assumption that nucleic acid maintenance processes drive compositional asymmetries to a minimum level within all metazoan mitochondria (Castellana et al. 2011). Our finding that median GC and AT bias values approximate the maximum values observed suggests that animal mtDNAs are typically near saturation for these types of mutations.

Strand asymmetry varies significantly between animal phyla. To investigate variation in the strength of compositional asymmetry among animals, we compared AT and GC skew within and between animal phyla. We reasoned that if GC and AT biases are introduced into mtDNA by the same or similar mutational processes, these values should be highly correlated for all animal mtDNAs. The strength of correlation between GC and AT skew values varied significantly between animal phyla as did median skew values (Figure 2b-c; Table 1). Notably, Mollusks displayed a near perfect correlation between GC and AT biases while no correlation was found within amphibian mtDNAs, defined as an absence of a statistically significant difference between the correlation

coefficient in a comparison of matched amphibian GC and AT skews with a set of GC and AT skews paired at random.

Chordates were highly sampled relative to all other animal taxa, yet their compositional asymmetry values clustered tightly, signifying low variation in strand asymmetry within these taxa (Figure 2d). Chordates tended to have a stronger C over G bias than non-chordate species, though the preference for A over T was no different relative to other animals. To investigate possible substructure within the chordate cluster, we compared strand asymmetry between the Mammalia, Aves and Amphibia, three classes each represented by over 100 complete mitogenome sequences in NCBI GenBank. Avian mtDNAs exhibited a strong GC bias and low variance about the median GC skew value, whereas mammals displayed a broad range of GC biases; amphibian GC skews were weaker than either birds or mammals (Figure 2e). Thus amphibian mtDNAs demonstrate that S_{GC} and S_{AT} biases can result from autonomous molecular processes, despite their apparent statistical correlation and among animal mitogenomes overall.

Reversals of compositional bias patterns*.* We identified 306 animal species with positive GC skews and negative AT skew values, considered indicative of a reversal in the pattern of mutation bias (Supplemental Table 3). We report the first cases of compositional bias “reversal” within three phyla

(Mollusca, Platyhelminthes, Nematoda), as well as the first non-mammalian bias "reversals" among the Chordata (Actinopterygii; Supplemental Table 3). In itself, this finding demonstrates that preference for G over C and T over A on the major transcribed strand, once thought to be uncommon, is present within the animal lineage at a higher frequency than previous studies suggest.

Reversals of mutational constraint had been identified within multiple animal taxa and are particularly well characterized within mammals and arthropods (Hassanin et al. 2005); for these taxa, mapping bias "reversals" onto the relevant phylogenies revealed polyphyly coincident with mitochondrial genome rearrangement. These findings resulted in the characterization of "reversals" as random events, perhaps related to rare instances of mtDNA recombination (Fonseca et al. 2008). Given these published predictions, we expected that if "reversals" truly are spontaneous in nature, they would be rare within each taxon, and in all cases be homoplasious. To test whether these hypotheses based on vertebrate and arthropod data are applicable to reversals in general, we analyzed in detail the nature and frequency of bias reversal within Mollusca, Platyhelminthes and Nematoda.

Compositional bias reversals are synapomorphic within both the nematodes and platyhelminths. Intriguingly, for both the Platyhelminthes and Nematoda, over 93% of the sequenced mitogenomes in each phylum exhibit

reversal of compositional biases. We note that reversals are not a common feature among other Ecdysozoan mtDNAs. For example, biases within arthropod mtDNAs closely resemble those of chordates (i.e. no reversal) in 92% of species (Figure 2b-c). Furthermore, compositional bias reversals did not preferentially occur among those species of Nematoda that had undergone mtDNA rearrangements. These results illustrate that within these particular lineages, bias reversals are neither rare nor homoplasious, but instead synapomorphic.

Inversion of the major non-coding region alone is not sufficient to produce reversal of compositional biases. Bony fish cluster tightly at very low (near zero) AT skew and strongly negative GC skew indicating an equal distribution of A and T, but not G and C, on the primary strand (Table 1, Figure 3b). The single notable outlier is *Tetrabrachium ocellatum*, (Tetrabrachiidae) a lophiiform species which harbors a highly rearranged mtDNA relative to other lophiiform fishes (Miya et al. 2010; Poulsen et al. 2013). However, *T. ocellatum* mtDNA does not contain an inversion of the major non-coding region (NCR), as has been frequently observed for other vertebrate species which have suffered “reversals” (Near et al. 2012). When mtDNA sequences are employed in phylogenetic reconstructions of lophiiform fishes, *T. ocellatum* is subject to severe long-branch attraction artifacts (Poulsen et al. 2013). Mitochondrial genomes of species from the sister Family Antennariidae are neither rearranged nor unusual in their AT or GC skews (falling within the inter-quartile range for fish), and the same is true for

their common ancestors, the Brachionichthyidae. Thus we attribute the recalcitrance of *T. ocellatum* to phylogenetic placement in part to lineage-specific reversal of nucleotide composition biases that are independent of NCR translocation or mtDNA rearrangement. In contrast, the mitogenomes of two other lophiiform species, *Ceratias uranoscopus* and *Cryptopsaras couesii* (Ceratiidae), have approximate median mtDNA skew values for teleosts (Supplemental Table 1) despite the fact that both have documented NCR translocations (Shedlock et al. 2004). These findings demonstrate that changes in the direction of compositional bias, i.e. the sign of the skew value (+ or -) are not specific to inversion or translocation of the NCR. Despite previous studies that predict an association between NCR translocation and skew reversal (Fonseca and Harris 2008), we demonstrate that compositional bias reversals may be associated with mtDNA rearrangement that does not involve the NCR.

Bias reversal co-occurs with non-maternal mtDNA inheritance among molluscan species. S_{GC} and S_{AT} values for mollusc mtDNAs may be subdivided into two clusters with skews of comparable strength, but opposite sign relative to each other (Figure 3c). This result is consistent with a reversal of mutational constraint for mtDNAs within one cluster. Bivalvia is clearly split, with species in which doubly-uniparental inheritance (DUI) of mtDNA is documented (Ghiselli et al. 2013), forming a tight cluster. This observation suggests a possible link between the mode of mtDNA inheritance and the accumulation of strand

asymmetric substitutions. Gastropods mtDNAs form a discrete cluster with three outliers (Figure 3c). One of these outliers, *Lottia digitalis,* fails to conform to the biases present within either the primary gastropod cluster or the bivalves. This deviation could indicate a recent change in the mutation pressure driving the sign of S_{AT} and S_{GC}, perhaps capturing *L. digitalis* in a transition from one cluster to the other. Of note, the *L. digitalis* mtDNA is the largest and most complex of the Prosobranchia sequenced to date, and contains repetitive sequences that may facilitate inter- and intra-molecular rearrangement, consistent with an association between skew value "reversals" and rearrangement of mitogenomic architecture (Jannotti-Passos et al. 2010).

Taken together, these data indicate that origin inversion can be one cause, but is not *the only* catalyst for reversals in AT and GC biases. Indeed, we identify here two additional, independent scenarios that may co-occur with skew reversal: mitogenomic rearrangement without NCR inversion, and non-maternal mitochondrial inheritance.

Nucleotide composition and strand asymmetry within nematode mitochondrial genomes

Nucleotide composition, strand asymmetry and mitogenomic architecture. We selected a single phylum, the Nematoda, for a more in-depth analysis of mtDNA nucleotide composition and strand asymmetry. Nematoda presents a particularly interesting case study for exploration of nucleotide asymmetry reversal because 93% of mtDNAs display negative S_{GC} and positive S_{AT} values. Furthermore, we previously identified highly rearranged nematode mtDNA architectures that involve rearrangements of protein-coding genes and sequence inversions (Hyman et al. 2011), enabling the direct testing of how strand asymmetry may be related to mitochondrial gene transcriptional polarity and proximity to non-coding regions predicted to contain the origin(s) of replication.

To investigate whether revered strand asymmetry in nematode mtDNAs could be attributed to unusual nucleotide composition, we calculated %GC and %AT for a catenated set of 12 protein-coding genes at all codon positions for 75 nematode mitochondrial genomes available in NCBI GenBank. Nematode mtDNAs are AT rich, averaging less than 30% GC (Rota-Stabelli et al. 2010); this value is consistent with the range of nucleotide compositions present in other

Ecdysozoan phyla, despite the departure of nematode strand asymmetries from those of other Ecdysozoan lineages (Gissi et al. 2008).

Nematode mtDNAs encode genes on both strands and exhibit elevated gene rearrangement and inversion rates relative to the other animal lineages (Gissi et al. 2008; Hyman et al. 2011). We asked if the "reversals" of S_{GC} and S_{AT} sign in nematode mtDNAs might be explained by a shared mtDNA rearrangement event that resulted in the inversion of a replication origin early in nematode divergence. To this end, we examined the order of protein-coding genes and their transcriptional orientation relative to the largest non-coding region of each nematode mtDNA. This analysis revealed that rearrangements involving the largest non-coding region, putatively containing sequences functional to mitochondrial gene expression and maintenance including a replication origin, occurred in 7 of 75 species, yet 73 of 75 nematode mtDNAs display bias reversal, regardless of gene order or NCR inversion. We conclude that for nematodes, as for the actinopterygiian fishes, inversion itself is not sufficient to alter nucleotide compositional asymmetries.

We next considered a possible relationship between gene transcriptional polarity and skew. To test whether genes encoded on opposite strands have opposite compositional biases, we repeated the skew analysis on a set of catenated protein-coding genes from 47 nematode species, 9 of which encoded

protein-coding genes on both mtDNA strands. We computed skew values first for the catenated protein-coding genes in their native transcriptional polarity (Figure 4a), and then with protein-coding genes encoded on the non-primary strand artificially inverted so that all coding sequences shared the same transcriptional orientation (Necşulea and Lobry 2007). Surprisingly, this treatment did not alter the sign of GC or AT skews, and strand asymmetry values for individual mitogenomes did not vary significantly regardless of transcriptional polarity (Supplemental Table 4); these data indicate that transcriptional orientation does not affect compositional biases within nematode mtDNAs.

Intriguingly, representatives of the nematode taxonomic class Enoplea carry mtDNAs with weaker skew values when compared to the more recently derived class Chromadorea, which includes *Caenorhabditis elegans* and its relatives. This observation held independent of gene transcriptional polarity. MtDNAs from multiple enoplean species contain large non-coding repeats within their mitogenomes that may facilitate homologus recombination (Hyman et. al. 2011). Thus it is possible that mutation rates are suppressed within some nematode species as a result of periodic gene conversion events promoted by recombination, consistent with our hypotheses regarding molecular evolution of plant and fungal mitogenomes.

Nematode mtDNA strand biases are equivalent at third codon positions and fourfold-degenerate sites. We next addressed the impact of compositional biases on phylogenetically informative sites within nematode mitogenomes. We measured S_{GC} and S_{AT} for mitochondrial protein-coding genes at all codon positions, third codon positions only, or fourfold degenerate sites only. We found that the compositional biases not only persist at 3rd codon positions but are stronger than values calculated for all three codon positions within the coding genes (see Supplemental Table 4 for p-values of binomial tests). Notably, these observations held when the analysis was restricted to fourfold-degenerate sites, though the significance was less (Supplementary Table 4). Consideration of the median and range of S_{GC} at 3rd codon positions or fourfold degenerate sites revealed highly similar trends between the two data subsets as compared to S_{GC} calculated across all codon positions (Figure 5a), and skews at fourfold degenerate sites were not significantly different from skews at all third codon positions (Supplementary Table 4). These data imply that fourfold-degenerate nucleotide positions, ostensibly free from selection, do not accumulate mutations at different rates or in different patterns than constrained positions. We propose that the persistence of biases, independent of selection, is consistent with our speculation as to the role of genetic processes, such as gene conversion, in shaping compositional biases within mitochondrial genomes; DNA strand invasion and subsequent crossover relies on homology, but does not distinguish among codon positions.

Previous work on compositional symmetries in the mitochondrial genomes of tardigrades, priapulids and onychophorans demonstrated that within these lineages, inverted mitochondrial genes tend to retain the strand asymmetry signature of their former transcriptional polarity at the 3rd codon position (Rota-Stabelli et al. 2010). Our data clearly demonstrate that this is not be the case for nematodes that encode genes on both mtDNA strands, as skews for the coding regions are not highly influenced by artificial inversion. Our data confirm that mitogenomic rearrangement, mutation, and/or DNA repair processes may act in a coordinate fashion to modulate compositional biases at all codon positions.

Variation in compositional biases among nematode mitochondrial genomes and implications for mtDNA maintenance. To determine whether compositional asymmetries varied within different genes or regions of nematode mtDNAs, we calculated the GC and AT skew across the coding region(s) of each mitogenome using a sliding window strategy. We reasoned that if adjacent genes harbored opposite compositional biases, the sign (+ or -) of GC and/or AT skew would be inverted at the gene junctions, as occurs in the nuclear genomes of several eukaryotes (Agier and Fischer 2012). In contrast, if biases were homogenized along the mtDNA, GC and AT skew values would also remain constant. Of the 75 nematode mtDNAs analyzed in this study, an example plot is shown in Figure 5a representing the Chromodorea (*Caenorhabditis elegans*), and the Enoplea, (*Hexamermis agrotis*). In *H. agrotis*, but not *C. elegans*, both

GC and AT skew values bracket 0.0 The differences in sliding window topologies for these two nematodes accurately reflect those observed for nematodes in each of the two nematode taxonomic classes. Moreover, these results reflect our earlier finding that compositional biases tend to be much stronger in chromadorean species. We did not detect any change in biases at gene junctions based on the absence of abrupt skew sign change across all genome windows, indicating that compositional biases do not vary by protein-coding gene within nematode mtDNAs.

We next calculated the *cumulative* GC skew along each nematode mitochondrial genome with a similar sliding window approach. Rather than calculating the GC and AT skew within each individual window (200 bp), window values are added to generate a running sum across the entire coding region (Arakawa and Tomita 2012; Mascher et al. 2013). When compared over the length of the mtDNA from an arbitrary starting point, the slope of the cumulative GC skew indicates the relative rate at which mutations accumulate in different genomic regions. These plots can be classified by their shape and have proven informative about replication-associated mutations in a variety of contexts, as inversion of the cumulative GC bias slope on this type of plot is indicative of the presence of a replication origin and/or termination site (Touchon and Rocha 2008). Of the 75 mtDNAs analyzed, example plots from 3 species, *Caenorhabditis elegans* (chromadorea), *Ascaris suum* (chromadorea), and

Romanomermis iyengari (enoplea) are shown in Figure 6a. Interestingly, for chromadorean species, cumulative GC skew slope is constant and positive (Figure 6a, leftmost and middle panels), whereas for the enoplean species, multiple inversions were observed (Figure 6a, rightmost panel). Within vertebrate mtDNAs, cumulative GC skews tend to increase from the site of the origin on a constant slope, but do not reach a single discrete maximum as origin and termination sites are adjacent (Lin et al. 2012). The presence of multiple slope inversions in the cumulative GC plot of enoplean, but not chromadorean species suggests that multiple mtDNA replication origins and termination sites are present in enoplean mtDNAs, and this impacts GC skew. Furthermore, these results could be interpreted to indicate the lack of a discrete replication origin and termination site in the mtDNAs of chromadoreans. The mechanism of mtDNA replication in nematodes is poorly understood (Bratic et al. 2009), and further investigations are needed to determine the utility of these speculations. It is also interesting to note that *A. suum* mtDNA has experienced a 4kb translocation of the major non-coding region presumed to contain a replication origin relative to the C. elegans mitogenome, yet the cumulative GC plots closely resemble each other. This result further supports our proposal that rearrangements involving the non-coding regions are not alone sufficient to impact mtDNA skews.

To enable a broad comparison among multiple nematode mtDNAs, the cumulative GC skew profiles for 75 nematode species were clustered based on

feature similarity, and represented as a heatmap (Figure 6b). This analysis identified three very different patterns of mutation within nematode mitogenomes. The vast majority of cumulative GC skew profiles conformed to the pattern of skew represented by *C. elegans* and *A. suum* in Figure 6a. As shown in Figure 6a, i.e. the plots were devoid of both minima and maxima, consistent with a static rate of compositional asymmetry accumulation throughout the entire mtDNA molecule. As we have suggested, this plot topology may suggest the absence of a discrete replication origin within the mitogenomes of these nematode species; alternatively, an mtDNA replication mechanism in which each initiation event produced multiple daughter molecules, such as the rolling-circle replication mechanism employed by bacteriophages, would also be consistent with such a pattern.

For nine mtDNAs of species belonging to the infraclass Dorylaimia (Enoplea; (Hyman et al. 2011) the rendered cumulative GC skew profiles display a pattern reminiscent of circular bacterial genomes with a single inflection point (figure 6b). The significance of this pattern is challenging to interpret, however it has been suggested that this pattern could imply the presence of a discrete replication origin without a specific termination site, or vice versa (Touchon and Rocha 2008).

The identification of mutational constraint reversal as synapomorphic for the nematodes and platyhelminths is highly unusual among animals and raises the question as to when reversal occurred for these phyla. As the relationships among basal Ecdysozoan phyla are yet unclear, it remains possible that these reversals have not arisen in parallel within the two clades, but are attributable to a single event in Ecdysozoan evolution (Rota-Stabelli et al. 2010; Bernt et al. 2012).

Discussion

Our expansion of compositional asymmetry analyses to incorporate more species within the Chordates and Arthropods, and new branches of the Eukaryota, has provided the unexpected result that bias strength and patterning is highly lineage specific. Indeed, the observation that compositional biases vary significantly among animals, plants and fungi, and within the animal lineage itself raises the possibility that multiple mitochondrial genome synthesis and repair mechanism have evolved within eukaryotes, opening new avenues for future study. While mitochondrial genome synthesis has primarily been studied within mammalian cells, recent work supports the presence of several mtDNA replication mechanisms, not all of which are efficient in protecting mtDNA from the mutagenic environment of the mitochondrial matrix (Pohjoismäki and Goffart 2011; Holt and Reyes 2012; Jõers and Jacobs 2013). Further studies on mtDNA

synthesis in species beyond the established mammalian models are necessary to evaluate possible correlation between replication mechanism, mutation rate and compositional biases.

Materials and Methods

MtDNA sequence data. The eukaryotic data set includes 2,428 mtDNA sequences retrieved from the National Center for Biotechnology Information FTP web server (ftp://ftp.ncbi.nlm.nih.gov/genomes/MITOCHONDRIA/Metazoa/) in October 2012 and analyzed in the R statistical computing language with both custom R scripts and functions from the *seqinR* package following formatting with the Bioconductor package *Biostrings* and/or MEGA 5.0 (Charif et al. 2005; Pages et al.). All coding and compositional bias analyses were executed in the RStudio Integrated Development Environment (RStudio 2012).

S_{GC} *and* S_{AT} *calculations.* GC and AT skews were calculated as described (Perna and Kocher 1995). For mtDNA sequences retrieved from NCBI Genbank, the strand present in the database annotated as the reference sequence was used for skew calculations. For metazoan mitochondrial genomes, convention typically employs the strand that encodes the greatest number of protein-coding genes. For nematode mtDNAs that may encode an equal number of protein-coding genes on each strand, both strands were analyzed for the presence of

skew; no significant difference was found in the absolute value of GC or AT skews between strands for those mitogenomes (data not shown).

Formatting and artificial mtDNA architecture rearrangement. Nematode mtDNAs were imported into MEGA 5.0 for extraction of sequences by codon position, by fourfold-degenerate sites, and for formatting for cumulative GC skew calculations. MtDNAs were reoriented such that the first nucleotide annotated to the gene immediately 3-prime of the major non-coding region was assigned position 0, and trimmed to include the catenated protein-coding region only. Nematode mitogenomes were aligned with the MUSCLE algorithm. Termination codons were excluded from the nematode analyses listed above, as they can vary from 1 to 3 nucleotides in length within this lineage and converted to triplets when necessary by post-transcriptional polyadenylation (Lemire 2005). For artificial rearrangement analysis, protein-coding genes were inverted in MEGA 5.0.

Cumulative GC skew and clustering. Cumulative GC skew was calculated essentially as described (Perna and Kocher 1995); a 200 nucleotide window was used with 50 nucleotide overlap between windows. Heirarchical clustering of cumulative GC skew profiles was accomplished with the Lance-Williams formula, using defaults of the *heatmap* function in R package *stats*.

Statistical analyses. All statistical analyses were performed in R. Correlations and significance tests were implemented using the *cor.test* function within the package *stats*. Both Pearson and Spearman correlations were performed unless

otherwise noted, as the linearity of the relationship between GC and AT skews was unknown at outset. References in the text to “weak” or “strong” skew refer to absolute values in compared to the absolute value of the median, unless otherwise noted.

References

Agier N, Fischer G. 2012. The mutational profile of the yeast genome is shaped by replication. Mol. Biol. Evol. 29:905–913.

Arakawa K, Tomita M. 2012. Measures of Compositional Strand Bias Related to Replication Machinery and its Applications. Curr. Genomics 13:4–15.

Backert S, Lynn Nielsen B, Börner T. 1997. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends in Plant Science 2:477–483.

Bendich AJ. 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr Genet 24:279–290.

Bernt M, Braband A, Schierwater B, Stadler PF. 2012. Genetic aspects of mitochondrial genome evolution. Molecular Phylogenetics and Evolution.

Bratic I, Hench J, Henriksson J, Antebi A, Burglin TR, Trifunovic A. 2009. Mitochondrial DNA level, but not active replicase, is essential for Caenorhabditis elegans development. Nucleic Acids Research 37:1817–1828.

Castellana S, Vicario S, Saccone C. 2011. Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein coding genes. Genome Biology and Evolution.

Charif D, Thioulouse J, Lobry JR, Perrière G. 2005. Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics 21:545–547.

Christensen AC. 2013. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biology and Evolution 5:1079–1086.

Denver DR, Morris K, Lynch M, Vassilieva LL. 2000. High Direct Estimate of the Mutation Rate in the Mitochondrial Genome of Caenorhabditis elegans. Science.

Fonseca MM, Harris DJ. 2008. Relationship between mitochondrial gene rearrangements and stability of the origin of light strand replication. Genet. Mol. Biol. 31:566–574.

Fonseca MM, Posada D, Harris DJ. 2008. Inverted Replication of Vertebrate Mitochondria. Mol. Biol. Evol. 25:805–808.

Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin SV, Passamonti M. 2013. Structure, transcription and variability of metazoan mitochondrial genome. Perspectives from an unusual mitochondrial inheritance system. Genome Biology and Evolution.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301–320.

Haag-Liautard C, Coffey N, Houle D, Lynch M. 2008. PLOS Biology: Direct Estimation of the Mitochondrial DNA Mutation Rate in Drosophila melanogaster. PLoS Biol.

Hassanin A, Léger N, Deutsch J. 2005. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. Systematic Biology 54:277–298.

Hassanin A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Molecular Phylogenetics and Evolution 38:100–116.

Holt IJ, Reyes A. 2012. Human Mitochondrial DNA Replication. Cold Spring Harbor Perspectives in Biology.

Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, Thermes C. 2007. Human gene organization driven by the coordination of replication and transcription. Genome Res. 17:1278–1285.

Hyman BC, Lewis SC, Tang S, Wu Z. 2011. Rampant gene rearrangement and haplotype hypervariation among nematode mitochondrial genomes. Genetica 139:611–615.

Jannotti-Passos LK, Ruiz JC, Caldeira RL, Murta SMF, Coelho PMZ, Carvalho OS. 2010. Phylogenetic analysis of Biomphalaria tenagophila (Orbigny, 1835) (Mollusca: Gastropoda). Mem. Inst. Oswaldo Cruz 105:504–511.

Jõers P, Jacobs HT. 2013. Analysis of replication intermediates indicates that Drosophila melanogaster mitochondrial DNA replicates by a strand-coupled theta mechanism. PLoS ONE 8:e53249.

Lemire B. 2005. Mitochondrial genetics. WormBook:1–10.

Lin Q, Cui P, Ding F, Hu S, Yu J. 2012. Replication-Associated Mutational Pressure (RMP) Governs Strand-Biased Compositional Asymmetry (SCA) and Gene Organization in Animal Mitochondrial Genomes. Curr. Genomics 13:28–36.

Mascher M, Schubert I, Scholz U, Friedel S. 2013. Patterns of nucleotide asymmetries in plant and animal genomes. BioSystems 111:181–189.

Miya M, Pietsch TW, Orr JW, et al. 2010. Evolutionary history of anglerfishes (Teleostei: Lophiiformes): a mitogenomic perspective. BMC Evol. Biol. 10:58.

Montooth KL, Rand DM. 2008. The spectrum of mitochondrial mutation differs across species. PLoS Biol 6:e213.

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proceedings of the ….

Necşulea A, Lobry JR. 2007. A new method for assessing the effect of replication on DNA base composition asymmetry. Mol. Biol. Evol. 24:2169–2179.

Nikolaou C, Almirantis Y. 2006. Deviations from Chargaff's second parity rule in organellar DNA: Insights into the evolution of organellar genomes. Gene.

Pages H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: String objects

representing biological sequences, and matching algorithms

. R version 2.13.1. Available from: http://www.bioconductor.org/packages/2.12/bioc/html/Biostrings.html

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358.

Pohjoismäki JLO, Goffart S. 2011. Of circles, forks and humanity: Topological organisation and replication of mammalian mitochondrial DNA. Bioessays 33:290–299.

Poulsen JY, Byrkjedal I, Willassen E, Rees D, Takeshima H, Satoh TP, Shinohara G, Nishida M, Miya M. 2013. Mitogenomic sequences and evidence from unique gene rearrangements corroborate evolutionary relationships of myctophiformes (Neoteleostei). BMC Evol. Biol. 13:111.

Rand DM. 2001. The Units of Selection on Mitochondrial DNA. Annual Review of

Ecology and Systematics 32:415–448.

Rocha EPC, Touchon M, Feil EJ. 2006. Similar compositional biases are caused by very different mutational effects. Genome Res. 16:1537–1547.

Rocha EPC. 2004. The replication-related organization of bacterial genomes. Microbiology 150:1609–1627.

Rota-Stabelli O, Kayal E, Gleeson D, Daub J, Boore JL, Telford MJ, Pisani D, Blaxter M, Lavrov DV. 2010. Ecdysozoan mitogenomics: evidence for a common origin of the legged invertebrates, the Panarthropoda. Genome Biology and Evolution 2:425–440.

Seligmann H. 2012. Coding constraints modulate chemically spontaneous mutational replication gradients in mitochondrial genomes. Curr. Genomics 13:37–54.

Sernova NV, Gelfand MS. 2008. Identification of replication origins in prokaryotic genomes. Briefings in Bioinformtics 9:376–391.

Shedlock AM, Pietsch TW, Haygood MG, Bentzen P. 2004. Molecular systematics and life history evolution of anglerfishes (Teleostei: Lophiiformes): Evidence from mitochondrial DNA. Steenstrupia.

Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A, Larsson N-G. 2008. Strong Purifying Selection in Transmission of Mammalian Mitochondrial DNA. PLoS Biol 6:e10.

Touchon M, Rocha EPC. 2008. From GC skews to wavelets: a gentle guide to the analysis of compositional asymmetries in genomic data. Biochimie 90:648–659.

Wei S-J, Shi M, Chen X-X, Sharkey MJ, van Achterberg C, Ye G-Y, He J-H. 2010. New views on strand asymmetry in insect mitochondrial genomes. PLoS ONE 5:e12708.

Wei S-J, Shi M, Sharkey MJ, van Achterberg C, Chen X-X. 2010. Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. BMC Genomics 11:371.

Kingdom	Phylum	Class	N	GC Skew median	AT Skew median	Pearson (R)	Significance	Spearman (rho)	Significance
Viridiplantae			278	0.00154	-0.00049	0.606407	***	0.3729	***
Fungi			121	0.06578	-0.00485	0.2805	*	0.19915	*
Animalia			2029	-0.268	0.05565	-0.84833	***	-0.676655	***
	Arthropoda		378	-0.2056	0.03444	-0.51376	***	-0.29466	***
	Chordata		1454	-0.285	0.06561	-0.63451	***	-0.61796	***
		Aves	151	-0.3879	0.1229	-0.180062	*	-0.1801	*
		Amphibia	106	-0.247	0.03199	-0.054675	none	-0.8088	none
		Mammalia	336	-0.3257	0.09352	-0.609545	***	-0.6261	***
		Actinopterygii	861	-0.2652	0.05000	-0.472813	***	-0.412	***
	Mollusca		86	0.1006	-0.1267	-0.936	***	-0.77306	***
	Nematoda		75	0.4016	-0.2401	-0.865271	***	-0.70178	***
	Platyhelminthes		36	0.4105	-0.3052	-0.756137	***	-0.63526	***

Table 1. Median skew values for eukaryotic taxa and tests for correlation between GC and AT skew.
Note: *p<0.05, **p<0.001, ***p<0.0001

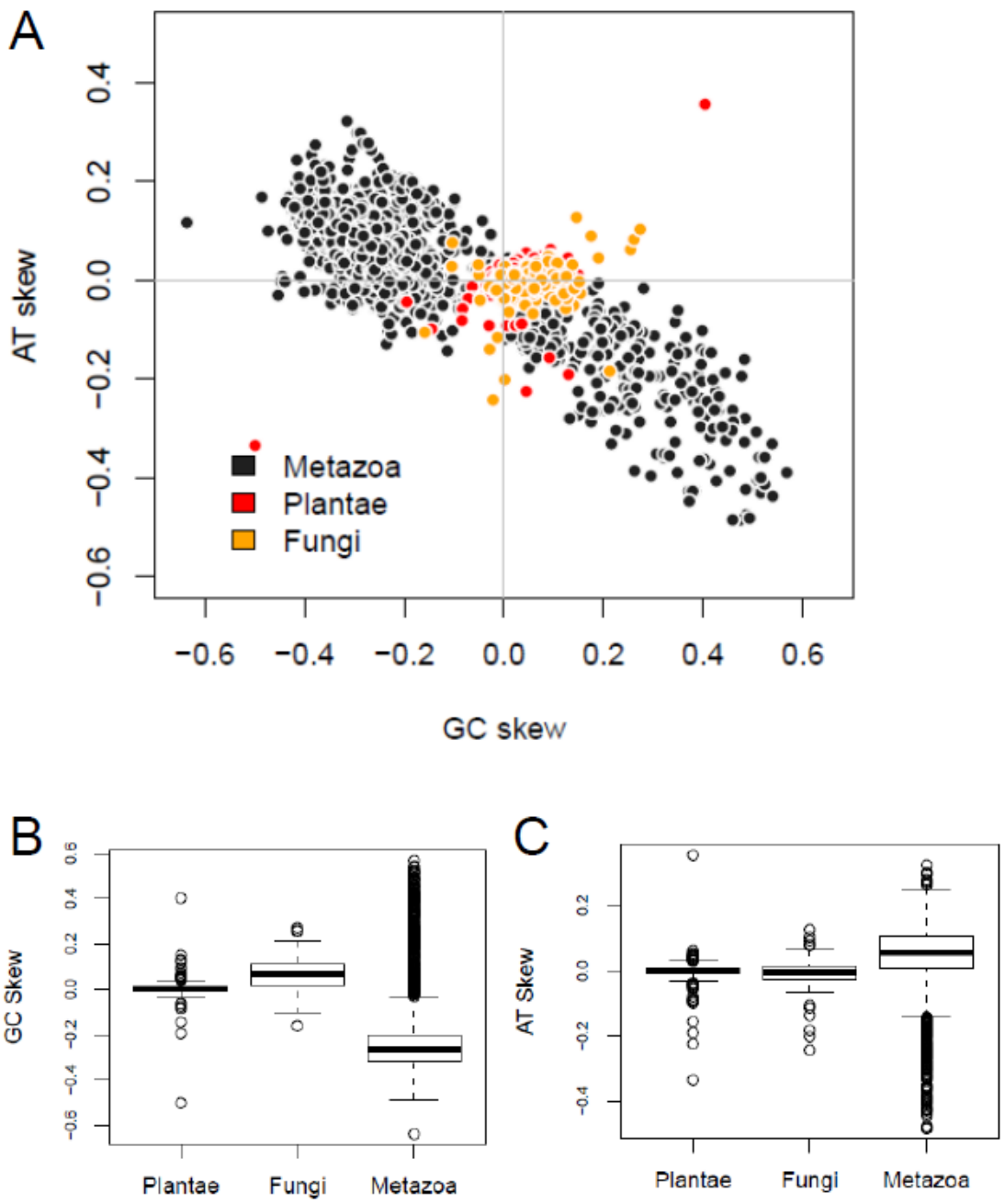

Figure 1. Compositional asymmetries in mitochondrial DNA. (A) Scatterplot of GC and AT skews calculated for 2,397 eukaryotic mtDNAs. GC and AT skews are negatively correlated within the metazoa, but positively correlated for plants and fungi. Metazoan mtDNA GC skews (B) and AT skews (C) are more varied than plant or fungal skew values.

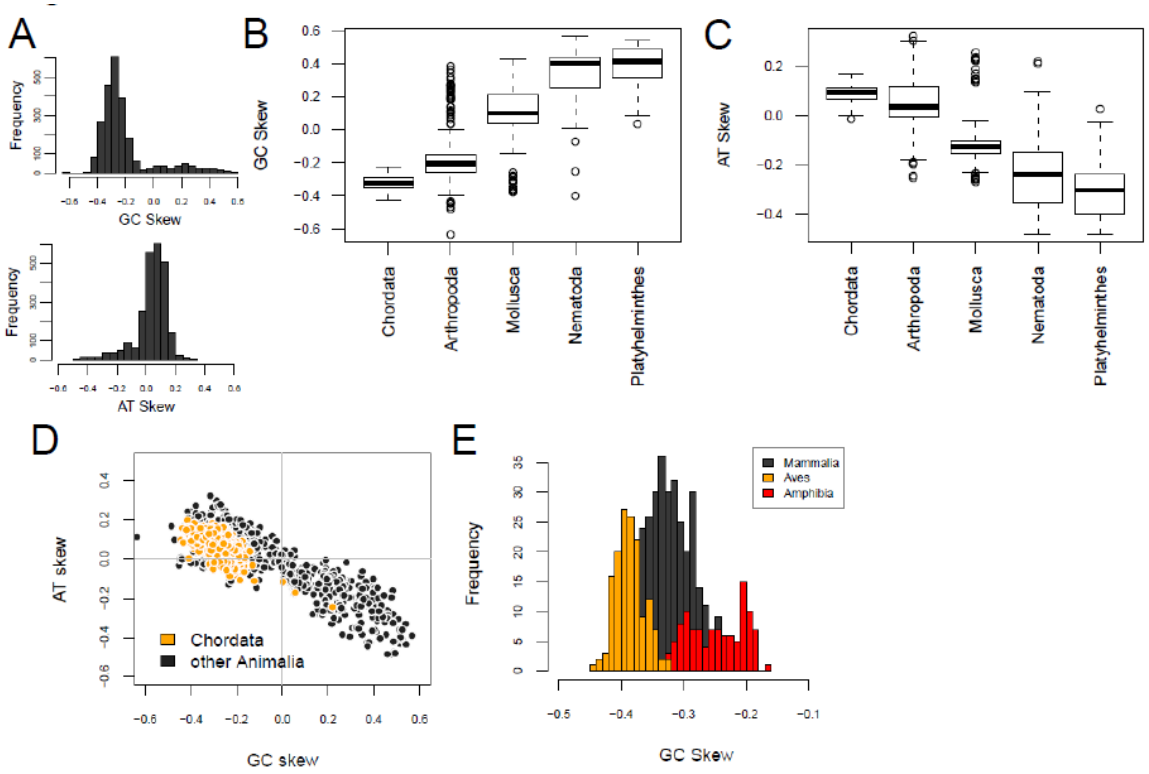

Figure 2. GC and AT skews vary among animal lineages. (A) Neither metazoan GC skews, top panel, or AT skews, bottom panel, are normally distributed. Boxplots of metazoan GC skews (B) and AT skews (C) by phylum demonstrate the presence of stronger skew values in nematodes and helminthes. (D) Scatterplot of metazoan GC and AT skews; chordates (yellow) display decreased variability in mtDNA skew compared to non-chordate species (grey). (E) Histogram plots of GC skew in bird (yellow), mammal (grey) and amphibian (red) mtDNAs indicate substructure within the Chordate cluster.

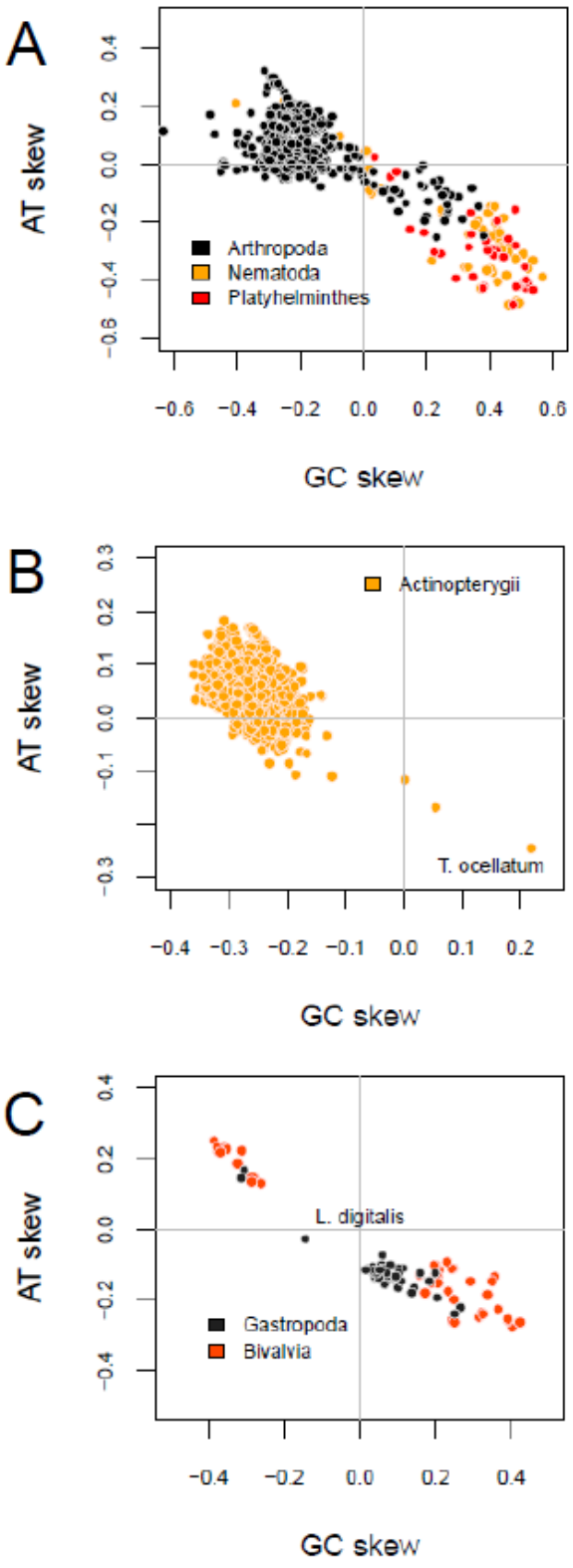
A
AT skew
GC skew
Arthropoda
Nematoda
Platyhelminthes
B
AT skew
GC skew
Actinopterygii
T. ocellatum
C
AT skew
GC skew
L. digitalis
Gastropoda
Bivalvia

Figure 3. Reversals of GC and AT skew are common in nematode and helminth mitochondria. (A) Scatterplot of nematode (yellow) and platyhelminth (red) mtDNAs display compositional bias reversal characterized by the presence of their skew values in the quadrant of positive GC skew values and negative AT skew values. In contrast, the majority of arthropod mtDNAs (grey) display negative GC skews and positive AT skews. (B) Scatterplot of Actinopterygiian fish mtDNA GC and AT skews with the single identified asymmetry reversal of *T. ocellatum* noted (lower right). (C) Scatterplot of molluscan GC and AT skews. Note the location of *L. digitalis* near the 0,0 skew position, between the two major Mollusc clades, Bivalvia and Gastropoda.

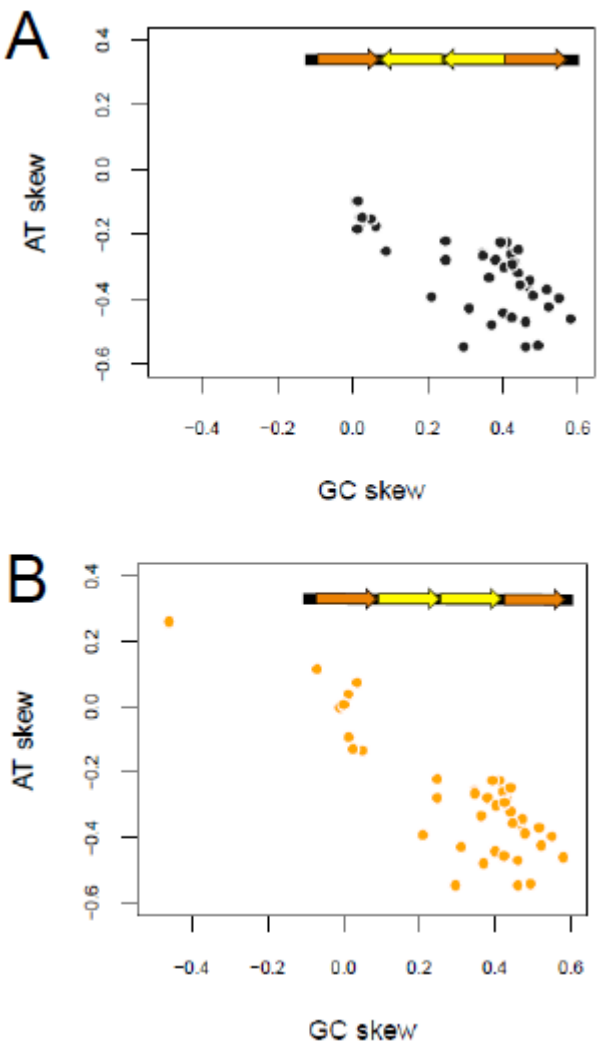

Figure 4. Transcriptional polarity is independent of skew in nematode mitochondrial genomes. (A) Scatterplot of the skews present in nematode mtDNAs with transcription from both mtDNA strands. (B) Aritifical inversion of mtDNA genes does not move nematode mtDNAs from the positive GC skew, negative AT skew quadrant.

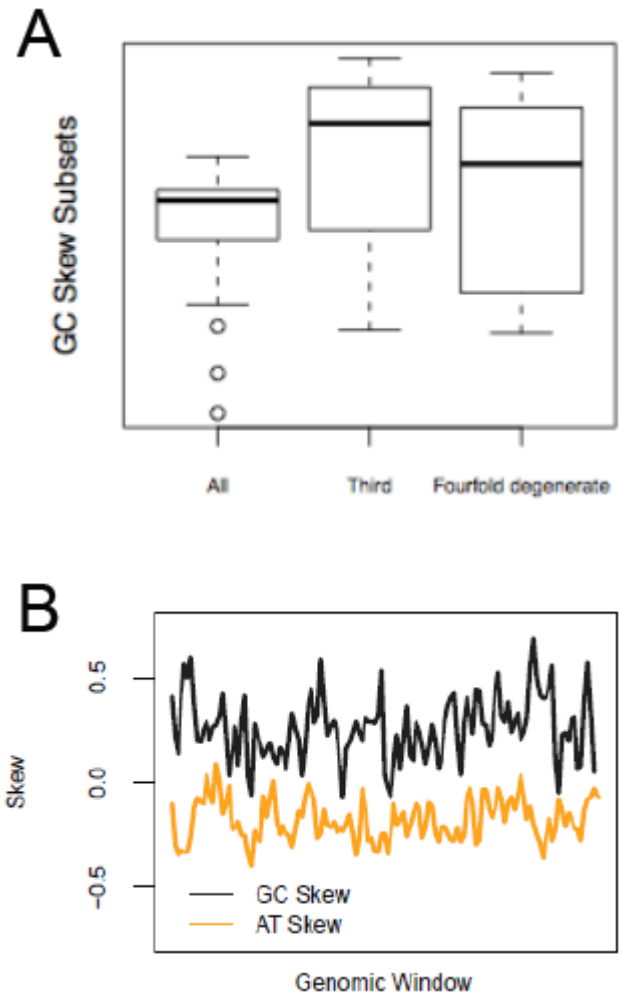

Figure 5. Compositional asymmetries are present at third codon positions, and fourfold degenerate sites. (A) Boxplot comparing GC skews of nematode mtDNAs calculated over all codon positions ("All"), third codon positions only ("Third"), or fourfold degenerate sites only ("Fourfold"). (B) Example plots of GC and AT skews per 200 nucleotide window of the mtDNAs of *C. elegans* (top panel) and *H. agrotis* (bottom panel).

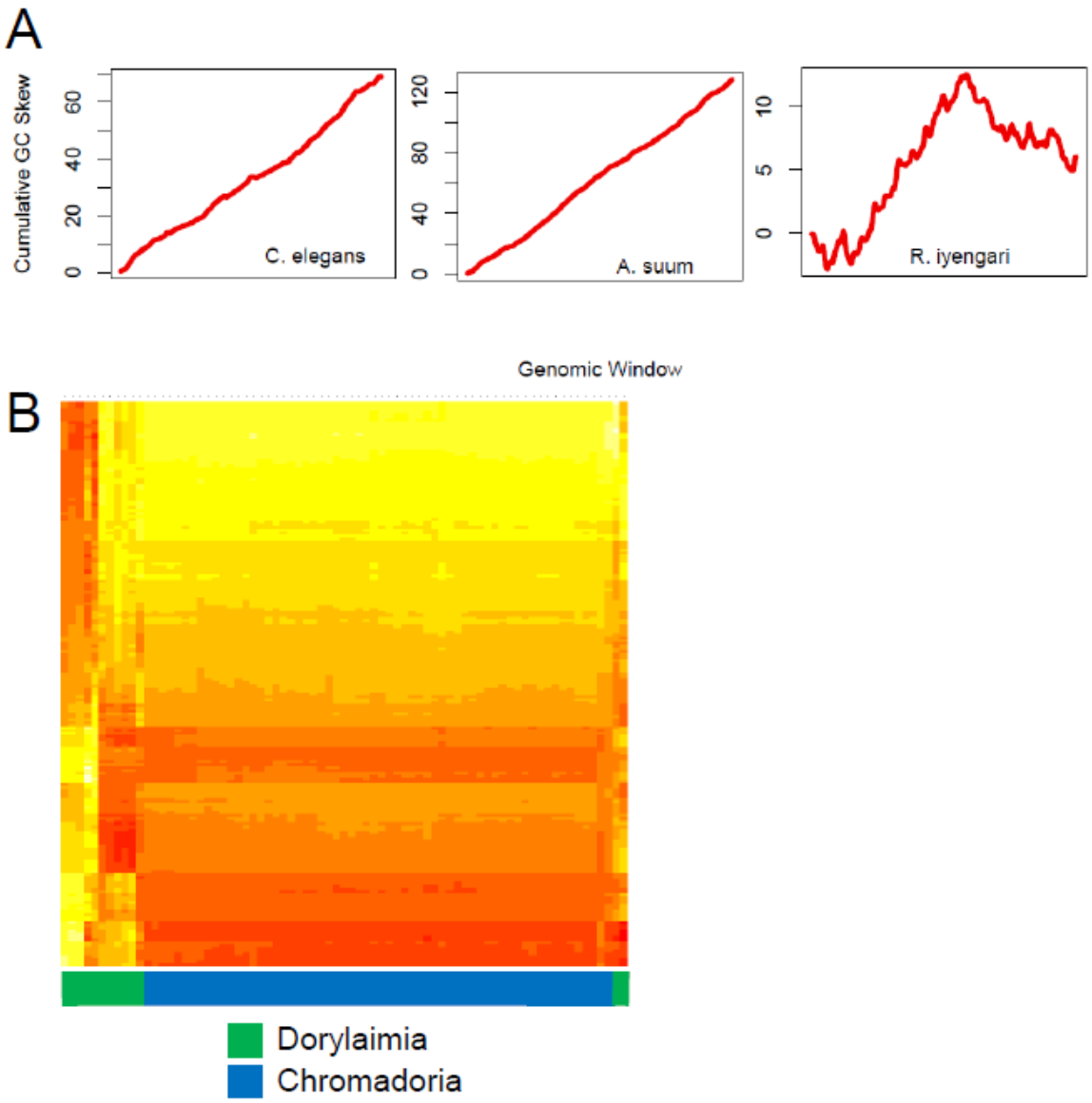

Figure 6. Cumulative GC skew in nematode mitochondrial DNA. (A) Example plots of GC skew for each 200 nucleotide window of the *C. elegans*, *A. suum* and *R. iyengari* mtDNAs. (B) Heatmap describing the cumulative GC skew profiles of all 75 nematodes included in this study. Dorylaims (green bar, below) display a divergent skew pattern in comparison to Chromadorians (blue bar, below).

Chapter 3. Topological analysis of *Caenorhabditis elegans* mitochondrial DNA

Introduction

Animal mitochondria contain a DNA genome that replicates independent of chromosomal DNA housed in the cell's nucleus. Across the metazoa, mitochondrial DNA (mtDNA) averages 16kb in size although lengths vary widely among taxa. MtDNA encodes protein subunits of the mitochondrial respiratory complexes, as well as mitochondrial transfer RNAs and ribosomal RNAs required for the translation of those genes(Falkenberg et al. 2007). The heritable monomer form of metazoan mitochondrial DNA (mtDNA) is typically a double-stranded circle (Falkenberg et al. 2007). Notable exceptions include the mitogenome of the cnidarian *Hydra attenuata,* which is composed of linear concatemers, and that of the nematode *Globodera pallida* which is an amalgam of multiple "minicircles", each containing one or a few mitochondrial genes (Armstrong et al. 2000; Pont-Kingdon et al. 2000).

Cellular respiration is essential for animal life, and in turn requires the activity of respiratory complex subunits encoded in mtDNA (Nunnari and Suomalainen 2012). Thus defects in the replication and maintenance of mtDNA cause a severe, heritable class of myopathies in humans (Falkenberg et al. 2007). The model nematode, *Caenorhabditis elegans*, is considered an up-and-coming experimental system for the study of mtDNA maintenance and specifically mtDNA replication-associated defects (Addo et al. 2010; Bratic et al.

2010). The nematode is small and easily grown on petri dishes or in liquid cultures in the laboratory, offering the practical advantages of a microorganism, yet is a motile multi-cellular animal with highly differentiated tissues that is easily manipulable by genetic techniques such as RNA interference and Caspase-mediated genome editing (Addo et al. 2010; Gaj et al. 2013).

The utility of *C. elegans* as a model for mtDNA replication defects is currently limited by a lack of knowledge about the native topology and replication of its mtDNA. MtDNA maintenance is well studied only in human cultured cells and mouse; it has long been assumed that other metazoans follow an identical or near-identical program of mtDNA synthesis as in cultured human cells (Gissi et al. 2008; John et al. 2010). However, recent studies have overturned this assumption, revealing that multiple mtDNA replication mechanisms operate within the mitochondria of some mice, birds and the fruitfly *Drosophila melanogaster* (Reyes et al. 2005; Yasukawa et al. 2005; Jõers and Jacobs 2013). Furthermore, in a groundbreaking series of studies, Pohjoismaki *et al.* recently demonstrated that mtDNA mechanisms may vary between cultured human cells and human tissues, as well as between human organs (Pohjoismäki et al. 2009; Pohjoismäki, Goffart, et al. 2010; Pohjoismäki, Holmes, et al. 2010).

Depletion of *C. elegans* mtDNA leads to a severe developmental arrest phenotype, suggesting that mtDNA is as important to metabolism in the worm as

in vertebrates and the fruitfly (Lemire 2005). The restriction map of the *C. elegans* mitochondrial genome as well as the complete nucleotide sequence are consistent with an intact circular structure (Okimoto et al. 1992). However, as in many animals where mtDNA has been characterized, the true *in vivo* topology of *C. elegans* mtDNA remains unknown. Given the multipartite nature of the mtDNA of the nematode *G. pallida*, it is important to establish whether the topology of *C. elegans* mtDNA is truly circular in nature. Here, we test the hypothesis that the mitogenome of *C. elegans* is circular double stranded (dsDNA) and provide evidence that this is indeed the case. Moreover, we demonstrate unique features of *C. elegans* mtDNA structure as compared to previously studied model organisms, most prominent being the absence of RNA-DNA hybrid tracts described in mammalian and fly mtDNA.

Results

The mitochondrial genome monomer of *C. elegans* is circular. If the *C. elegans* mitochondrial genome monomer is circular, mtDNA should migrate in an agarose gel as two discrete bands containing relaxed and supercoiled forms respectively, as is true for the mtDNA of vertebrates (Falkenberg et al. 2007). To test this prediction, we performed one-dimensional (1D) electrophoresis of enriched mitochondrial nucleic acid followed by Southern blot hybridization. Gels stained by ethidium bromide indicated the presence of DNA in at least two visible

forms (Figure 1A, lane “U”), a slowly migrating band coincident with the migration of a 30kb linear DNA marker, and a fast migrating band co-migrant with a 10kb linear DNA standard. In addition, fluorescence was observed in the well of each gel (Figure 1A, yellow arrow), indicating that a fraction of enriched mitochondrial nucleic acid was either too large to migrate into the gel, too topologically complex to migrate into the gel, or both. In contrast, nucleic acid treated with the restriction enzyme *ClaI,* that is predicted to cleave once per mtDNA monomer (Okimoto et al. 1992) (Figure 2), migrated as a single major band greater than 10kb but less than 23kb (Figure 1A, lane “ClaI”), consistent with the size of the *C. elegans* mitogenome at 13.8kb as inferred from restriction mapping and DNA sequencing (Okimoto et al. 1992).

Weaker staining bands were also observed; to test whether these bands were the result of mtDNA degradation after incubation at 37 degrees, mock-digested nucleic acid was co-electrophoresed (Figure 1A, lane “M”). As mock-digested nucleic acid closely resembled untreated DNA, the minor bands present in lane “ClaI” likely represent contaminating nuclear DNA, or *Escherichia coli* DNA ,the nematode food source in laboratory culture (). Consistent with these interpretations, only a single ~13.8kb band appeared on exposures of the corresponding Southern blot (Figure 1B, lane “ClaI”), and two major bands are clearly visible in the other lanes containing DNA undigested by restriction enzymes (Figure 1B, lanes “U” and “M”). These bands also indicate that the

probe used for hybridization was indeed mtDNA sequence-specific. The collapse of the major and minor nucleic acid bands into a single ~13.8kb band following *ClaI* treatment is consistent with the electrophoretic behavior of circular DNA upon cleavage by a restriction enzyme cutting once in the genome to generate a unit length linear molecule.

To test whether the major and minor mtDNA forms were susceptible to degradation by enzymes that cleave *C. elegans* mtDNA other than *ClaI*, we repeated the 1D analysis with an additional set of three restriction digests: *EcoRI*, *HindIII* and *StyI*, all of which are predicted to produce two restriction fragments (Figure 1C, 1D). These treatments each produced both two prominent bands (Figure 1C) in addition to many minor bands whose sizes, when summed, total 13.8kb. Only a subset of prominent bands was apparent on the corresponding Southern blot (Figure 1D), indicating that the prominent bands are mtDNA containing the probe sequence while faster migrating bands that do not share homology with the mtDNA hybridization probe were not detected. These results support our conclusion that *C. elegans* mtDNA is indeed circular in its non-replicative form.

We confirmed the circularity of mtDNA *in vivo* by examining *C. elegans* mitochondrial nucleic acid by transmission electron microscopy. Two classes of circular molecules were present: fully double-stranded DNA circles, and DNA

circles with regions of complex secondary structure suggesting gaps of single stranded DNA (ssDNA) (Griffith and Christiansen 1978) (Figure 3A, 3B black arrows). Fully double-stranded circles and circles with putative ssDNA gaps were observed in near equal proportion. Both classes of circular DNA measured 13.4kb +/- .5kb in length on average as calibrated to a co-spread 5kb plasmid, in line with the predicted *C. elegans* mtDNA size of 13.8kb. To test whether the complex secondary structures were truly ssDNA, we incubated *C. elegans* mitochondrial nucleic acid with the *Escherichia coli* single-stranded DNA binding protein (SSB). We reasoned that if the structures were ssDNA, mtDNA molecules treated with SSB would appear with thickened, punctate regions where the SSB had bound to the mtDNA (Thresher and Griffith 1992). We observed circular DNA with bound protein following SSB treatment (Figure 3C, black arrows). These experiments establish that *C. elegans* mtDNA is circular *in vivo* with some molecules containing ssDNA gaps.

MtDNA topisomers lack RNA and single-stranded regions, but harbor Holliday junctions. Next, we examined the response of *C. elegans* mtDNA topoisomers to a battery of nucleic-acid modifying endonucleases with different substrate specificities (Table 1). Following enzyme treatments, mitochondrial nucleic acid was subjected to 1D electrophoresis, Southern transfer and hybridization to detect mtDNA (Figure 4). Recent studies on metazoan mtDNA architecture have identified RNA-DNA hybrid replication intermediates in some

species, including humans. This hybrid structure stands in contrast to the partially single-stranded DNA (ssDNA) intermediates predicted by the canonical mtDNA replication mechanism operative in mammalian cells (Clayton 1982). These newer findings led to the proposal of an alternate mtDNA replication mechanism, which transiently incorporates RNA tracts into synthesis products, termed RNA Incorporation Throughout the Lagging Strand (RITOLS) (Holt and Reyes 2012; Reyes et al. 2013). To determine whether RITOLS or canonical partially-ssDNA replication intermediates were present in *C. elegans* mitochondria, we treated mitochondrial nucleic acid with RNaseH or S1 nuclease.

Surprisingly, the slow and fast migrating topological forms identified in Figure 1 were resistant to both RNaseH and S1. Only a small fraction of mtDNA was converted to linear molecules of a single genome length by these treatments while the slow migrating mtDNA band persisted (compare the *Cla*I digest in Figure 4 to the sample “RNaseH”), consistent with low frequency nicking of fast migrating supercoiled mtDNA. This result suggests that the RITOLS mechanism does not operate within *C. elegans*, as previous work has demonstrated that removal of RNA-DNA hybrid tracts by RNAseH results in the complete linearization of relaxed circular monomers when RITOLS is present (Jõers and Jacobs 2013; Reyes et al. 2013). In addition, these results suggest that gapped circular monomers containing lengths of ssDNA do not accumulate to a substantial degree in *C. elegans*, in contrast to mammals in which the majority of

replicating mtDNA molecules are in the gapped-circular form, at least in muscle and brain tissue (Pohjoismäki, Holmes, et al. 2010).

MtDNA recombination has been viewed as rare or absent within animal mitochondria (Falkenberg et al. 2007). However, an emergent recent literature has documented mtDNA recombination within a handful of species including a nematode, *Meloidogyne javanica* (Lunt and Hyman 1997; Gissi et al. 2008). Furthermore, many nematode mtDNAs, including *C. elegans*, contain repetitive non-coding region sequence that could facilitate homologous recombination (Okimoto et al. 1992; Hyman et al. 2011). Thus, we next investigated if mtDNA recombination intermediates were present in *C. elegans*. Intriguingly, we found that treatment of mitochondrial nucleic acid with the *E. coli RusA* resolvase, known to specifically cleave Holliday junctions, led to the degradation of circular mtDNA and the production of linear monomer mtDNA molecules (Figure 4, lane "RusA"). The depletion of supercoiled circular DNA may be explained by nicking of circular mtDNA during resolvase treatment (Materials and Methods). The generation of linear molecules by RusA treatment suggests the presence of tandemly repeated multimeric mtDNA molecules. While bands of electrophoretic mobility larger than the open circular form are absent from Figure 4A, molecules of multiple genome lengths would be refractory to fractionation in 0.4% agarose, and likely remain in the vicinity of the well. To test whether molecules greater than 13.8kb mtDNA were present, we examined a longer exposure of the blot in

Figure 4A, specifically focusing on the well region (Figure 4B). Indeed, high molecular weight forms ("HMW") were now observed and their abundance as judged by relative hybridization signal intensity was altered by treatment with RusA.

Linear mtDNA products were also generated by T7 endonuclease with efficiency equal to RusA; this enzyme specifically, targets generic cruciforms and nucleotide mismatches within dsDNA. This result strongly suggests that a large subset of mtDNAs contain true Holliday junctions, as opposed to hairpin secondary DNA structures, which may be cleaved by T7, but not *RusA*.

Lastly, to confirm the identity of the fast migrating circular mtDNA band, we treated *C. elegans* mtDNA with Topoisomerase I, which relaxes negative supercoiling and compared the qualitative effect of this treatment with that of Topoisomerase IV, which decatenates dsDNA circles. Topoisomerase I treatment resulted in the enhancement of autoradiographic signal from the slower migrating band which we infer contains relaxed circular monomers, with a concomitant loss of signal from the fast migrating mtDNA band. In addition, we observed an increase in the frequency of mtDNA molecules in the linear monomer form following topoisomerase treatment, which we attribute to artifactual nicking of supercoiled mtDNA during the one-hour incubation at 37 degrees Celsius.

Discussion

Here we show that the mitochondrial genome of *C. elegans* is a supercoiled circle in its non-replicative monomer form. Animal mitochondrial DNA topology has been most thoroughly studied in human cultured cells, mouse and fruitfly, all of which share a circular topology with *C. elegans* (Falkenberg et al. 2007; Jõers and Jacobs 2013). Unlike vertebrate systems, mtDNA synthesis in non-vertebrate animals is poorly understood (Gissi et al. 2008). However, now that the topology of mtDNA in the model nematode is confirmed to be circular in this study, we may infer some basic characteristics of mtDNA maintenance in this organism. No matter what method of mtDNA synthesis may be employed in the nematode, termination and disjunction must ultimately produce monomeric circular molecules as our electron micrographs show predominantly non-replicating unit circles.

Furthermore, our finding that mitochondrial nucleic acid consists primarily of dsDNA without substantial RNA tracts suggests that the RITOLS mtDNA replication mechanism is absent from *C. elegans*, though more detailed analyses of replication fork structure is needed, and detailed characterization of *C. elegans* replication intermediates is found in the next chapter.

Electron microscopy (Figure 3) revealed regions in some mitochondrial genome circles that are bound by *E. coli* SSB, indicating the presence of ssDNA gaps in

the double stranded mtDNA molecules; the genesis of those regions is not clear. The strand-displacement model of mtDNA replication predicts daughter molecules containing ssDNA gaps, which in mammalian species may be of substantial length (Gillum and Clayton 1978). However, it is possible that the conditions of DNA treatment and spreading for TEM may influence regional propensity for DNA denaturation, and especially so within a molecule of over 80% A+T (Okimoto 1992; Thresher and Griffith 1992). As such, the ssDNA tracts reported here may be artifactual. The possible single-stranded nature of the lagging strand during *C. elegans* mtDNA replication is addressed definitively in Chapter 4.

Our discovery that *C. elegans* mtDNA is sensitive to Holliday junction resolvases presents an exciting avenue for future study, as it strongly suggests the occurrence of mtDNA recombination as a component of the mtDNA replication process. This possibility is also addressed in the next chapter. The nematodes employed in this study were young adult wildtype animals of the ubiquitous ‘N2 Bristol’ strain cultured under standard conditions, indicating that mtDNA recombination in this nematode cannot be attributed to the consequences of aging or a response to severe stress, as is true for recombination events in human mitochondria.

Materials and Methods

Caenorhabditis elegans strain N2 Bristol was obtained from the *Caenorhabditis* Genetics Center (CGC; Minneapolis, USA) and maintained as described (Stiernagle 1999). For preparation of intact mitochondria, nematodes were collected, dounce-homogenized and subjected to differential centrifugation to generate a crude mitochondria-enriched fraction; further enrichment was achieved by sucrose-step gradient centrifugation (1:1.3 M sucrose), all procedures were conducted at 4°C (Jõers and Jacobs 2013). Freshly isolated mitochondria were lysed with SDS in the presence of Proteinase K. Following phenol-chloroform extraction; mitochondrial nucleic acid was precipitated with ethanol in the presence of sodium acetate, washed in ethanol, and re-suspended in Tris-EDTA pH 7.6 for further manipulation.

Treatments (restriction endonucleases, RNases, nucleases) were executed according to manufacturer instructions; RusA treatment was as described (Bolt and Lloyd 2002), with the addition of 10uM $MgCl_2$ to the reaction buffer.

Gel electrophoresis and Southern hybridization was performed as described (Jõers and Jacobs 2013); for all Southern hybridization panels, one-hour exposures are presented. Ethidium bromide stained agarose gels and Southern hybridizations in Figures 1 and 4 are representative of three biological replicates and two technical replicates each.

For transmission electron microscopy, aliquots of RNase I-treated mitochondrial nucleic acid were mounted directly on parlodion-coated copper grids and co-

spread with 5kb dsDNA pgLGAP plasmid. Mounting followed the Kleinschmidt method and imaging was performed as described (Griffith and Christiansen 1978; Thresher and Griffith 1992). Of the 1104 circular DNA molecules imaged by TEM, 50 were chosen at random and measured with Gatan ElectronMicrograph© three times each to generate the average mtDNA size value reported herein.

References

Addo MG, Cossard R, Pichard D, Obiri-Danso K, Rötig A, Delahodde A. 2010. Caenorhabditis elegans, a pluricellular model organism to screen new genes involved in mitochondrial genome maintenance. Biochim. Biophys. Acta 1802:765–773.

Armstrong MR, Blok VC, Phillips MS. 2000. A multipartite mitochondrial genome in the potato cyst nematode Globodera pallida. Genetics 154:181–192.

Bolt EL, Lloyd RG. 2002. Substrate specificity of RusA resolvase reveals the DNA structures targeted by RuvAB and RecG in vivo. Mol. Cell 10:187–198.

Bratic I, Hench J, Trifunovic A. 2010. Caenorhabditis elegans as a model system for mtDNA replication defects. Methods 51:437–443.

Clayton DA. 1982. Replication of animal mitochondrial DNA. Cell.

Falkenberg M, Larsson N-G, Gustafsson CM. 2007. DNA Replication and Transcription in Mammalian Mitochondria. Annu. Rev. Biochem. 76:679–699.

Gaj T, Gersbach CA, Barbas CF III. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology:1–9.

Gillum AM, Clayton DA. 1978. Displacement-loop replication initiation sequence in animal mitochondrial DNA exists as a family of discrete lengths. Proc. Natl. Acad. Sci. U.S.A. 75:677–681.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301–320.

Griffith JD, Christiansen G. 1978. Electron microscope visualization of chromatin and other DNA-protein complexes. Annual Review of Biophysics and Bioengineering:19–35.

Holt IJ, Reyes A. 2012. Human Mitochondrial DNA Replication. Cold Spring Harbor Perspectives in Biology 4:a012971–a012971.

Hyman BC, Lewis SC, Tang S, Wu Z. 2011. Rampant gene rearrangement and haplotype hypervariation among nematode mitochondrial genomes. Genetica 139:611–615.

John JCS, Facucho-Oliveira J, Jiang Y, Kelly R, Salah R. 2010. Mitochondrial DNA transmission, replication and inheritance: a journey from the gamete

through the embryo and into offspring and embryonic stem cells. Human Reproduction Update 16:488–509.

Jõers P, Jacobs HT. 2013. Analysis of replication intermediates indicates that Drosophila melanogaster mitochondrial DNA replicates by a strand-coupled theta mechanism. PLoS ONE 8:e53249.

Lemire B. 2005. Mitochondrial genetics. WormBook:1–10.

Lunt DH, Hyman BC. 1997. Animal mitochondrial DNA recombination. Nature 387:247–247.

Nunnari J, Suomalainen A. 2012. Mitochondria: in sickness and in health. Cell 148:1145–1159.

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, Caenorhabditis elegans and Ascaris suum. :471–498.

Pohjoismäki JLO, Goffart S, Taylor RW, Turnbull DM, Suomalainen A, Jacobs HT, Karhunen PJ. 2010. Developmental and pathological changes in the human cardiac muscle mitochondrial DNA organization, replication and copy number. PLoS ONE 5:e10426.

Pohjoismäki JLO, Goffart S, Tyynismaa H, et al. 2009. Human heart mitochondrial DNA is organized in complex catenated networks containing abundant four-way junctions and replication forks. J. Biol. Chem. 284:21446–21457.

Pohjoismäki JLO, Holmes JB, Wood SR, et al. 2010. Mammalian Mitochondrial DNA Replication Intermediates Are Essentially Duplex but Contain Extensive Tracts of RNA/DNA Hybrid. Journal of Molecular Biology 397:1144–1155.

Pont-Kingdon G, Vassort CG, Warrior R, Okimoto R, Beagley CT, Wolstenholme DR. 2000. Mitochondrial DNA of Hydra attenuata (Cnidaria): A Sequence That Includes an End of One Linear Molecule and the Genes for l-rRNA, tRNAf-Met, tRNATrp, COII, and ATPase8. J Mol Evol 51:404–415.

Reyes A, Kazak L, Wood SR, Yasukawa T, Jacobs HT, Holt IJ. 2013. Mitochondrial DNA replication proceeds via a “bootlace” mechanism involving the incorporation of processed transcripts. Nucleic Acids Research 41:5837–5850.

Reyes A, Yang MY, Bowmaker M, Holt IJ. 2005. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. Journal of Biological

Chemistry.

Stiernagle T. 1999. Maintenance of C. elegans. C. elegans: a practical approach

Thresher R, Griffith J. 1992. Electron microscopic visualization of DNA and DNA-protein complexes as adjunct to biochemical studies. Meth. Enzymol. 211:481–490.

Yasukawa T, Yang MY, Jacobs HT, Holt IJ. 2005. A Bidirectional Origin of Replication Maps to the Major Noncoding Region of Human Mitochondrial DNA. Mol. Cell 18:651–662.

WormBook : the online review of C. elegans biology.

Table 1. Qualitative summary of *C. elegans* mtDNA endonuclease sensitivity.

Treatment	**Action**	**Sensitivity**
RNase H	Specifically degrades RNA-DNA hybrids	-
S1 nuclease	Specifically degrades ssDNA	-
RusA resolvase	Degrades Holliday junctions, but not non-Holliday cruciforms	++
Topoisomerase I	Removes negative supercoils from dsDNA	+
Topoisomerase IV	Decatenation of dsDNA, including termination intermediates following DNA replication	+
T7 endonuclease	Cleavage of cruciforms, including Holliday junctions, heteroduplex and imperfectly-matched DNA	+++

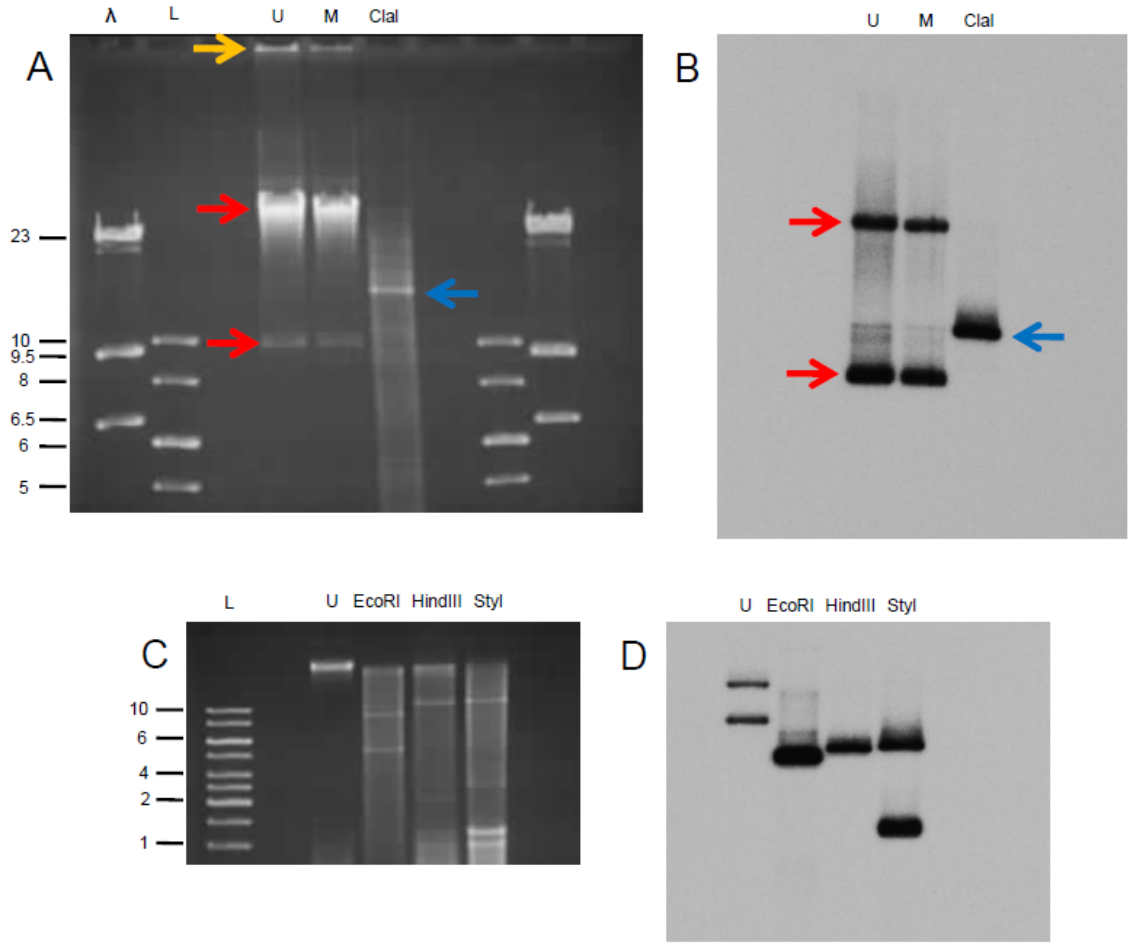

Figure 1. Restriction analysis of *C. elegans* mtDNA. (A) Enriched mitochondrial nucleic acid was fractionated on a 0.5% agarose gel without treatment, lane“U”; following mock digestion by incubation at 37 degrees C 1 hour, lane “M”; following cleavage by *ClaI*, “ClaI”. Samples were co-fractionated with both a Lambda/HindIII linear standard (λ, outermost) and a 1kb linear DNA standard (“L”, innermost); ladder fragment sizes as indicated at left. The yellow arrow indicates well-bound DNA, while red arrows indicate putative mitochondrial DNA, and the blue arrow indicates the prominent restriction product (see text). (B) Southern hybridization of the agarose gel in panel A probed with the sequence of the mitochondrial gene *ND5*. For genomic probe location, see Figure 2. Arrows as in panel A. (C) Enriched mitochondrial nucleic acid was

fractionated on a 0.8% agarose gel either without treatment (“U”) or following restriction enzyme cleavage by *EcoRI*, *HindII*I or *StyI*, respectively. All enzymes cleave twice in the mtDNA. Samples were co-fractionated with a 1kb linear DNA standard. (D) Southern hybridization of the agarose gel in panel C probed with the same mtDAN sequence as panel B. For EcoRI and HindIII, only a single band is evident due to the absence of the probe hybridization sequence from faster migrating bands. For sample StyI, two bands are visible, as one restriction site cleaves inside the *ND5* gene.

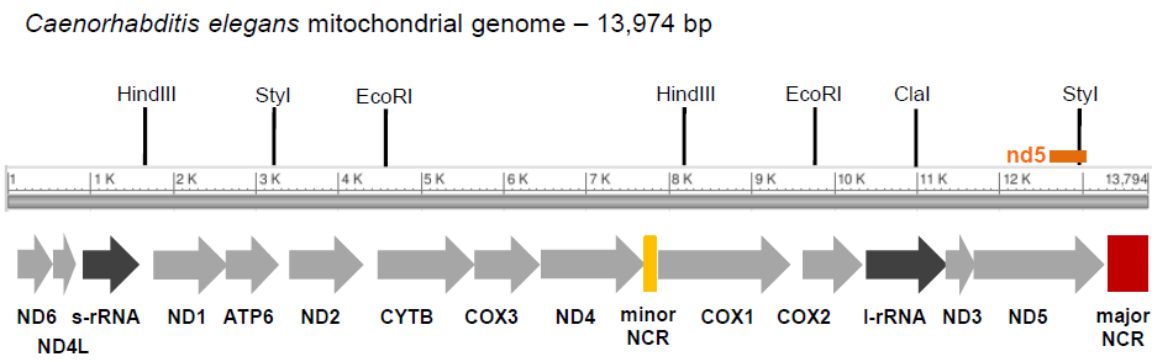

Figure 2. *C. elegans* mtDNA architecture. Physical and transcriptional map of the *Caenorhabditis elegans* mitochondrial genome. Arrows depict transcriptional orientation and relative lengths of protein coding genes (light grey arrows), large and small ribosomal RNA genes (dark grey arrows), and two non-coding regions (yellow and red bars). ND5 hybridization probe location is indicated by the orange bar.

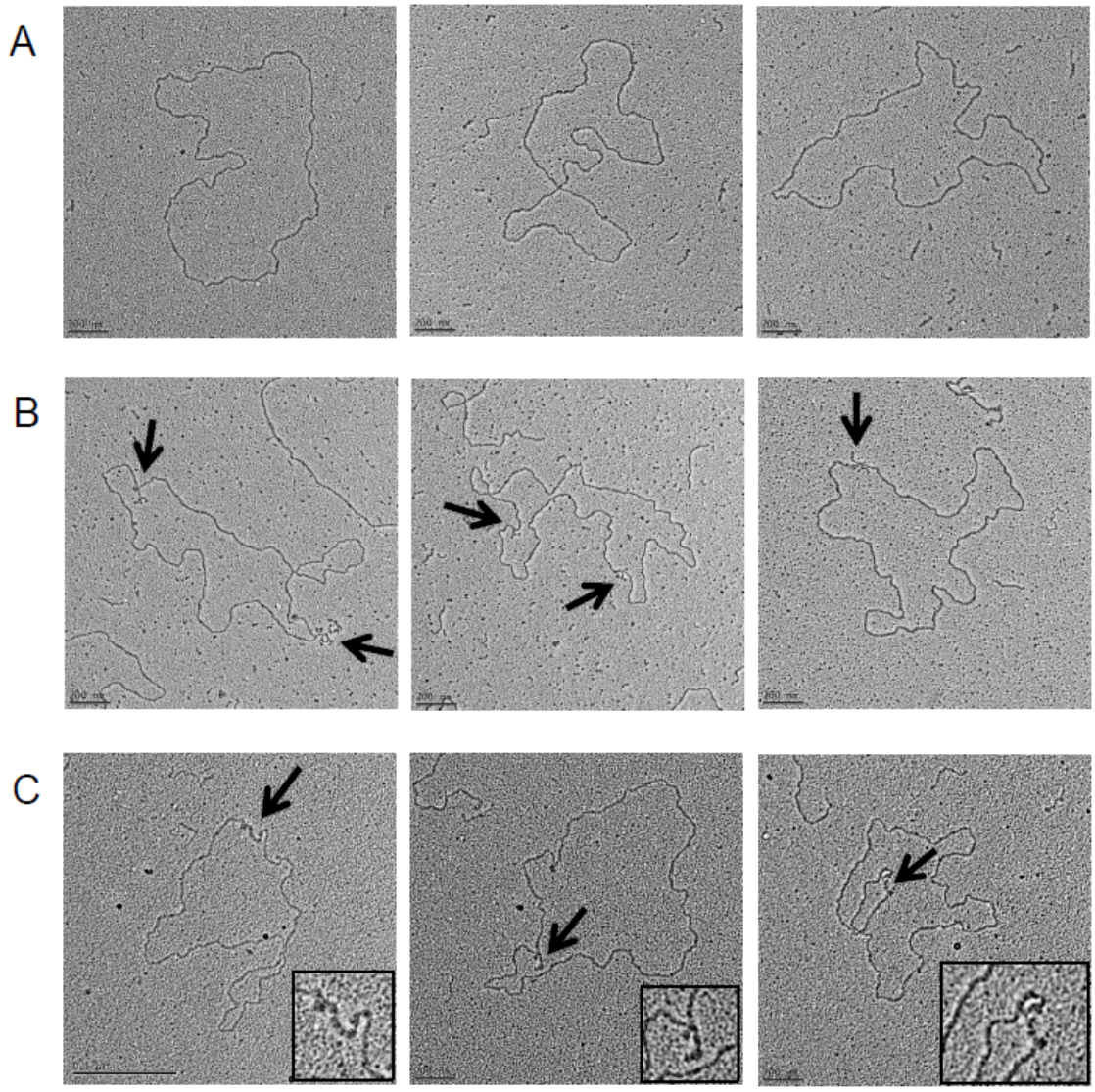

Figure 3. Direct observation of circular mitochondrial DNA. RNase I-treated mitochondrial nucleic acid was spread on parlodion-coated copper grids and visualized by TEM. (A) Fully double-stranded DNA circles. (B) DNA circles with regions of complex secondary structure (black arrows). (C) DNA circles incubated with the *E. coli* single-stranded binding protein (SSB) display thickened regions consistent with protein binding (black arrows); SSB regions inset at twice magnification. Scale bars, 0.2 μm. Data generated from three independent mtDNA isolations.

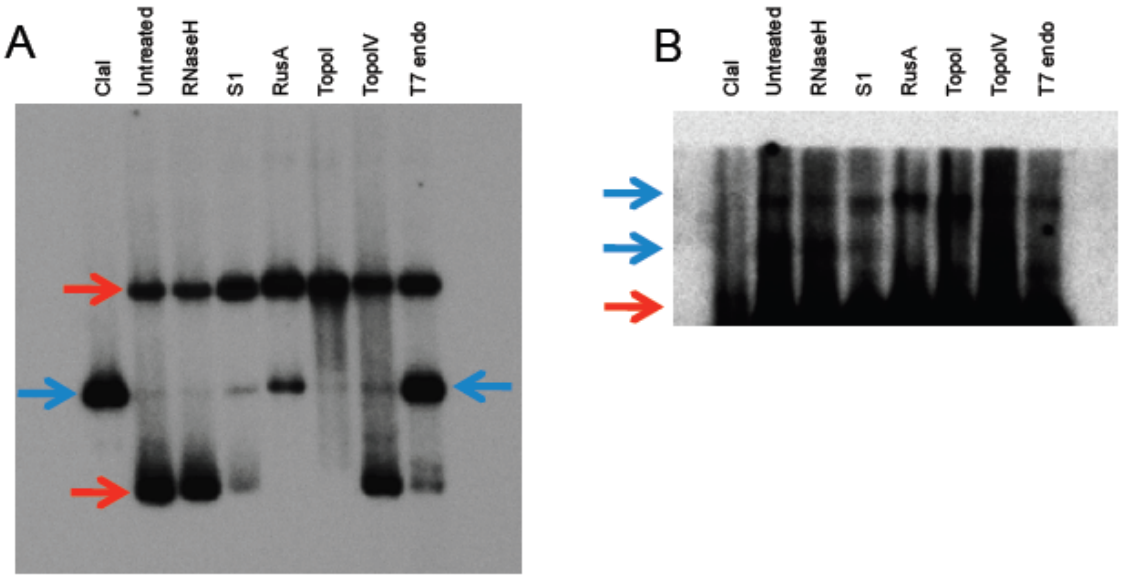

Figure 4. Topological analysis of endonuclease-treated *C. elegans* mtDNA. Enriched mitochondrial nucleic acid was fractionated on a 0.4% agarose gel following indicated enzyme treatments followed by Southern blot hybridization employing an *ND5* probe to specifically detect mtDNA as in Figure 1. (A) Blue arrows indicate bands predicted to be 13.8 kb linearized; red arrows indicate slow and fast migrating circular mtDNA. Note the increase in linear mtDNA after RusA Holliday junction resolvase treatment (lane designated RusA), or T7 endonuclease (rightmost lane); Image represents a two-hour autoradiographic exposure. (B) An 22 hour exposure of the gel well region from the blot described in (A). Note the hybridization signal from high molecular weight (HMW, blue arrows) mtDNA forms. RusA treatment reduces a high molecular weight band (lane, "RusA"). The red arrow marks the position of the open circular form; the lower portion of the blot has been cropped due to overexposure.

Chapter 4. Rolling circle replication of *C. elegans* mitochondrial DNA generates concatemeric lariat structures *in vivo*

Introduction

Perpetuation of respiratory competent mitochondria is essential to life in most eukaryotes, and requires faithful duplication of the mitochondrial chromosome (Falkenberg et al. 2007). Early studies of mitochondrial DNA (mtDNA) replication in mammalian cultured cells supported a unidirectional strand displacement or 'asymmetric' model of animal mtDNA replication, producing partially single-stranded-DNA (ssDNA) intermediates (Holt and Reyes 2012). Strand-coupled 'theta' replication has also been proposed (Holt et al. 2000). In addition, significant support has been amassed for a temporally synchronous mode of replication involving provisional RNA incorporation throughout the lagging strand (RITOLS), a model in which expanding double-stranded-DNA replication bubbles indicate origin activity (Reyes et al. 2013).

All previously described animal mtDNA replication models share two features. First, initiation of replication relies on elongation from a transcript-primed displacement loop (D-loop). Second, each successful synthesis cycle from the circular template results in but two circular daughter molecules. Previous work on mtDNA synthesis has focused primarily on mammalian species; mtDNA maintenance elsewhere in the animal lineage remains poorly understood.

Caenorhabditis elegans has proven an attractive model for studies of aging and metabolic disease, processes intimately associated with mitochondrial

health, yet little is known of mtDNA maintenance in this organism. MtDNA is required for nematode development beyond the early larval stages, and perturbations that result in mtDNA depletion during embryonic development commonly result in larval arrest (Lemire 2005). The mitochondrial complement in somatic cells of the adult nematode largely results from distribution of oocyte mtDNA molecules during development, segregation which does not require the mtDNA polymerase POLG-1 (Bratic et al. 2009). Thus mtDNA replication is expected to occur primarily in the adult gonad.

C. elegans mtDNA harbors two non-coding regions (NCRs), delimiting coding regions of 5.5 and 7.7 kb respectively (Figure 1A; (Okimoto et al. 1992). By analogy with the mammalian mtDNA organization, both NCRs are proposed to play a role in *C. elegans* mtDNA replication, one as the first-strand origin (akin to the mammalian D-loop) and the other serving as a second-strand origin (Lemire 2005; Bratic et al. 2010). To test this assumption, we investigated the mechanism of *C. elegans* mtDNA replication *in vivo* and the possible function of the two NCRs therein.

Results

Non-coding regions lack bubble origin activity

To determine if either NCR could function as a replication origin, we examined mtDNA fragments containing each one of the two NCRs for origin activity using two-dimensional neutral agarose gel electrophoresis (2DNAGE) (Lorimer 2002). Y arcs formed by progressing forks, as well as cruciform structures, were readily apparent (Figure 1B and 1C, charateristic arcs depicted in 1E). Analysis of replication intermediates (RIs) derived from restriction fragments lacking both NCRs also revealed full Y arcs and cruciforms, consistent with active replication of the entire mtDNA (Figure 1D). However, a bubble arc indicative of theta type replication initiation was not detected from any region of the genome, even after long autoradiographic exposures (Figure S1A-E).

We next considered that first-strand replication initiation might occur from more than one site in the genome. Such low frequency bubble intermediates may go undetected by fragment 2DNAGE, as only a subset of intermediates are analyzed in each experiment (Lorimer 2002; Holt and Reyes 2012; Reyes et al. 2013). To address this possibility, we performed 2DNAGE analysis after digestion with restriction enzymes cutting only once in the mitochondrial genome. We reasoned that analysis of RIs spanning the complete mtDNA would pool molecules containing D-loop initiation structures along a single arc regardless of initiation site, facilitating detection. Consonant with our 2DNAGE data on sub-

genomic fragments, linearization and subsequent 2DNAGE of the full-length genome demonstrated clear Y and X shaped intermediates (Figure S1F-J), yet no bubble arc was observed. We therefore conclude that initiation of *C. elegans* mtDNA synthesis does not involve the formation of a bubble intermediate at levels detectable by blot-hybridization.

Bioinformatic analysis of cumulative GC skew in *C. elegans* mtDNA revealed a profile absent of local minima or maxima, considered the molecular signatures of origin and termination activity respectively (Figure S1K; (Touchon and Rocha 2008). This finding is consistent with the lack of initiation bubbles; the absence of a single origin of first-strand replication effectively excludes strand-coupled initiation from a D-loop, i.e. the theta replication mode (Holt and Reyes 2012).

Strand-synchronous replication produces dsDNA replication intermediates

We next investigated whether *C. elegans* mtDNA RI structure was consistent with the asymmetric strand-displacement or RITOLS models of mtDNA replication. Temporally asynchronous replication of the two template strands is predicted to generate partially single-stranded RIs (Holt and Reyes 2012)and references therein). In 2DNAGE, such ssDNA regions block endonuclease cleavage, producing slow-moving Y arcs greater than twice the unit length fragment size, and render RIs sensitive to the action of the single-strand specific nuclease S1 (Holt and Reyes 2012). In contrast, RNA:DNA

hybrid-containing RIs typical of the RITOLS mode can be detected based on their sensitivity to degradation by RNase H, which exposes ssDNA regions sensitive to S1 nuclease (Yasukawa et al. 2005; Pohjoismäki et al. 2010). Thus these treatments are expected to dramatically alter the electrophoretic migration properties of RITOLS intermediates in 2D gels (Yasukawa et al. 2006; Holt and Reyes 2012). These intermediates represent a transient step in replication, preceding the synthesis of the DNA of the definitive lagging strand.

Strand-displacement and/or RITOLS intermediates were identified by systematic treatment of mtDNA fragments representing the complete mitochondrial genome with S1 nuclease, RNase H, or RNase H followed by S1 nuclease, followed by 2DNAGE analysis. For all mtDNA fragments analyzed, Y arc and cruciform spike migration patterns persisted treatment with either S1 or RNase H alone (Figure 2A-C, see also Figure S2). The Y arc hybridization signal was somewhat affected by RNase H exposure, yet the majority of fragment RIs remained following subsequent treatment with S1 nuclease (Figure 2D; see quantification in Figure 2E).

These experiments indicate that *C. elegans* RIs lack the extensive ssDNA character expected from strand-displacement synthesis. Furthermore, slow-moving Y arcs were not observed, and depletion of the Y arc signal after treatment with RNase H and S1 was no greater than with RNase H treatment alone (Figure 2E). Thus RNase H failed to ‘unmask’ substantial ssDNA regions, as would be expected if extensive RNA:DNA hybrid tracts were present. These

data are consistent with synchronous (or very near-synchronous) replication of the two mtDNA strands independent of a D-loop, and eliminate from further consideration both asymmetric and RITOLS-mode strand-displacement replication.

mtDNA replication intermediates are branched-circular lariats with concatemer synthesis products

The absence of theta form, RITOLS and partially ssDNA strand-displacement intermediates led us to consider alternate DNA replication mechanisms. The detection of Y arcs, but not bubble arcs, by 2DNAGE of fragments derived from a circular template is consistent with a rolling circle replication (RCR) mechanism (Han and Stachow 1994; Belanger et al. 1996). According to the RCR model, sustained elongation on a circular template produces linear DNA molecules greater than template unit-length that may become resolved to monomers in a variety of ways, or remain concatemeric linear networks (Kornberg and Baker 1992). A central prediction of the RCR model is the presence of “lariat” DNA forms *in vivo*. For *C. elegans* mtDNA, we hypothesized the occurrence of one-genome unit length circular templates, from which multimeric linear tails would extend. Alternatively, a second replication mode can be envisaged that would involve strand-invasion of a template by linear molecules, as occurs in some bacteriophages and the mtDNA of the

fungus *Candida albicans* (Kreuzer 2000; Gerhold et al. 2010). This alternative would predict Y-form RIs in the absence of bubble RIs, but not lariat structures.

To determine whether lariat molecules consistent with rolling circle intermediates were present, we directly examined *C. elegans* mtDNA using transmission electron microscopy (TEM). We observed both circular and branched-circular lariat molecules (Figure 1, Table S1). Most prominent were dsDNA circles with a mean measured length of 13.61 kb +/- .407 kb, consistent with the sequenced mitochondrial genome size of 13.794 kb (Okimoto et al. 1992) (Figure 3A, Figure 3B). Though the *C. elegans* mtDNA restriction map has long been thought to be circular (Okimoto et al. 1992), to our knowledge this is the first evidence that non-replicative *C. elegans* mtDNA exists in a topologically circular form. As predicted by our 2DNAGE and bioinformatic analyses, none of these circular mtDNAs contained a visible displacement loop. Lariats with linear tails ranging from < 1 kb to 48.2 kb in length, more than three genome units, were the next most frequently observed class of molecules (Figure 3C and 3D, Table S1). The mean length of the lariat circular portion measured 13.64 kb (Figure 3A). Fifty-six percent of lariat molecules appeared fully double-stranded at the circle-branch junction (Figure 3E and 3F), though ssDNA tracts of less than ~500 bases are not readily visible under the imaging conditions used.

A subset of lariat molecules contained visible interspersed regions of collapsed secondary structure, considered diagnostic for ssDNA (Figure 3G) (Griffith and Christiansen 1978). Within individual molecules, these regions

occurred at the junction of the circular and linear branch portions of lariats, further along the linear branch only, or in both locations (Figure 3H and 3I). Neither strand-displacement nor theta RIs were observed among the 1262 molecules analyzed by TEM, while lariats made up approximately 4% of mtDNA molecules, in line with previous reports of the proportion of replicating mtDNA in mammalian cells and *Drosophila melanogaster* (Falkenberg et al. 2007; Jõers and Jacobs 2013).

The major non-coding region is a hotspot for homologous recombination

The high frequency of non-replicative circular monomers we detected by TEM suggested conversion between the circular monomer and the lariat mtDNA forms. In the course of fragment 2DNAGE, we noted that the hybridization intensity of cruciform structures was most prevalent, relative to the monomer spot, in fragments harboring the 465 bp 'major' NCR (compare Figure 1B to Figure 1C and 1D). This observation suggested that site-specific cruciform structure formation could play a role in the maintenance *C. elegans* mtDNA and/or the production of monomer circles (Kornberg and Baker 1992; Backert 2002).

We further addressed cruciform architecture by 2DNAGE following treatment with RusA, an *Escherichia coli* resolvase highly specific for Holliday junctions (HJ) (Chan et al. 1997; Bolt and Lloyd 2002). This analysis was performed using RusA alone or in combination with S1 nuclease. RusA treatment

reduced the hybridization signal of the cruciform spike by 48 % relative to the untreated control (Figure 4A and 4B, see also 4G) while S1 nuclease alone had no significant effect (Figure 4C). Treatment with S1 after RusA further reduced the cruciform signal, and revealed a subclass of cruciforms resistant to both RusA and S1, that persisted after S1 nuclease treatment (Figure 4D). These molecules formed a just off-vertical 2DNAGE spike (Figure 4E), a migration pattern typical of hemicatenanes (Lopes et al. 2003). The collapse of adjacent HJs commonly results in the production of resolvase-resistant hemicatenanes (represented in Figure 4F) that migrate in 2D gels as cruciforms, further supporting the presence of a recombination hotspot within the major NCR (Lucas and Hyrien 2000; Lopes et al. 2003). These data provide strong enzymatic evidence for homologous recombination of *C. elegans* mtDNA *in vivo*.

Discussion

Taken together, our findings indicate that synthesis of *C. elegans* mtDNA proceeds by rolling circle replication. We propose that multiple replication cycles on single template circles generate lariat structures with multimer tails composed primarily, although not entirely, of dsDNA and resolved to monomers through recombination events. In the context of a rolling circle, each initiation event will give rise to multiple genome units, making the identification of a specific start site challenging. Neither the biochemical nor microscopic methods used here

revealed a specific site of replication initiation, nor did our bioinformatic analysis identify molecular signatures of site-specific initiation (Touchon and Rocha 2008). It remains possible that replication initiates from a site-specific nick in the mtDNA, and further investigation is needed.

The sporadic ssDNA regions observed along lariat molecules (Figure 3H) are consistent with yet-to-be-ligated gaps between lagging-strand synthesis products. Similar to the mechanism of elongation in the T7 bacteriophage rolling-circle synthesis model (Griffith and Christiansen 1978; Park et al. 1998), looping at the lariat circle-to-tail junction (Figure 3I) would then suggest a single replisome engaged in coordinate synthesis of the leading and lagging strands, as confirmed by the 2DNAGE data presented here (Figure 2). This work indicates that while mechanistically variable, metazoan mtDNA synthesis favors DNA protection from a single-stranded state, with the decided advantage of minimizing damage to vulnerable ssDNA by the oxidative environment of the mitochondrion.

Several mtDNA maintenance factors are highly homologous to bacteriophage proteins, including the mtDNA polymerase POLGA, RNA polymerase POLRMT, and TWINKLE helicase; moreover, antiviral drugs are commonly mitotoxic (Falkenberg et al. 2007). These observations raise the possibility of a phage-like mtDNA maintenance mechanism in the ancestral endosymbiont (Falkenberg et al. 2007). Our identification of RCR in the nematode, a basal animal, may then implicate strand-displacement and RITOLS as derived mechanisms.

RCR has been proposed as a mechanism of fungal and plant mtDNA maintenance. Our findings differ from these descriptions in that the monomer circle remains the most common mtDNA topology. Among the fungi, RCR of *Saccharomyces cerevisiae* mtDNA may produce linear molecules and the mtDNA of *Candida albicans* forms concatameric linear networks (Ling and Shibata 2002; Gerhold et al. 2010). Chloroplast genomes are particularly complex, often consisting of a mix of branched, linear and circular topologies that may be many genome multimers in size, rendering monomer circles a rare occurrence (Bendich 1993). The C. elegans mtDNA replication mechanism is distinct from models of bacteriophage replication as well, due to the apparent absence of linear mtDNA monomers as bacteriophage DNA replication produces linear monomers with either distinct (phage T7) or permuted (phage T4) ends (Kornberg and Baker 1992; Park et al. 1998).

It had been assumed that *C. elegans* mtDNA adheres to the strand-displacement replication mechanism in which the NCRs contain first- and second-strand origins; here we demonstrate that this cannot be the case. Rather, the HJ intermediates coincident with the major NCR are the only structures specific to either of the NCRs. While the details of concatemer resolution remain to be determined, we suggest that the identified recombination intermediates may represent termination/resolution structures, in which strand-invasion or branch migration arrest occurs in a site-specific manner, facilitated by the short repeats present in the major NCR region (Okimoto et al. 1992). Invasion by the

unreplicated ssDNA 3' end of the lariat tail at the major NCR sequence would create a recombination junction (Figure 4H, i). Subsequent migration of the junction would provide an opportunity for formation of a second HJ (Figure 4H, ii). Cleavage and HJ resolution (Figure 4H, iii) would generate a gapped circular monomer and lariat tail with 3' overhang, one genome unit-length shorter (Figure 4H, iv). Such a mechanism could also produce rare uni-circular multimers (see Table S1) in which snapback does not occur at an adjacent concatemer.

This mode of resolution, while speculative, is consistent with two intriguing aspects of our results. First, a minority of Y arc RIs are sensitive to degradation by RusA (compare Figure 4A and 4B); these are consistent with recombination/resolution intermediates excised by the restriction digest, in which the arms of the Y are joined by HJs. Second, as noted our 2DNAGE analysis demonstrated the presence of major NCR-specific hemicatenanes, as expected to form by convergence of adjacent HJs. We did not observe molecules simultaneously involved in elongation and resolution by TEM, as would be consistent with this model. However, we note that if present *in vivo*, such molecules could possibly exceed 75 kb in size and are thus likely fragile to DNA isolation techniques.

Whether RCR occurs elsewhere in the animal lineage remains to be explored. Unlike previous studies characterizing mtDNA RIs that involved transcriptional organizations in which both strands encode mitochondrial genes, all protein-coding genes are transcribed from one *C. elegans* mtDNA strand

(Figure 1), thus raising the intriguing possibility that RCR may be linked to the transcriptional architecture of the mitochondrial genome. Fortunately, the Nematoda are an excellent model system in which to test such hypotheses. The mtDNAs of many nematode species have been sequenced, revealing considerable architectural variation and enabling comparative studies in which the mtDNAs differ by gene inversion, scrambled gene orders, or translocation of the major NCR in the genome map (Gissi et al. 2008). As such, the phylum presents a powerful new model system for probing the relationship between mitochondrial genome architecture and replication mode.

C. elegans itself offers a compelling new model for the study of rolling circle replication in animal cells from both mechanistic and genetic perspectives. We anticipate that the future characterization of factors involved in this process will provide many new insights into animal mtDNA maintenance and the evolution of replication mechanisms.

Materials and Methods

Nematode maintenance and mtDNA isolation. *Caenorhabditis elegans* strain N2 Bristol was obtained from the *Caenorhabditis* Genetics Center (CGC; Minneapolis, USA) and maintained as described (Stiernagle 1999). For preparation of intact mitochondria, nematodes were collected, dounce-homogenized and subjected to differential centrifugation to generate a crude mitochondria-enriched fraction; further enrichment was achieved by sucrose-step gradient centrifugation (1:1.3 M sucrose), all procedures were conducted at 4°C (Jõers and Jacobs 2013). Freshly isolated mitochondria were lysed with SDS in the presence of Proteinase K. Following phenol-chloroform extraction, mitochondrial nucleic acid was precipitated with ethanol in the presence of sodium acetate, washed in ethanol, and re-suspended in Tris-EDTA pH 7.6 for further manipulation.

Enzymatic treatment. Treatments (restriction endonucleases, RNases, nucleases) were executed according to manufacturer instructions; RusA treatment was as described (Bolt and Lloyd 2002).

2D neutral agarose gel electrophoresis and blot hybridization. 2DNAGE, blot-transfer and probe hybridization were carried out as previously described (Lockshon et al. 1995; Jõers and Jacobs 2013); see Supplementary Table 2 for genomic locations and sequences of mtDNA probes. For all 2DNAGE panels, four-hour exposures are presented. For detailed information on probes see Supplemental Experimental Procedures.

Transmission electron microscopy. Aliquots of RNase I-treated mitochondrial nucleic acid were mounted directly on parlodion-coated copper grids following the Kleinschmidt method and imaged as described(Griffith and Christiansen 1978; Thresher and Griffith 1992). Molecule lengths were measured in Gatan DigitalMicrograph™ and calibrated by measurement of a co-spread 3.5 kb pglGAP plasmid. Each mtDNA molecule was measured 3 times to obtain mean values as reported in Fig. 3.

Bioinformatic analysis. Cumulative GC skew analysis was carried out as described(Gerhold et al. 2010) using a custom *R* script.

References

Backert S. 2002. R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. The EMBO Journal 21:3128–3136.

Belanger KG, Mirzayan C, Kreuzer HE, Alberts BM, Kreuzer KN. 1996. Two-dimensional gel analysis of rolling circle replication in the presence and absence of bacteriophage T4 primase. Nucleic Acids Research 24:2166–2175.

Bendich AJ. 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr Genet 24:279–290.

Bolt EL, Lloyd RG. 2002. Substrate specificity of RusA resolvase reveals the DNA structures targeted by RuvAB and RecG in vivo. Mol. Cell 10:187–198.

Bratic I, Hench J, Henriksson J, Antebi A, Burglin TR, Trifunovic A. 2009. Mitochondrial DNA level, but not active replicase, is essential for Caenorhabditis elegans development. Nucleic Acids Research 37:1817–1828.

Bratic I, Hench J, Trifunovic A. 2010. Caenorhabditis elegans as a model system for mtDNA replication defects. Methods 51:437–443.

Chan SN, Harris L, Bolt EL, Whitby MC, Lloyd RG. 1997. Sequence specificity and biochemical characterization of the RusA Holliday junction resolvase of Escherichia coli. J. Biol. Chem. 272:14873–14882.

Clayton, DA. 2003. Mitochondrial DNA replication: what we know. IUBMB Life Apr-May;55(4-5):213-7.

Falkenberg M, Larsson N-G, Gustafsson CM. 2007. DNA Replication and Transcription in Mammalian Mitochondria. Annu. Rev. Biochem. 76:679–699.

Gerhold JM, Aun A, Sedman T, Jõers P, Sedman J. 2010. Strand invasion structures in the inverted repeat of Candida albicans mitochondrial DNA reveal a role for homologous recombination in replication. Mol. Cell 39:851–861.

Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101:301–320.

Griffith JD, Christiansen G. 1978. Electron microscope visualization of chromatin

and other DNA-protein complexes. Annual Review of Biophysics and Bioengineering:19–35.

Han Z, Stachow C. 1994. Analysis of Schizosaccharomyces pombe mitochondrial DNA replication by two dimensional gel electrophoresis. Chromosoma 103:162–170.

Holt IJ, Lorimer HE, Jacobs HT. 2000. Coupled Leading- and Lagging-Strand Synthesis of Mammalian Mitochondrial DNA. Cell 100:515–524.

Holt IJ, Reyes A. 2012. Human Mitochondrial DNA Replication. Cold Spring Harbor Perspectives in Biology.

Jõers P, Jacobs HT. 2013. Analysis of replication intermediates indicates that Drosophila melanogaster mitochondrial DNA replicates by a strand-coupled theta mechanism. PLoS ONE 8:e53249.

Kornberg A, Baker TA. 1992. DNA replication.

Kreuzer KN. 2000. Recombination-dependent DNA replication in phage T4. Trends in Biochemical Sciences 25:165–173.

Lemire B. 2005. Mitochondrial genetics. WormBook:1–10.

Ling F, Shibata T. 2002. Recombination-dependent mtDNA partitioning: in vivo role of Mhr1p to promote pairing of homologous DNA. EMBO J. 21:4730–4740.

Lockshon D, Zweifel SG, Freeman-Cook LL, Lorimer HE, Brewer BJ, Fangman WL. 1995. A role for recombination junctions in the segregation of mitochondrial DNA in yeast. Cell 81:947–955.

Lopes M, Cotta-Ramusino C, Liberi G, Foiani M. 2003. Branch Migrating Sister Chromatid Junctions Form at Replication Origins through Rad51/Rad52-Independent Mechanisms. Mol. Cell 12:1499–1510.

Lorimer HE. 2002. 2D gel electrophoresis of mtDNA. Mitochondrial DNA.

Lucas I, Hyrien O. 2000. Hemicatenanes form upon inhibition of DNA replication. Nucleic Acids Research 28:2187–2193.

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR. 1992. The mitochondrial genomes of two nematodes, Caenorhabditis elegans and Ascaris suum. :471–498.

Park K, Debyser Z, Tabor S, Richardson CC, Griffith JD. 1998. Formation of a DNA loop at the replication fork generated by bacteriophage T7 replication proteins. J. Biol. Chem. 273:5260–5270.

Pohjoismäki JLO, Holmes JB, Wood SR, et al. 2010. Mammalian Mitochondrial DNA Replication Intermediates Are Essentially Duplex but Contain Extensive Tracts of RNA/DNA Hybrid. Journal of Molecular Biology 397:1144–1155.

Reyes A, Kazak L, Wood SR, Yasukawa T, Jacobs HT, Holt IJ. 2013. Mitochondrial DNA replication proceeds via a “bootlace” mechanism involving the incorporation of processed transcripts. Nucleic Acids Research:1–14.

Stiernagle T. 1999. Maintenance of C. elegans. C. elegans: a practical approach

Thresher R, Griffith J. 1992. Electron microscopic visualization of DNA and DNA-protein complexes as adjunct to biochemical studies. Meth. Enzymol. 211:481–490.

Touchon M, Rocha EPC. 2008. From GC skews to wavelets: a gentle guide to the analysis of compositional asymmetries in genomic data. Biochimie 90:648–659.

Yasukawa T, Reyes A, Cluett TJ, Yang MY, Bowmaker M, Jacobs HT, Holt IJ. 2006. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. The EMBO Journal 25:5358–5371.

Yasukawa T, Yang MY, Jacobs HT, Holt IJ. 2005. A Bidirectional Origin of Replication Maps to the Major Noncoding Region of Human Mitochondrial DNA. Mol. Cell 18:651–662.

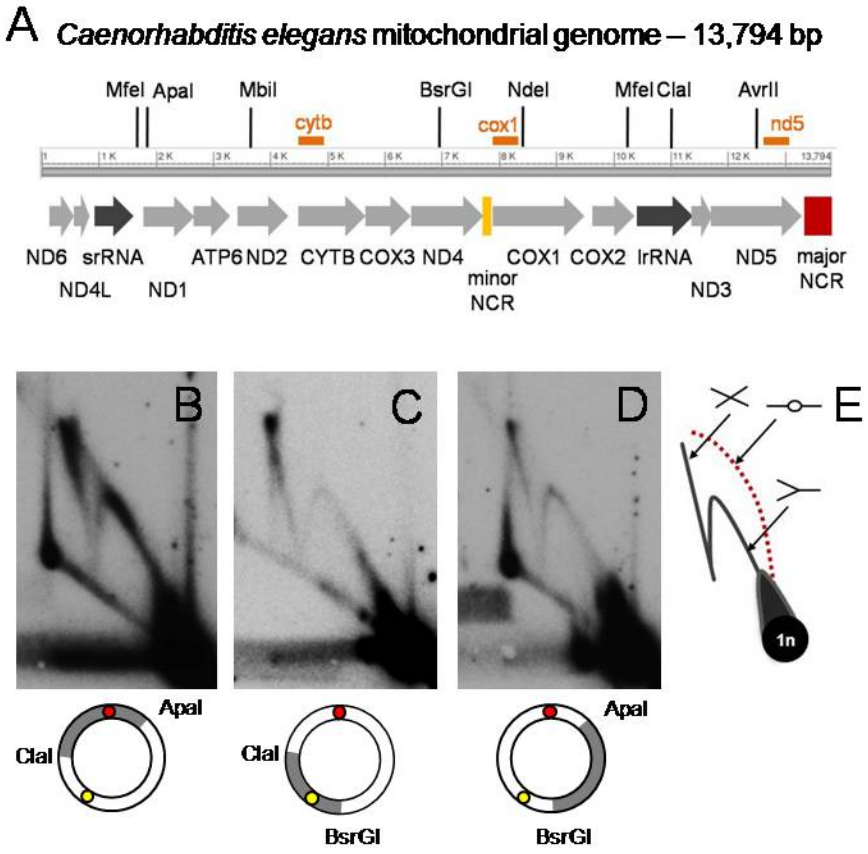
A Caenorhabditis elegans mitochondrial genome – 13,794 bp
MfeI ApaI MbiI BsrGI NdeI MfeI ClaI AvrII
cytb
cox1
nd5
ND6 srRNA ATP6 ND2 CYTB COX3 ND4 COX1 COX2 lrRNA ND5 major NCR
ND4L ND1 minor NCR ND3
B
C
D
E
1n
ClaI ApaI
ClaI BsrGI
BsrGI ApaI

Figure 1: 2DNAGE analysis of *C. elegans* mitochondrial DNA reveals prominent replication and recombination intermediates but no initiation bubbles. (A) Physical and transcriptional maps of the *Caenorhabditis elegans* mitochondrial genome. Arrows depict transcriptional orientation and relative lengths of protein coding genes (light grey arrows), large and small ribosomal RNA genes (dark grey arrows), and two NCRs (yellow and red bars). Hybridization probe locations are indicated by orange bars; see also Supplementary Experimental Procedures. MtDNA isolated from a sucrose gradient-enriched mitochondrial preparation was restriction enzyme cleaved as indicated to generate 3.8 or 5 kb fragments that collectively cover the entire genome, subjected to 2DNAGE and probed to identify intermediates containing (B) the 465 bp 'major' NCR (red circle); (C) the 104 bp 'minor' NCR (yellow circle); or a coding region only (D). Idealized replication intermediate migration within 2DNAGE gels (E) is depicted (counter-clockwise from lower right: 1n spot; Y-shaped replication forks; bubble-shaped initiation intermediates; cruciforms). Panels are representative of three independent experiments per fragment. See also Figure S1.

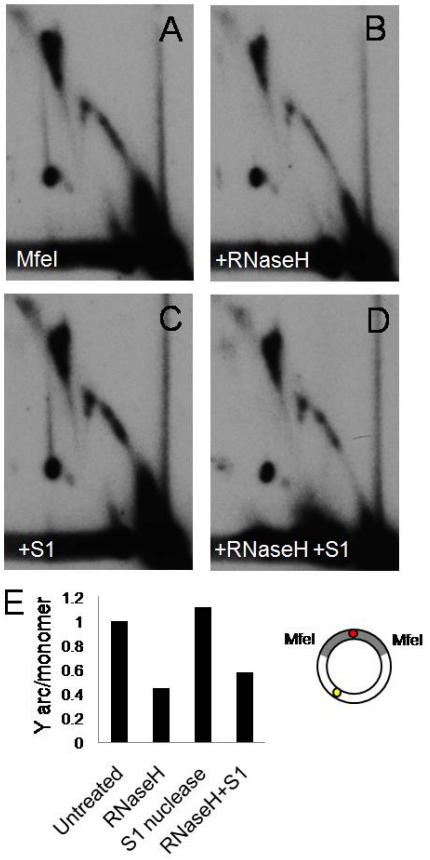

Figure 2: Advancing forks are engaged in strand-synchronous replication. MtDNA was treated with RNase H, S1 nuclease, or both, followed by 2DNAGE and blot-hybridization with a probe specific to the major-NCR. (A) 2DNAGE of untreated, MfeI-digested mtDNA. (B) RNase H treatment diminishes replication intermediates, yet a full Y arc persists. (C) Treatment by S1 nuclease does not alter migration of replication intermediates. (D) Treatment with S1 nuclease after RNase H does not unmask single- stranded DNA, consistent with the presence of double-stranded DNA intermediates. (E) ImageJ-quantified mean hybridization intensity of the Y arc derived from the major NCR-containing region, relative to the monomer spot. See also Figure S2 for analysis of fragments containing the remainder of the genome.

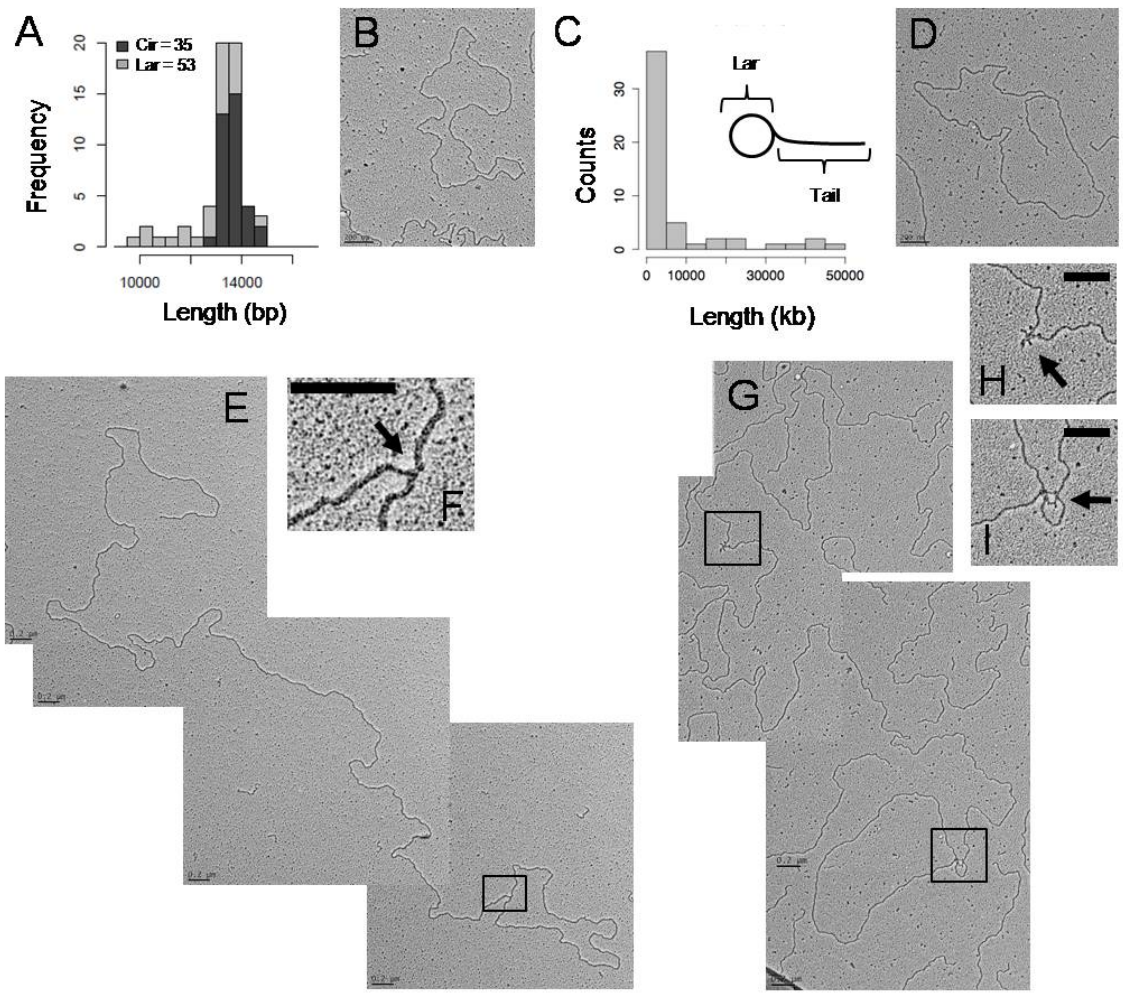
A
Cir = 35
Lar = 53
Frequency
0
5
10
15
20
10000
14000
Length (bp)
B
C
Lar
Tail
Counts
0
10
20
30
0
10000
30000
50000
Length (kb)
D
H
I
E
F
G

Figure 3: *C. elegans* mtDNA forms lariat-shaped rolling circle replication intermediates. RNase I-treated mitochondrial nucleic acid was spread on parlodion-coated copper grids and visualized by TEM. (A) Length distribution frequencies of monomeric circle mtDNAs (Cir) and lariat mtDNAs (Lar) in bp. (B) Representative electron micrograph of a double-stranded DNA monomer circle. (C) Frequency distribution of lariat tail lengths in kb. (D) Representative electron micrograph of an mtDNA lariat. (E) Multimeric lariat molecule with one genome unit-length circle (1n; 13.8 kb) with concatemeric linear tail (3.5n; 48.2 kb). (F) Higher magnification of the apparently double-stranded circle-tail junction. (G) Lariat with partial ssDNA structure consisting of a genome unit-length circle (1n; 13.8 kb) and concatemer tail (2.6n; 35.7 kb). Magnification of the single-stranded regions along the tail (H), as well as at the tail-circle junction (I); arrows indicate ssDNA. Scale bars, 0.2 μm. Data generated from three independent mtDNA isolations.

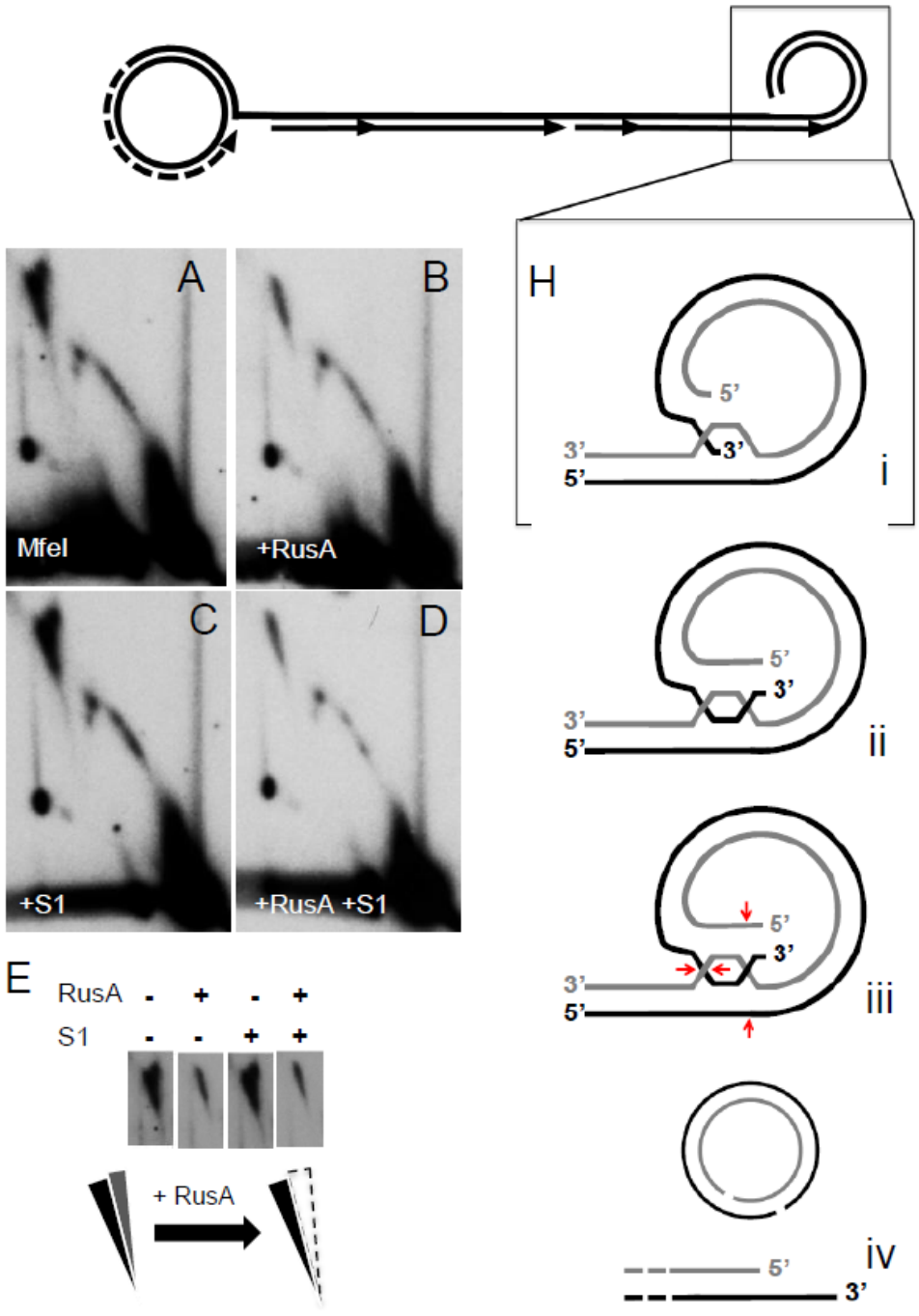
A
MfeI
B
+RusA
C
+S1
D
+RusA +S1
E
RusA - + - +
S1 - - + +
+ RusA
F
G
1.2
1
0.8
0.6
0.4
H
5'
3'
3'
5'
i
ii
iii
iv

Figure 4: Prominent Holliday junctions and hemicatenanes indicate a recombination hotspot at the major non-coding region. 2DNAGE patterns of MfeI-digested mtDNA treated with RusA resolvase, S1 nuclease, or both; panels are representative of three independent experiments. (A) Enhanced cruciform spike intensity is specific to RIs of the major-NCR fragment. (B) RusA treatment degrades the cruciform spike. (C) S1 nuclease treatment alone does not effect migration of replication intermediates. (D) RusA and S1 nuclease in combination further reduced the cruciform signal. (E) Two converging spikes of X-shaped intermediates are visible; the rightmost cruciform spike displays enhanced sensitivity to RusA while the leftmost spike remains. (F) Two Holliday junctions converge to form a hemicatenane. (G) ImageJ quantification of cruciform structure sensitivity to RusA, S1 nuclease, or both relative to the 1n monomer signal. Quantification represents the mean from three independent experiments +/- s.d. (H) Lariat mtDNAs have unreplicated ssDNA 3' ends which may 'snap back' and invade repeats in the major NCR (i); branch migration creates a double HJ (ii); HJ cleavage (iii) and resolution generate (iv) a gapped circle and new free 3' end.

Conclusion

Several conclusions may be drawn from these initial studies of nematode mitochondrial DNA (mtDNA) replication. First, the biochemical data indicate that the phylum Nematoda, as represented by *Caenorhabditis elegans*, employs an mtDNA synthesis mechanism heretofore unique among animals; rolling-circle replication. Second, the computational data on mtDNA substitutions suggests that this alternative DNA replication mechanism impacts the spectrum of mutations that accumulate in mtDNA in a heritable manner. The analyses of strand asymmetry presented herein further indicate that the mitochondrial genomes of other animal phyla, particularly the platyhelminths, contain patterns of substitution inconsistent with any known mtDNA replication model, including the rolling-circle model; this result is significant and warrants further study, as it indicates that additional mtDNA replication mechanisms may exist that have not yet been described. Third, these studies highlight a fundamental difference between the mitochondrial metabolism of the model nematode and mammals, bringing to light a critical weakness in the utility of *C. elegans* as a model for mtDNA replication defects. Further examination of assumptions, regarding the suitability of model organisms as stand-ins for the study of human disease, may prove fruitful.

Appendix. mTERF protein family members mTTF and mTerf5 have opposing roles in the coordination of mitochondrial DNA synthesis

Introduction

The mitochondrial transcription termination factor (mTERF) family comprises a set of mitochondrial DNA-binding proteins with diverse, documented roles in mitochondrial gene expression (Linder et. al. 2005; Roberti et. al. 2009). These proteins harbor multiple TERF motifs (I-IX), which form left-handed helical repeats that function as superhelical DNA-binding domains (Yakubovskaya et. al. 2010; Spahr et. al. 2010; Spahr et. al. 2012; Yakubovskaya et. al. 2012). mTERF family members have been implicated in the regulation of transcriptional initiation (Martin et. al.2005; Park et. al. 2007; Wenz et. al. 2009) as well as attenuation (Matrin et. al. 2005; Robeti et. al. 2003; Hyvarinen et. al. 2010), and have also been shown to participate in mitoribosome assembly and translation (Roberti et. al. 2006; Camara et. al. 2011; Wredenberg et. al. 2013). In the mouse, Mterf3 and Mterf4 are essential genes (Park et. al. 2007; Camara et. al. 2011), while Mterf2 is not (Wenz et. al. 2009). Four proteins of this family have been identified in *Drosophila*, of which the best characterized is mTTF (CG18124). mTTF binds two sequence elements in *Drosophila* mtDNA (Roberti et. al. 2003), each located at the junctions of convergently transcribed blocks of genes (see Fig. 1A). mTTF binding facilitates transcriptional termination bidirectionally *in vitro* and is required for transcriptional attenuation *in vivo* (Roberti et. al. 2006; Roberti et. al. 2005). The level and activity of mTTF therefore influences the steady-state levels of mitochondrial RNAs whose coding sequences lie between the mTTF binding sites and the putative promoters (Roberti et. al. 2006). Human MTERF1 also terminates transcription bidirectionally *in vitro* at its major

binding site downstream of the rRNA genes (Christiansen et. al. 1986; Kruse et. al. 1989; Shang et. al. 1994), but manipulation of its activity in cultured cells has been found to have only subtle effects on transcript levels (Hyvarinen et. al. 2010). Knockdown of the insect-specific paralog of mTTF, mTerf5, was found to have opposite effects on transcript levels to knockdown of mTTF, despite the fact that mTerf5 binds to the same sites in mtDNA in an mTTF-dependent manner (Bruni et. al. 2012).

Metazoan mitochondrial genomes are generally highly compact in their organization, with little or no non-coding sequences between genes. Typically, genes are encoded on both strands, although in vertebrates the majority are on one strand, whose transcription polarity is oppositely oriented to DNA replication. This type of organization means that encounters between the transcription and replication machineries, which compete for the same template, are inevitable. As in other genetic systems, these processes should be subject to regulation, in order to minimize and resolve potential conflicts, including both co-directional and anti-directional collisions between the two molecular machineries. Defects in collision regulation have been shown to cause abortive DNA synthesis, mutagenesis and genomic instability in a wide range of organisms (Poveda et. al. 2010; Mirkin et. al. 2007; Mirkin et. al. 2005; Srivatsan et. al. 2010; Knott et. al. 2009; Pomerantz et. al. 2010; Washburn et. al. 2011).

As DNA binding proteins with an established role in the regulation of mtDNA transcription, mTERF family members are strong candidates for mediating

conflicts between the replisome and transcription complexes. Moreover, MTERF proteins likely play a role in mtDNA metabolism, considering that alterations in the levels of MTERF1 or its homologs MTERF2 (MTERFD3) and MTERF3 (MTERFD1) modulate the levels of paused replication intermediates in cultured human cells (Hyvarinen et. al. 2007; Hyvarinen et. al. 2011). The sea urchin mTERF protein mtDBP has been demonstrated *in vitro* to have contrahelicase activity (Polosa et. al. 2005). This feature, commonly seen in replication termination proteins, is shared also by the mammalian nuclear rDNA transcription terminator TTF-1, which has been suggested to regulate entry of the replication machinery into an actively transcribed region (Putter et. al. 2002). The possible correspondence of the mTTF binding sites in *D. melanogaster* mtDNA with the regions of replication pausing identified in our earlier study (Joers and Jacobs 2013) suggests that mTERF family proteins have a role in the regulation of mtDNA synthesis. Moreover, the RITOLS mode (RNA incorporation throughout the lagging strand) of mtDNA replication in mammals has been demonstrated to require RNA transcripts for efficient synthesis of the lagging strand (Yang et. al. 2002; Yasukawa et. al. 2006). This type of productive interaction between mitochondrial transcription and replication suggests a potential regulatory role for mTERF family members in linking these two mechanisms together.

To test the possible involvement of mTTF and mTerf5 in mtDNA replication, we investigated their effects on mtDNA metabolism after manipulation of their expression by RNAi, both in cultured cells and *in vivo*. Here we show that both mTTF and mTerf5 are required for normal mtDNA topology and maintenance.

Lack of mTTF, mTerf5, or both factors is associated with developmental arrest at the L3 larval stage. Depletion of mTTF mRNA in S2 cells led to the accumulation of nicks, dsDNA breaks and recombination junctions. Analysis of mtDNA replication intermediates by two-dimensional neutral agarose gel electrophoresis (2DNAGE) revealed stalled replication forks over broad zones surrounding the TERF motif binding sites, and an accumulation of aberrant replication intermediates with extended segments of RNA/DNA hybrid, indicating a failure to complete lagging-strand DNA synthesis. Knockdown of mTTF's paralog, mTerf5, had an opposite effect on mtDNA replication intermediates, bringing about an increase in replication pause strength when compared to wild-type, a decrease in fragile replication intermediates containing single-stranded segments, and the disappearance of species even with the short segments of RNA/DNA hybrid that we were able to detect in wild-type cells.

Because of their opposing but essential roles in mtDNA expression and synthesis, we propose that the correct balance of these two mTERF family members facilitates the orderly and productive passage of oppositely moving replication and transcription complexes, preventing collisions that would otherwise result in abortive replication and loss of genome integrity.

Results

Replication pauses at TERF binding sites. The two major replication pause regions of *D. melanogaster* mtDNA were previously mapped approximately to 1/3 and 2/3 of genome length from the replication origin, which is located in the NCR (Joers and Jacobs 2013). More detailed analysis of these regions using 2DNAGE of ~5kb mtDNA fragments revealed that the location of these pause signals coincides with the positions of the two TERF motifs where mTTF binds the genome directly, at the cyt b/ND1 gene boundary (here designated bs1) and the ND3/ND5 gene boundary (bs2) (Fig. 1 and S1). The HindIII fragment beyond bs2, encompassing the remainder of the coding region, did not reveal any discrete pause signals. In accordance with the observed replication slow zone in this region (Joers and Jacobs 2013), an enhanced signal was evident at the start of the Y arc in this fragment. However, the nature of this slow zone was markedly different from pauses at mTTF binding sites, suggesting a separate mechanism causing the replication impediment. Treatment with the single strand-specific nuclease S1 had no effect on the migration of replication intermediates in any of the fragments tested, consistent with the previous inference that DNA replication in these regions is fully strand coupled (Joers and Jacobs 2013).

mTTF knockdown in S2 cells causes mtDNA depletion and altered topology. To investigate a possible role of mTTF in mtDNA maintenance we used RNA interference to knock down mTTF in S2 cells. Reduction of mTTF mRNA levels by ~70% (Fig. S2A) resulted in altered mitochondrial transcript levels

consistent with the previous report by Roberti et al. (2006; Fig S2B). Analysis of mtDNA copy number by qRT-PCR indicated that mtDNA levels fell to approximately 20% of control levels following 4-5 days of mTTF knockdown (Fig. 2A). This decrease was also confirmed by PicoGreen staining of mtDNA nucleoids (Fig. S2C) and Southern blotting of both digested and undigested mtDNA (Fig. 2B and S2D). The reduction in mtDNA amount by mTTF knockdown detected by Southern blots normalized for total nucleic acid concentration (Fig. 2B, S2D) was accompanied with progressive disappearance of the supercoiled mtDNA topoisomer form and increase in fragmentation relative to intact molecules (Fig. 2B). Analysis of linearized mtDNA by 2DNAGE revealed a relative increase both of recombination structures and broken RIs (Fig. 2C). Such broken intermediates are normally found only at a low-level in control cells after treatment with S1 nuclease (Joers and Jacobs 2013), which cuts the region that remains single-stranded in some replicating molecules, that extends from the replication origin across the rRNA gene locus. After mTTF knockdown, these broken intermediates were much more abundant, and further treatment with S1 nuclease had only a minor effect (compare NdeI digest in Fig. 2C with Fig. 6 in Joers and Jacobs 2013).

mTTF knockdown in developing flies causes mtDNA depletion, broken RIs and larval arrest. To investigate whether the effects of mTTF knockdown on mtDNA maintenance seen in S2 cells were consistent with knockdwon effects

in the whole organism, we expressed a (hairpin) dsRNA transgene targeted to mTTF, using a ubiquitous and constitutive *daughterless*-GAL4 (da-GAL4) driver. RNA interference *in vivo* produced ~90% knockdown of mTTF at the mRNA level and led to depletion of protein levels at larval stage (Fig. S2 A and B). mTTF knockdown larvae gained weight more slowly than wild-type larvae of the same genetic background (Fig. 3A). More than 90% of individuals failed to develop beyond the L3 larval stage although larval weight exceeded the critical range for developmental progression (Fig. 3A and S3C, De Moed et. al. 1999). None of the few aberrant pupae progressed to the late pupal stages. The persistent larval stage lasted approximately 30 days, during which time the larvae became progressively inactive until 100% mortality was reached.

Mitochondrial RNA levels were altered in a similar manner as in mTTF knockdown cells, e.g. COX2 mRNA was downregulated, whereas cytochrome b mRNA was elevated (Fig. S3D). Mitochondrial DNA copy number failed to increase as typical for wild-type development, remaining at 40% of the wild-type level 3 days after hatching (Fig. 3A). During the persistent larval stage, mtDNA copy number steadily declined to approximately 25% of the maximum level observed in wild type L3 larvae, by 25 days after hatching.

A similar accumulation of broken RIs was observed in mTTF knockdown larvae as in S2 cells (Fig. 3C). The control strain (w^{1118} x da-GAL4) displayed an identical pattern of RIs to that described previously for the Oregon-R wild-

type strain (Joers and Jacobs 2013) ruling out any confounding effect due to genetic background.

mTTF knockdown causes replication stalling in a broad zone, with failure to complete lagging-strand synthesis in RITOLS replication. The observed drop in mtDNA copy number and topological changes following mTTF knockdown prompted us to characterize mtDNA RIs in cells and in larvae knocked down for mTTF in more detail. In each of the two ClaI fragments that contained the mTTF binding sites, replication pauses were observed to fade out and spread over a wider region during 4 days of RNAi in S2 cells (Fig. 4A, red arrows). In accordance with data from linearized mt genomes (Fig. 2C) we observed a transient increase in cruciform DNA species, particularly a subclass of Holliday junction-like molecules (Fig. 4A, blue arrows). A similar decrease and spreading of the pauses was seen in mTTF knockdown larvae, including accumulation of recombination junctions (Fig. S4), although to a lesser extent than in mTTF knockdown cells.

Replication forks stalled by collisions have a tendency to regress along the template, thus forming a “chickenfoot” structure around a Holliday junction (see drawing in Fig. 4B). If such fork regression is relatively limited, then these RIs would still migrate close to a classical Y-arc but should become sensitive to nucleases targeting Holliday junctions. Consistent with this prediction, treatment with the bacterial cruciform-cutting enzyme RusA removed relatively more material from the region of the Y-arc in cells knocked down for mTTF compared to control cells (Fig. 4B, also Fig. S5 for comparison of original

data). Thus, mTTF knockdown resulted in an accumulation of stalled, regressed replication forks, which are considered as a signature of collisions. Note the decrease of the recombination structures migrating on the X-arc in response to RusA treatment, confirming its functionality in this experiment (blue arrows in Fig. 4B).

Since this data strongly suggests that mTTF binding acts as a barrier to fork progression, we reconsidered the issue of the role of RNA in mtDNA replication. Previous 2DNAGE analyses indicated that mtDNA RIs in *D. melanogaster* were fully double-stranded (Joers and Jacobs 2013), and that restriction sites across the genome were completely digestible (with the exception of the rRNA locus), which would not be the case in presence of RNA-DNA hybrids. This implies that in *Drosophila*, RITOLS RIs either contain only a few limited patches of RNA-DNA hybrid, and/or cannot be resolved by 2DNAGE. We investigated the issue further by treating mtDNA after restriction digestion with RNase H, which digests regions of RNA hybridized to DNA, thus modifying the migration pattern of RIs. This analysis revealed a prominent, novel arc, migrating just below the Y-arc (Fig. 5, red arrows), whose trajectory is consistent with the presence of one or more short segments of ssDNA arising from the removal of RNA from some replicating molecules. Such intermediates were missing in untreated samples (compare Fig. 5 Cla digests with corresponding digests in untreated samples in Fig. 4A). The prominence of this novel arc exceeded intesities of other intermediates, suggesting that additional material must have been released by the RNase H treatment which was previously unable to enter or be resolved in the gel

(compare intensities of RIs before and after RNase H treatment in samples hybridized on the same membrane in Fig. S6). The trajectory of the novel arc released by RNase H treatment differed markedly after knockdown of mTTF. The RITOLS arc migrating just below the standard Y-arc was changed to much shallower sub-Y arc, extending beyond the limit of the fragment analysed (Fig. 5, blue arrows). Its trajectory indicates much more extensive ssDNA regions (i.e. much longer segments of RNA/DNA hybrid prior to RNase H treatment, RITOLS-heavy in Fig. 5) than in the RIs that formed the sub-Y arc generated by RNase H treatment in untreated cells (RITOLS-light in Fig. 5).

mTerf5 is required for mtDNA maintenance. To test whether the mTTF partner protein mTerf5 could modulate the effects of mTTF on replication in addition to it's role in transcription, we investigated the effect of mTerf5 knockdown on mtDNA copy number in S2 cells (Fig. 6A). We observed a substantial depletion of mtDNA to a similar extent (~70%), and with similar kinetics, as mTTF knockdown. Simultaneous knockdown of both factors produced a more nuanced result, with an initial increase in mtDNA copy number, followed by a gradual decline to the same low mtDNA level as produced by knockdown of either factor alone, after 5 days of treatment. mTerf5 alone and simultaneous knockdown of mTerf5 and mTTF in the developing fly produced the same phenotype as mTTF knockdown, i.e. a persistent larval stage with failure of pupariation (Fig S2B).

Despite the fact that mTerf5 knockdown produced similar effects on mtDNA copy number and development as mTTF knockdown, 2DNAGE analysis of mtDNA from mTerf5 knockdown cells revealed different effects on the RI patterns. We observed enhanced replication pausing at mTTF binding site (compare the difference in intensity between pause site and Y-arc between Fig. 6C and corresponding digest in Fig 4A). The broken intermediates that accumulated in case of mTTF knockdown were absent (compare Fig. 6B with Fig. 2E), and treatment with S1 nuclease failed to release such intermediates in comparable levels to wt (compare Fig. 6B to Fig. 4 in Joers and Jacobs 2013). Treatment with RNase H had no discernible effect (compare Fig. 6C with the corresponding digests of Fig. 5). Loss of mTerf5 thus had opposite effects on RIs as mTTF knockdown, enhancing replication pausing at specific sites and shifting the balance of mtDNA replication to the fully dsDNA replication mode at the expense of RITOLS-type synthesis.

Discussion

Two members of the mTERF family, mTTF and mTerf5, were found to have specific and reciprocal effects on mtDNA metabolism, similar to their functional interaction in transcription. Contrary to a previous report (Roberti et. al. 2005), our data demonstrate that mTTF is required *in vivo* to maintain mtDNA levels. Prolonged application of either of two different dsRNAs targeted on mTTF, including one used by the previous investigators (Roberti et. al. 2005), resulted in mtDNA depletion in S2 cells, as well as in developing flies (Fig. S2E). Both dsRNAs produced similar results on mt transcription: the

loss of transcriptional termination activity resulting from mTTF knockdown upregulates transcripts downstream of each mTTF binding site, but downregulates upstream transcripts. The minor discrepancies between our findings on transcript levels and those reported elsewhere likely reflect the fact that changes in mtDNA levels compound those on RNA alone. The apparent difference in intensity of pausing after knockdown between larvae and cells (compare Fig.4A with Fig S3) is attributable to a difference in requirement of mitochondrial function; while S2 cells readily convert to a glycolytic metabolism, no such option is available for organism. Thus the observed conversion of pause site to stalling region is more pronounced in cultured cells.

The developmental arrest at larval L3 stage in response to loss of either mTTF, CG7175 or both factors appears to be a penetrant phenotype of mitochondrial dysfunction, as this phenotype is also shared by knockdown of many genes involved in mitochondrial function, including those encoding mitochondrial transcription factor 2 (mTFB2), mitochondrial single-strand binding protein (mtSSB) and CCDC56, a protein required for the assembly of cytochrome c oxidase (Maier et. al. 2001; Adan et. al. 2008; Peralta et. al. 2012; Rudolph et. al. 2007).

Although we previously found no evidence for RNA-containing RIs in *Drosophila*, the finer scale analysis conducted in the present study indicated the presence of short patches of RNA scattered around the mt genome, based on the prominent sub-Y arcs seen on 2DNAGE after treatment with

RNase H. Standard Y-arcs, which were already present before RNase H treatment, also remained after the treatment, albeit with decreased intensity. Thus, mtDNA RIs in *Drosophila* consist of two classes, as in vertebrates. One class is composed entirely of dsDNA. The second class contains short tracts of RNA/DNA hybrid, akin to the RITOLS intermediates seen in vertebrates (Yang et. Al. 2002; Yasukawa et. Al. 2006), except that the proportion of RNA/DNA hybrid in such molecules is much more limited in *D. melanogaster*, such that the arcs remaining after RNase H digestion migrate much closer to the standard Y-arc. Prior to RNase H digestion much of this RNA-containing material remains in complexes that cannot be resolved on standard 2DNAGE, even after restriction enzyme digestion. We suggest that the short segments of RNA that are hybridized to DNA are covalently joined to transcripts, forming complex networks unable to enter gels. RIs trapped in such tangles should be systematically released by RNase H treatment. Whether these two groups of RIs reflect two separate modes of replication or perhaps different stages of maturation remains to be determined.

mTTF and mTerf5 knockdown produce specific and reciprocal effects on mtDNA synthesis. Lack of mTTF causes random stalling, whilst causing a reduction in those molecules specifically paused at the binding site itself. This was accompanied by an increase in recombination junctions and in the accumulation of regressed forks across the regions surrounding bs1 and bs2. We observed a topological shift towards relaxed circular and linear nonreplicating molecules with concomitant accumulation of broken RIs, akin to those that can be created in material from unmanipulated cells by S1

nuclease treatment. This indicates that the ssDNA region in the rRNA locus is systematically broken, possibly being more pervasive or persistent than in control cells. In addition, a novel class of intermediates with extensive RNA:DNA hybrid segments accumulates. These replace the forms with only short RNA:DNA tracts, that are seen in control cells. The observed changes are consistent with a failure of site-specific pausing, presence of replication fork collisions and an inability to complete lagging-strand synthesis. Conversely, the absence of mTerf5 has exactly opposite effects; enhanced pausing at the mTTF binding sites, a decrease in the abundance of RIs broken at the rRNA locus and disappearance of the RNA-containing species.

The reciprocal nature of all these changes suggests that replication fork pausing and lagging-strand synthesis are related phenomena, with mTTF and mTerf5 implicated in their regulation. The bootlace mtDNA replication model (Yasukawa et. Al. 2006), recently inferred to operate in mammals (Reyes et. al. 2013) involves the incorporation of preformed, processed, lagging-strand RNA molecules, via hybridization with the displaced parental strand, as the replication fork progresses. The incorporated RNA is considered to be a possible source of primers for the synthesis of lagging-strand DNA. Our data suggest that replication of mtDNA in *Drosophila* follows a similar course, and that proteins belonging to the mTERF family are crucial factors in its execution and/or regulation, as illustrated in Fig. S7. The proposed model postulates that the balance of mTTF and mTerf5 nurses the productive interaction of the replication/transcription machineries moving in opposite directions, and that replication pausing is vital for ensuring the incorporation of RNA transcripts

into RIs at the replication fork (Fig. S7). Capture of a new bootlace, resulting from the arrival of a transcription complex that undergoes termination, is also proposed to be essential for the priming of lagging-strand DNA synthesis, not only at the immediate site of mTTF/mTerf5 binding, but also further downstream, as the replication fork progresses further. Consistent with this model, the intensity of the sub-Y arcs revealed by RNase H treatment in different regions of the mt genome correlates crudely with the level of transcription in the direction opposite to that of replication. Thus, the sub-Y is most prominent in the HindIII fragment mapping between mTTF bs2 and the NCR, near the start of the transcription unit that has the same polarity as the lagging strand (compare intensities of RNase H-sensitive intermediates between different digests in Fig. S6).

The prevention and/or regulation of collisions between the transcription and replication machineries appears to be indispensable for all genetic systems (Rudolph et. al. 2007), in particular to avoid knotting of the daughter molecules (Olavarrieta et. al. 2002), generation of recombinogenic ends (Takeuchi et. al. 2003) and other types of genomic instability (Mirkin et. al. 2007). Given their interconnected roles in mtDNA transcription and replication, these functions of mTTF and CG7175 are most likely related, providing a functional link between transcription and replication. The two proteins may also be considered as an example of an antagonistic pair of proteins that together control a specific process, a type of regulation widespread in biological systems, especially in development.

Materials and Methods

Flies, cell-lines and culture. Schneider's S2 cell-line (42) was cultured in Schneider's Medium (Sigma-Aldrich) at 25 ˚C. Cells were passaged every 3-4 d at a density of 0.5 x 10^6 cells/ml. Standard *Drosophila* strains, plus the mTTF RNAi line 101656 from the Vienna Drosophila RNAi Center (VDRC), were cultured as described previously (Fernandez et. al. 2009).

dsRNA constructs and transfection. Gene-specific dsRNAs were synthesized from templates created from S2 cell cDNA by a nested PCR strategy, which introduced the T7 promoter sequence on both sides of each final amplicon. S2 cells were transfected with 4 µg of each dsRNA added to 0.5 ml of culture medium, and grown for the times indicated in figures and legends. Where transfections were to be continued for > 3 d, cells were passaged every 3 d and fresh dsRNA was added. For visualization of nucleoids, dsRNA against Tfam was added for the final 2 d, as described in Supplementary Data, Nucleoids were detected by fluorescence microscopy, after staining with Quant-iT™ PicoGreen® (Invitrogen).

DNA and RNA extraction. Nucleic acids for mtDNA copy-number analysis, 2DNAGE and Q-RT-PCR were isolated from S2 cells, adult flies, larvae or purified mitochondria thereof using variants of standard methods. 2DNAGE used total nucleic acids isolated from sucrose density gradient-purified mitochondria (see Supplemental Information).

Q-PCR. Q-RT-PCR to measure RNA levels was performed essentially as described earlier (Roberti et. al. 2006), using cDNA prepared by random

priming or, where indicated, by gene- and strand-specific primers as detailed in Table S1. Assays always included three or more independent biological replicate samples, with normalization to the transcript of nuclear gene RpL32. Relative quantitation of mtDNA content was performed similarly, using total DNA as template, plus primers for mitochondrial 16S rDNA (Table S1), also with normalization to RpL32.

1D and 2D neutral agarose gel electrophoresis and Southern blot-hybridization. Standard one-dimensional electrophoresis used 0.6% agarose gels in TBE buffer. 2DNAGE and blot-hybridization were conducted essentially as described earlier, using slightly different conditions for resolving large and small DNA fragments (see Supplementary Data). Treatments of nucleic acid prior to 2DNAGE analysis with different enzymes are detailed in Supplemental Information. Radioactive probes for specific fragments of *Drosophila* mtDNA were generated by PCR, with [α-^{32}P]-dCTP (Perkin-Elmer, 3000 Ci/mmol) in the reaction mix (see Table S1 for primers).

References

Adan C, et al. (2008) Mitochondrial transcription factor B2 is essential for metabolic function in Drosophila melanogaster development. J Biol Chem 283(18):12333-12342.

Bruni F, et al. (2012) D-MTERF5 is a novel factor modulating transcription in Drosophila mitochondria. Mitochondrion 12(5):492-499.

Camara Y, et al. (2011) MTERF4 regulates translation by targeting the methyltransferase NSUN4 to the mammalian mitochondrial ribosome. Cell Metab 13(5):527-539.

Christianson TW & Clayton DA (1986) In vitro transcription of human mitochondrial DNA: accurate termination requires a region of DNA sequence that can function bidirectionally. Proc Natl Acad Sci U S A 83(17):6277-6281.

Fernandez-Ayala DJ, et al. (2009) Expression of the Ciona intestinalis alternative oxidase (AOX) in Drosophila complements defects in mitochondrial oxidative phosphorylation. Cell Metab 9(5):449-460.

He J, et al. (2007) The AAA+ protein ATAD3 has displacement loop binding properties and is involved in mitochondrial nucleoid organization. J Cell Biol 176(2):141-146.

Hyvarinen AK, et al. (2007) The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. Nucleic Acids Res 35(19):6458-6474.

Hyvarinen AK, Kumanto MK, Marjavaara SK, & Jacobs HT (2010) Effects on mitochondrial transcription of manipulating mTERF protein levels in cultured human HEK293 cells. BMC Mol Biol 11:72.

Hyvarinen AK, Pohjoismaki JL, Holt IJ, & Jacobs HT (2011) Overexpression of MTERFD1 or MTERFD3 impairs the completion of mitochondrial DNA replication. Mol Biol Rep 38(2):1321-1328.

Joers P & Jacobs HT (2013) Analysis of replication intermediates indicates that Drosophila melanogaster mitochondrial DNA replicates by a strand-coupled theta mechanism. PloS one 8(1):e53249.

Knott SR, Viggiani CJ, & Aparicio OM (2009) To promote and protect: coordinating DNA replication and transcription for genome stability. Epigenetics 4(6):362-365.

Kruse B, Narasimhan N, & Attardi G (1989) Termination of transcription in human mitochondria: identification and purification of a DNA binding protein factor that promotes termination. Cell 58(2):391-397.

Linder T, et al. (2005) A family of putative transcription termination factors shared amongst metazoans and plants. Curr Genet 48(4):265-269.

Maier D, et al. (2001) Mitochondrial single-stranded DNA-binding protein is required for mitochondrial DNA replication and development in Drosophila melanogaster. Mol Biol Cell 12(4):821-830.

Martin M, Cho J, Cesare AJ, Griffith JD, & Attardi G (2005) Termination factor-mediated DNA loop between termination and initiation sites drives mitochondrial rRNA synthesis. Cell 123(7):1227-1240.

Mirkin EV & Mirkin SM (2005) Mechanisms of transcription-replication collisions in bacteria. Mol Cell Biol 25(3):888-895.

Mirkin EV & Mirkin SM (2007) Replication fork stalling at natural impediments. Microbiol Mol Biol Rev 71(1):13-35.

Olavarrieta L, Hernandez P, Krimer DB, & Schvartzman JB (2002) DNA knotting caused by head-on collision of transcription and replication. J Mol Biol 322(1):1-6.

Park CB, et al. (2007) MTERF3 is a negative regulator of mammalian mtDNA transcription. Cell 130(2):273-285.

Peralta S, et al. (2012) Coiled coil domain-containing protein 56 (CCDC56) is a novel mitochondrial protein essential for cytochrome c oxidase function. J Biol Chem 287(29):24174-24185.

Polosa PL, Deceglie S, Roberti M, Gadaleta MN, & Cantatore P (2005) Contrahelicase activity of the mitochondrial transcription termination factor mtDBP. Nucleic Acids Res 33(12):3812-3820.

Pomerantz RT & O'Donnell M (2010) What happens when replication and transcription complexes collide? Cell Cycle 9(13):2537-2543.

Poveda AM, Le Clech M, & Pasero P (2010) Transcription and replication: breaking the rules of the road causes genomic instability. Transcription 1(2):99-102.

Putter V & Grummt F (2002) Transcription termination factor TTF-I exhibits contrahelicase activity during DNA replication. EMBO Rep 3(2):147-152.

Reyes A, Yang MY, Bowmaker M, & Holt IJ (2005) Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. J Biol Chem 280(5):3242-3250.

Roberti M, Bruni F, Polosa PL, Gadaleta MN, & Cantatore P (2006) The Drosophila termination factor DmTTF regulates in vivo mitochondrial transcription. Nucleic Acids Res 34(7):2109-2116.

Roberti M, et al. (2003) DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of Drosophila melanogaster mitochondrial DNA. Nucleic Acids Res 31(6):1597-1604.

Roberti M, et al. (2005) In vitro transcription termination activity of the Drosophila mitochondrial DNA-binding protein DmTTF. Biochem Biophys Res Commun 331(1):357-362.

Roberti M, et al. (2006) MTERF3, the most conserved member of the mTERF-family, is a modular factor involved in mitochondrial protein synthesis. Biochim Biophys Acta 1757(9-10):1199-1206.

Roberti M, et al. (2009) The MTERF family proteins: mitochondrial transcription regulators and beyond. Biochim Biophys Acta 1787(5):303-311.

Rudolph CJ, Dhillon P, Moore T, & Lloyd RG (2007) Avoiding and resolving conflicts between DNA replication and transcription. DNA Repair (Amst) 6(7):981-993.

Shang J & Clayton DA (1994) Human mitochondrial transcription termination exhibits RNA polymerase independence and biased bipolarity in vitro. J Biol Chem 269(46):29112-29120.

Spahr H, Habermann B, Gustafsson CM, Larsson NG, & Hallberg BM (2012) Structure of the human MTERF4-NSUN4 protein complex that regulates mitochondrial ribosome biogenesis. Proc Natl Acad Sci U S A 109(38):15253-15258.

Spahr H, Samuelsson T, Hallberg BM, & Gustafsson CM (2010) Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain. Biochem Biophys Res Commun 397(3):386-390.

Srivatsan A, Tehranchi A, MacAlpine DM, & Wang JD (2010) Co-orientation of replication and transcription preserves genome integrity. PLoS Genet 6(1):e1000810.

Takeuchi Y, Horiuchi T, & Kobayashi T (2003) Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev 17(12):1497-1506.

Washburn RS & Gottesman ME (2011) Transcription termination maintains chromosome integrity. Proc Natl Acad Sci U S A 108(2):792-797.

Wenz T, Luca C, Torraco A, & Moraes CT (2009) mTERF2 regulates oxidative phosphorylation by modulating mtDNA transcription. Cell Metab 9(6):499-511.

Wredenberg A, et al. (2013) MTERF3 Regulates Mitochondrial Ribosome Biogenesis in Invertebrates and Mammals. PLoS Genet 9(1):e1003178.

Yakubovskaya E, et al. (2012) Structure of the essential MTERF4:NSUN4 protein complex reveals how an MTERF protein collaborates to facilitate rRNA modification. Structure 20(11):1940-1947.

Yakubovskaya E, Mejia E, Byrnes J, Hambardjieva E, & Garcia-Diaz M (2010) Helix unwinding and base flipping enable human MTERF1 to terminate mitochondrial transcription. Cell 141(6):982-993.

Yang MY, et al. (2002) Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. Cell 111(4):495-505.

Yasukawa T, et al. (2006) Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. Embo J 25(22):5358-5371.

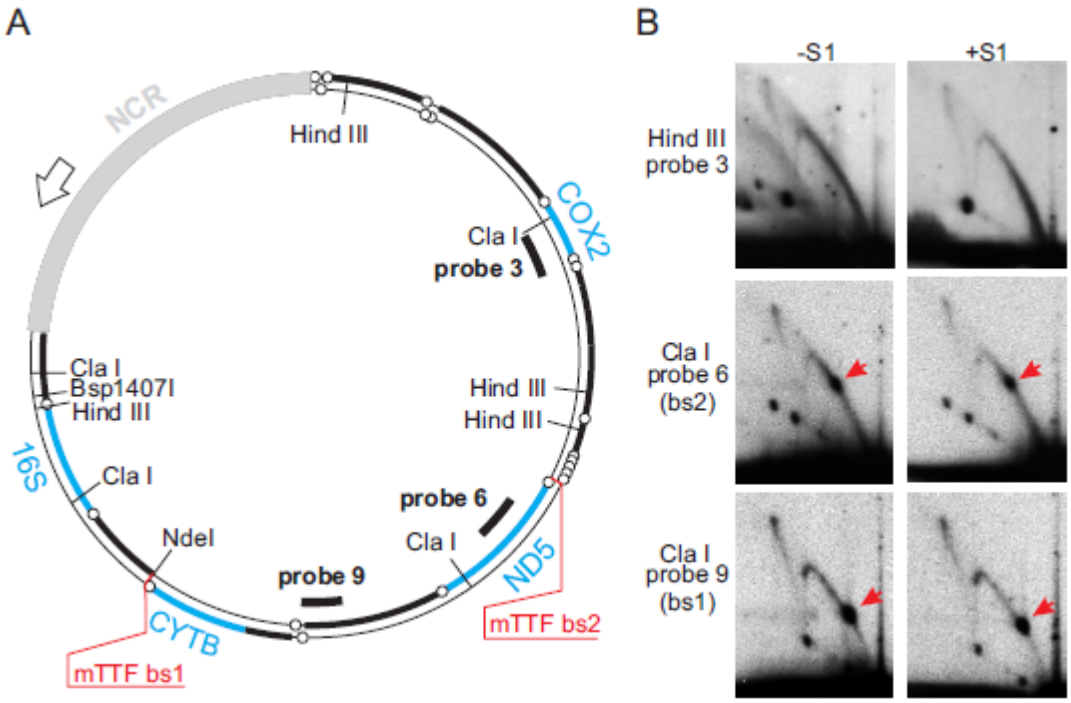

Figure 1: Replication pauses at mTTF binding sites.

A. Schematic drawing of *D. melanogaster* mtDNA with positions of probes, mTTF binding sites, gene clusters (in bold), tRNA genes (open circles) and restriction endonuclease sites for Hind III, Cla I, Nde I and Bsp 1407I. Positions of genes for which expression was analyzed are shown in blue. Open arrow points to the direction of replication from putative origin. B. 2DNAGE of ClaI- or HindIII-digested mtDNA. Red arrow points to the major pause site.

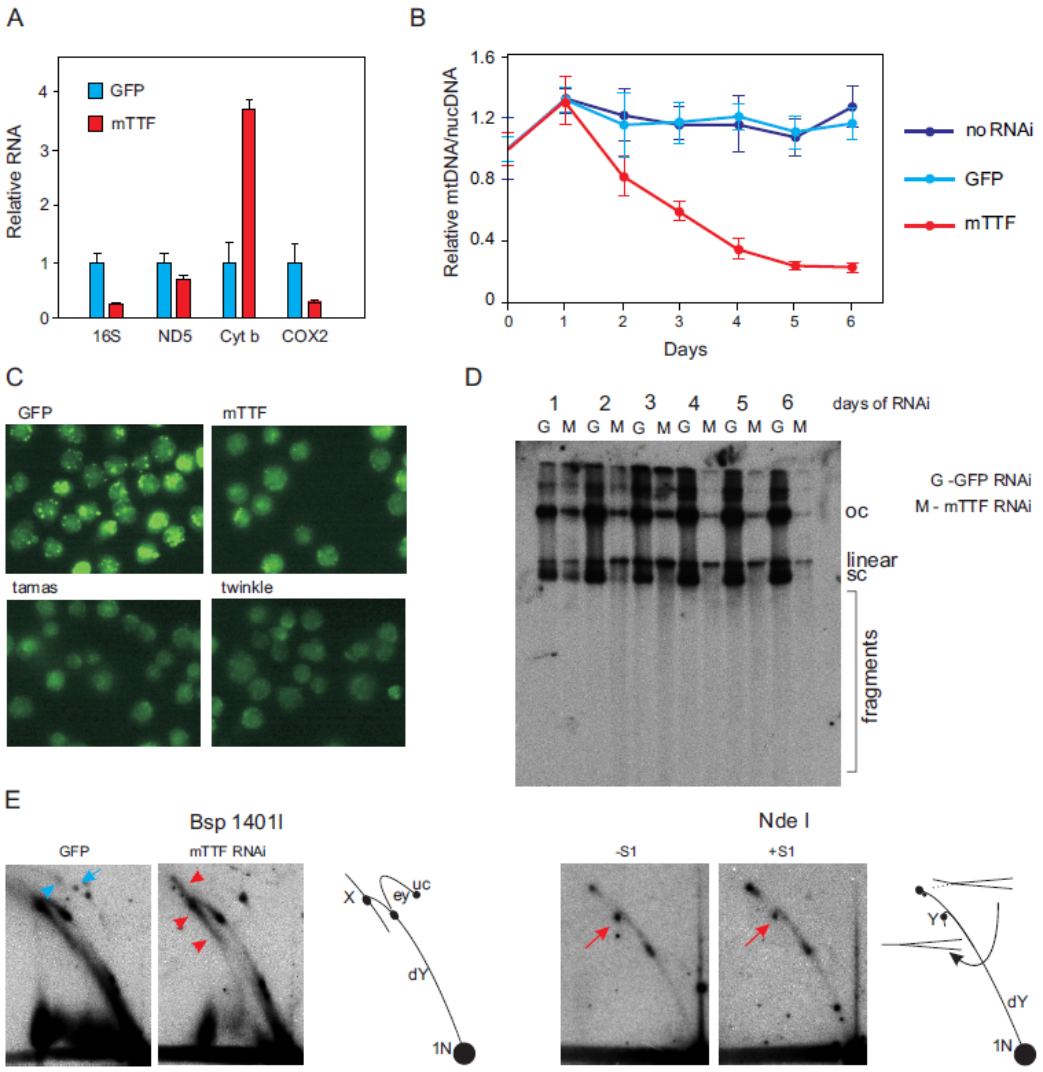

Figure 2: Effects of mTTF knockdown in S2 cells.

A. Mitochondrial transcript levels after 3 days of dsRNA treatment, either against mTTF or GFP, measured from three biological replicates. Error bars represent standard deviation. B. mtDNA copy number changes measured by qPCR, following treatment with dsRNA against GFP, mTTF and mock-transfected cells from three biological replicates. Error bars represent standard deviation. C. Images of PicoGreen-stained cell culture after 3 days of dsRNA against GFP, mTTF, and gened with established role in mtDNA synthesis: tamas and CG5924 (*D. melanogaster* homologue of mitochondrial helicase Twinkle). D. Agarose gel of uncut mtDNA, normalized for total nucleic acid concentration, showing three major forms (oc

– open circles, linear – genome-length linears, sc – supercoiled circles). E. 2DNAGE of mtDNA cut by restriction enzymes with single digestion sites in the genome. Bsp 1407I: note the disappearance of the characteristic, partially single-stranded eyebrow arc (blue arrows) and the increase in recombination intermediates or X-forms (red arrows) 3 days after start of mTTF dsRNA treatment. Nde I: note the accumulation of replication intermediates broken at the origin/rRNA locus. For detailed explanation of RI structures and conversions see Fig. S2G and (32).

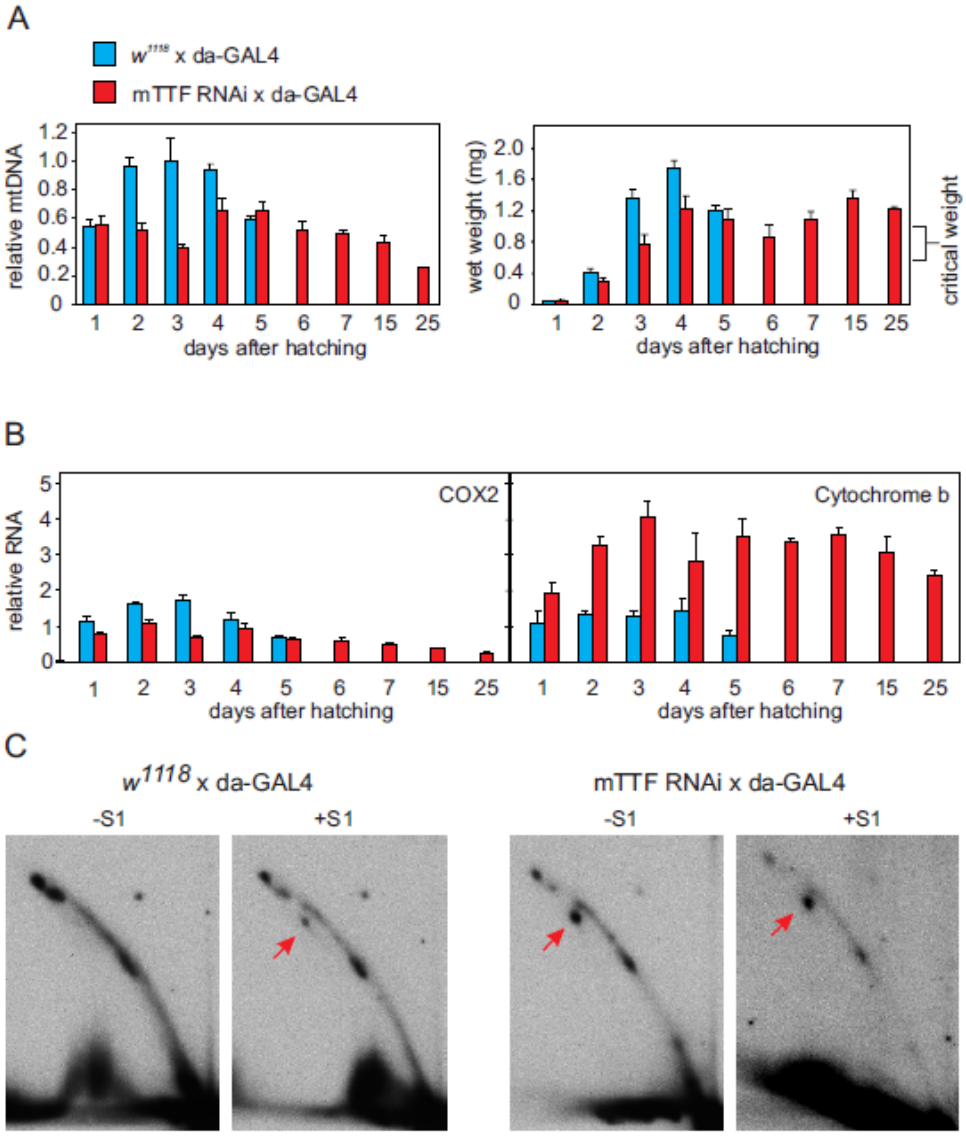

Figure 3: Effects of mTTF knockdown in flies.

A. Changes in mtDNA copy number, measured by qPCR, and in wet weight of larvae from mTTF RNAi (red bars) and control strain (blue bars) from three biological replicates, error bars represent standard deviation. B. Changes in expression of COX2 and cyt b genes in larvae from mTTF RNAi (red bars) and control strain (blue bars), measured from three independent replicates. Error bars represent standard deviation C. 2DNAGE of Nde I-digested mtDNA with or without S1 treatment. Note the appearance of broken replication intermediates in mTTF RNAi background (red arrows) which are only visible in material from control larvae after S1 treatment (see also Figure 2E).

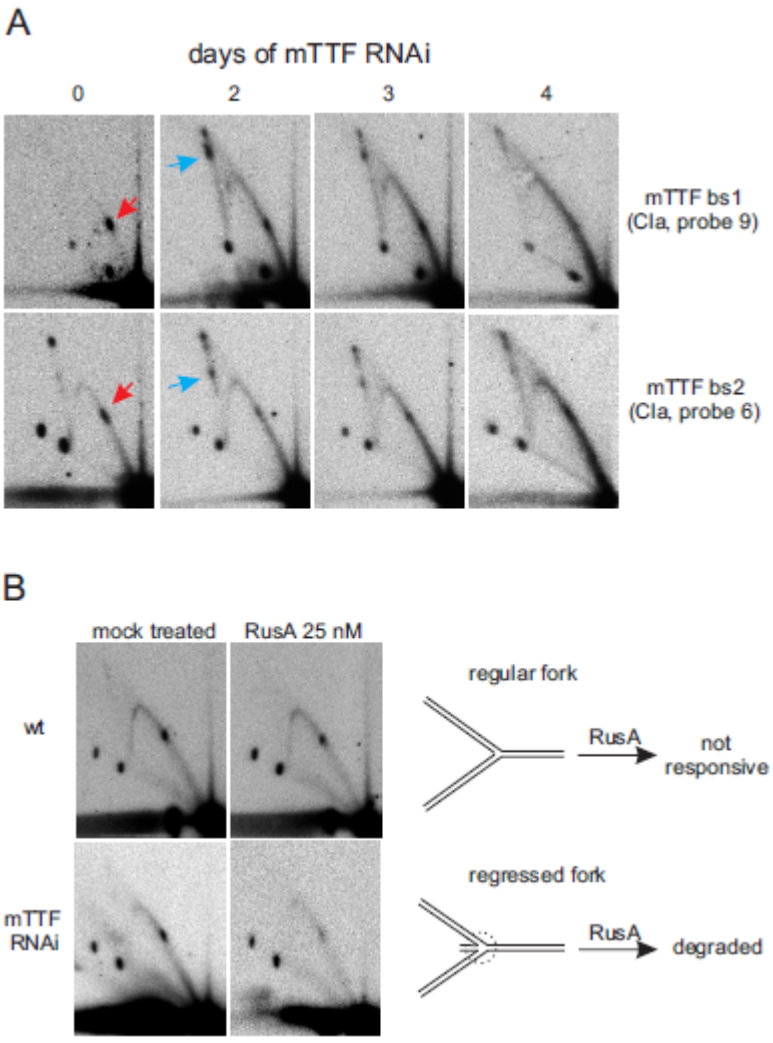

Figure 4: Aberrant replication fork stalling resulting from mTTF RNAi.

A. 2DNAGE of ClaI-digested mtDNA, time-course with mTTF dsRNA treatment, showing effects on replication intermediates in fragments containing the mTTF binding sites: substitution of defined pauses with wider regions of stalling across both mTTF binding sites. Red arrows point to pause site and blue arrows to increase in recombination intermediates. B. 2DNAGE analysis of replication intermediates of fragment containing mTTF bs2 before and after RusA-treatment. Note the increased removal of signal from the Y-arc after 3 days of mTTF dsRNA treatment and the RusA-mediated disappearance of X-forms in both cases (blue arrows). Drawing describes how regressed replication forks become sensitive to cruciform-cutting enzymes.

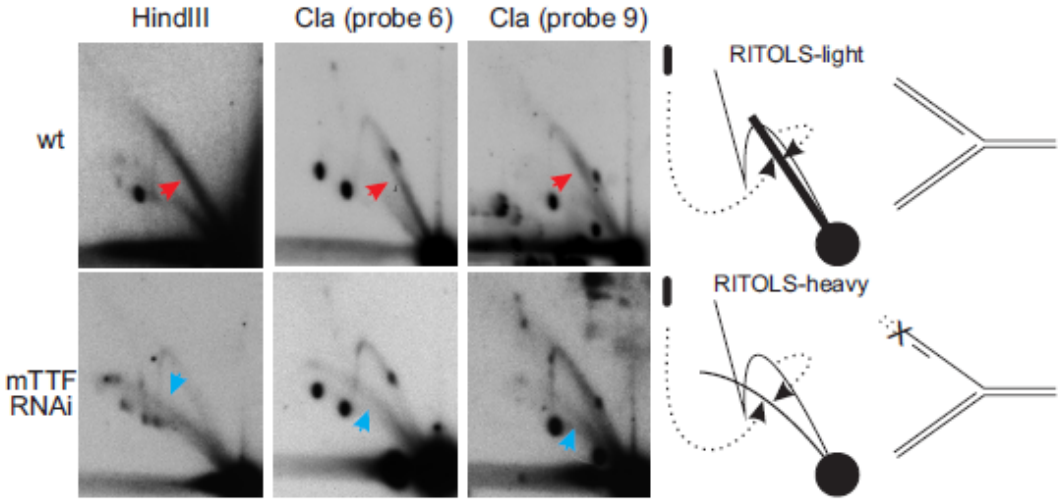

Figure 5: The effect of mTTF RNAi on RNA-dependent replication.

2DNAGE analysis of RNase H-treated replication intermediates from untreated and mTTF dsRNA-treated cell lines (after 3 days of treatment). Note the appearance of structures with limited segments of RNA/DNA hybrid in control cells and more extensive RNA/DNA hybrid segments in mTTF dsRNA-treated cells. Drawing shows the structure and provenance of aforementioned replication intermediates.

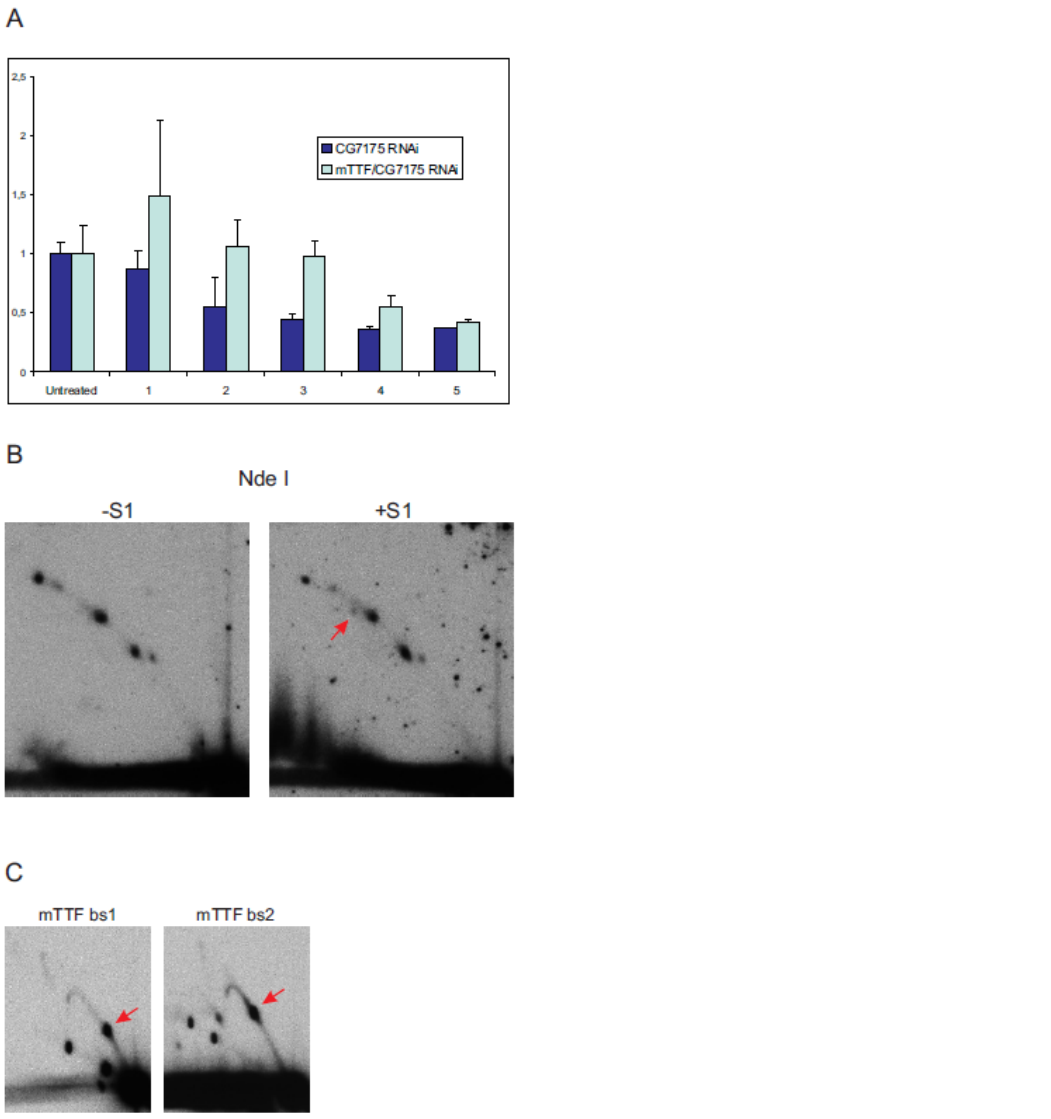

Figure 6: Effects of mTerf knockdown in S2 cells.

A. mtDNA copy number measured from three biological replicates by qPCR following treatment with dsRNAs against mTerf5, or both mTTF and mTerf5. Error bars represent standard deviation. B. NdeI-digested mtDNA after 3 days of mTerf dsRNA. C. ClaI-digested mtDNA +/- RNaseH, after 3 days mTerf dsRNA.

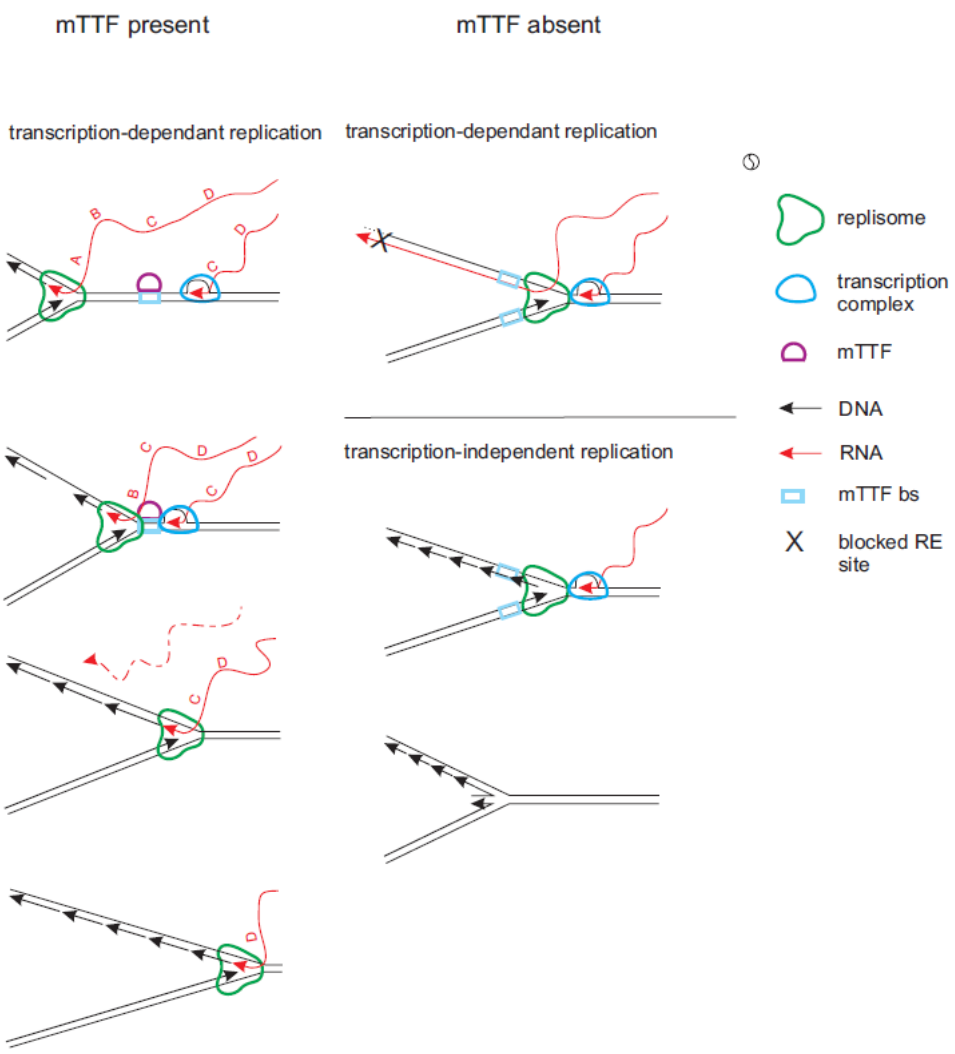

Figure 7: A tentative model for mTTF function.

A. Normal replication pauses at mTTF binding sites, where orderly passage of transcription and replication is mediated. mTTF binding sites may additionally be the places where an RNA bootlace is supplied, providing for lagging-strand priming. Letters mark corresponding positions on RNA and DNA strands. B. In case of mTTF RNAi, both proteins are absent from mtDNA and uncontrolled collisions take place outside of the mTTF binding site, leading to regressed forks and failure of normal lagging strand synthesis in RITOLS replication. C. Lack of mTerf5 leaves mTTF alone to excert stronger-than-normal pausing activity leading to downregulation of RITOLS replication.

Notes on Supplementary Information and Materials

All Supplementary Information referenced herein is available online through ProQuest Dissertation Services, www.proquest.com.