The Complete Mitochondrial Genome Sequence of the Spider *Habronattus oregonensis* Reveals Rearranged and Extremely Truncated tRNAs

Susan E. Masta[1] *and Jeffrey L. Boore*

Department of Energy Joint Genome Institute, Department of Evolutionary Genomics, Walnut Creek, California

We sequenced the entire mitochondrial genome of the jumping spider *Habronattus oregonensis* of the arachnid order Araneae (Arthropoda: Chelicerata). A number of unusual features distinguish this genome from other chelicerate and arthropod mitochondrial genomes. Most of the transfer RNA (tRNA) gene sequences are greatly reduced in size and cannot be folded into typical cloverleaf-shaped secondary structures. At least nine of the tRNA sequences lack the potential to form TΨC arm stem pairings and instead are inferred to have TV-replacement loops. Furthermore, sequences that could encode the 3′ aminoacyl acceptor stems in at least 10 tRNAs appear to be lacking, because fully paired acceptor stems are not possible and because the downstream sequences instead encode adjacent genes. Hence, these appear to be among the smallest known tRNA genes. We postulate that an RNA editing mechanism must exist to restore the 3′ aminoacyl acceptor stems to allow the tRNAs to function. At least seven tRNAs are rearranged with respect to the chelicerate *Limulus polyphemus*, although the arrangement of the protein-coding genes is identical. Most mitochondrial protein-coding genes of *H. oregonensis* have ATN as initiation codons, as commonly found in arthropod mtDNAs, but cytochrome oxidase subunits 2 and 3 genes apparently use TTG as an initiation codon. Finally, many of the gene sequences overlap one another and are truncated. This 14,381-bp genome, the first mitochondrial genome of a spider yet sequenced, is one of the smallest arthropod mitochondrial genomes known. We suggest that posttranscriptional RNA editing can likely maintain function of the tRNAs, while permitting the accumulation of mutations that would otherwise be deleterious. Such mechanisms may have allowed for the minimization of the spider mitochondrial genome.

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

Introduction

There are roughly 38,000 described species of spiders (Araneae), a group that diversified over 350 MYA. Despite this, no mitochondrial genomes and only a few mitochondrial genes have previously been sequenced from spiders. Therefore, little is known about the patterns and processes of mitochondrial genome evolution within this major group of arachnids.

Among the 11 extant orders of arachnids, Araneae and Acari (ticks and mites) are by far the most species rich. Arachnida and Merostomata (horseshoe crabs) are the two extant subclasses that make up the Chelicerata, and Chelicerata, Hexapoda, Crustacea, and Myriapoda together constitute the phylum Arthropoda. The origin of Chelicerata predates the divergence of Crustacea and Hexapoda, and divergence among chelicerate groups was underway in the Cambrian, with all extant orders well established by the Devonian (Weygoldt 1998). Yet, of this ancient and diverse group, the only chelicerate mitochondrial genome sequences published to date are from a horseshoe crab (Lavrov, Boore, and Brown 2000) and from ticks and mites within the arachnid order Acari (Black and Roehrdanz 1998; Navajas et al. 2002).

A typical metazoan mitochondrial genome is about 16 kb in size and contains 37 genes (22 transfer RNAs, 2 ribosomal RNAs, and 13 protein-coding genes), along with some noncoding sequences necessary for initiating and regulating transcription and replication (Clayton 1982; Wolstenholme 1992; Boore 1999). The arrangement of genes within the mitochondrial genome is well conserved within some major metazoan lineages, and therefore the evolutionarily infrequent rearrangements that have occurred hold potential for resolving systematic relationships among metazoan taxa (Boore and Brown 1998). Mitochondrial gene arrangement has proved to be a useful phylogenetic character for delimiting major lineages of invertebrates, including arthropods (Smith et al. 1993; Boore et al. 1995; Boore, Lavrov, and Brown 1998; Dowton 1999; Morrison et al. 2002). Because little is known about chelicerate mitochondrial gene arrangements, it is not clear whether this type of data could help to resolve the contentious and poorly resolved relationships among these taxa (for a review of proposed relationships, see Wheeler and Hayashi [1998]). However, given the ancient divergences among arachnid orders, we might expect to find gene arrangements that differ from horseshoe crabs, the sister group to arachnids. From just the few acarid mitochondrial genomes that have been examined, it has been shown that gene order can be used to delimit a major division (i.e., prostriate versus metastriate) within ticks (Black and Roehrdanz 1998; Roehrdanz, Degrugillier, and Black 2002).

In addition to gene arrangement, another feature of mitochondrial genomes known to vary among some metazoan groups is the secondary structure of their transfer RNA molecules. Transfer RNA has a highly conserved cloverleaf-shaped secondary structure, present in virtually all eukaryotes and prokaryotes (Dirheimer et al. 1995). Exceptions to this cloverleaf structure have been found in mitochondrial tRNAs from metazoans, with one to a few of the 22 tRNAs in a genome sometimes lacking one of the arms of the cloverleaf. Metazoan mitochondrial $tRNA^{Ser(AGN)}$ genes lack the potential to form stem base pairs in the dihydrouridine arm (DHU, or D, arm) of the cloverleaf, which are instead replaced by D loops (Wolstenholme 1992). Loss of the TΨC arm (T arm) has been described for a few of the mitochondrial tRNAs from

[1] Present address: Department of Biology, Portland State University, Portland, Oregon.

E-mail: smasta@pdx.edu.

Key words: mitochondrial genome, tRNA editing, aminoacyl acceptor stem, gene rearrangement, genome size, *Habronattus oregonensis*.

Mol. Biol. Evol. 21(5):893–902. 2004
DOI:10.1093/molbev/msh096
Advance Access publication March 10, 2004

 © Society for Molecular Biology and Evolution 2004; all rights reserved.

diverse metazoans, including gastropods (Terrett, Miles, and Thomas 1996), lizards (Macey et al. 1997), and brachiopods (Helfenbein, Brown, and Boore 2001). However, the most unusual situation occurs in a group of nematodes, where it has been shown that all mitochondrial tRNAs, except those coding for serine, lack T arms and variable loops. These are instead replaced by T arm–variable loops (TV-replacement loops), whereas the serine tRNAs have D-replacement loops (Wolstenholme et al. 1987; Okimoto and Wolstenholme 1990).

Because the cloverleaf secondary structure of tRNA is so widespread among all life forms, it can be argued that loss of the T arm is likely deleterious and that such tRNAs are not optimal. In general, mitochondrial tRNAs appear to be accumulating deleterious mutations, based on comparisons of bond strengths and substitution rates in mitochondrial versus nuclear tRNAs (Lynch 1996, 1997). Further evidence for the accumulation of deleterious mutations in mitochondrial tRNAs is provided by the fact that they are smaller and display almost no invariant nucleotide positions relative to their nuclear counterparts (Lynch 1996). Loss of a T arm further reduces both the size and the number of conserved nucleotide positions present in a given tRNA. This lack of sequence conservation also suggests that mitochondrial tRNAs may be more likely to display unpaired nucleotides in their stem regions, compared with their nuclear counterparts.

However, one possible solution that may overcome the effects of the accumulation of deleterious mutations in the DNA is posttranscriptional modification of the RNA transcripts. In this scenario, the bases would be modified so that the RNA contains paired bases in its stem regions. Such RNA editing has been found to change the nucleotide base pairs in the aminoacyl acceptor stems of metazoan mitochondrial tRNAs (e.g., Lonergan and Gray [1993], Börner et al. [1997], and Lavrov, Brown, and Boore [2000]).

A previous population genetic study revealed that the mitochondrial genes for $tRNA^{Val}$ and $tRNA^{Leu(CUN)}$ are truncated in jumping spiders in the genus *Habronattus* (Masta 2000). These tRNA sequences lack the potential to form the typical cloverleaf-shaped tRNA, and instead have TV-replacement loops. There is also intraspecific variation in the regions of the sequences that typically would encode the aminoacyl acceptor stems, suggesting that RNA editing occurs to correct nucleotide mispairings in these regions. Additionally, the gene for $tRNA^{Leu(CUN)}$ has a different location in the mtDNA molecule than that found in horseshoe crabs and prostriate ticks.

In the present study, we seek to determine whether and to what extent the other 20 mitochondrial tRNAs may be truncated, may possess mispaired acceptor stem sequences, and may be rearranged with respect to other chelicerates, with the long-term goal of understanding patterns of molecular evolution in arachnids.

Materials and Methods

Mitochondrial DNA was extracted from one individual of the jumping spider *Habronattus oregonensis* from the Santa Rita Mountains of Arizona, as previously described (Masta 2000). The mitochondrial genome was amplified in two large overlapping fragments by long PCR, using the primers HbCB-J (5′-TGGATTTA-GAGTTTTTTTTATTTT-3′) with HbC2-3300-J (5′-TAG TAT TTA TTG CTT TTC CTT CTC-3′) and Hb12S-N (5′-CAAAATAAGGCAAGTCGTAACA-3′) with C1-J-2309. Primers HbCB-J, HbC2-3300-J, and Hb12S-N were specifically designed to amplify *H. oregonensis* from this location, based on previously reported sequences (Masta 2000) (GenBank accession numbers AF239951 and AF241478), and primer C1-J-2309 was designed specifically for spiders by Hedin (1997*a*). Long-PCR amplification was performed using the Stratagene Herculase™ DNA polymerase kit. Reaction conditions were similar to the manufacturers' suggestions, with final concentrations of 0.16 mM of each dNTP, 0.4 mM of each primer, 1X Herculase™ polymerase buffer, 1 μl of mitochondrial DNA (concentration not determined), and 2.5 Units of Herculase™ polymerase per 100 μl reaction. Reactions were cycled at 92°C for 30 s, 50°C for 25 s, and 68°C for 12 min, for 37 cycles, followed by a final extension at 72°C for 15 min. The PCR products were electrophoresed in 0.8% agarose gels to estimate their size and concentration, then purified by precipitating with 0.3 M sodium acetate and 100% ethanol, followed by resuspension in water.

These PCR products were then processed for DNA sequencing by the production facility of the Department of Energy Joint Genome Institute. First, DNA was mechanically sheared randomly into fragments of about 1.5 to 2 kb. After enzymatic end repair and gel purification, these fragments were ligated into pUC18 and transformed into *E. coli* to create plasmid libraries, all using standard techniques (Sambrook and Russell 2001). Automated colony pickers were used to select colonies into 384-well plates. After overnight incubation, a small aliquot was processed robotically through amplification of plasmids, sequencing reactions, reaction clean-up, and electrophoretic separation on ABI 3730XL or Megabace 4000 automated DNA sequencer machines to produce a sequencing read from each end of each plasmid.

Sequences were fed automatically from sequencing machines into a UNIX-based folder system, where they were processed using Phred, trimmed for quality, and assembled using Phrap. Quality scores were assigned automatically, and the electropherograms and assembly were viewed and verified for accuracy using Sequencher (GeneCodes).

Typically, about 30X coverage of a mitochondrial genome is achieved in this way, meaning that any given region has been sequenced an average of 30 times, a redundancy that provides a deep and accurate assembly. The only nucleotides included in the assembly were those computer-scored at Phred Q20 or higher, a very stringent quality standard.

A region from each of the ends of the PCR product generated by the primer pair Hb12S-N with C1-J-2309 was not included in the assembly of sequences, presumably because of some bias in the shearing and/or enzymatic repair and cloning. These regions were subsequently sequenced by designing specific internal

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

primers for this long-PCR product. Chromatograms from sequences resulting from primer walking (and, therefore, not having the extensive coverage of the rest of the genome) were checked for accurate base calls by eye before aligning with the remainder of the sequences. Figure 1 is a map of the mitochondrial genome, indicating the locations of the primers used to amplify the genome for this study.

A consensus sequence was made from all the contigs in Sequencher and imported into MacVector, which was used as an aid in annotating the sequences. Protein-coding genes were identified by similarity of inferred amino acid sequences to those of other arthropod mtDNAs (*Drosophila yakuba*, *Anopheles gambiae*, *Artemia franciscana*, *Daphnia pulex*, *Lithobius forficatus*, *Narceus annularis*, *Limulus polyphemus*, *Ixodes hexagonus*, and *Rhipicephalus sanguineus*; GenBank numbers X03240, NC002084, NC001620, NC000844, AF309492, NC003343, AF216203, AF081828, and AF081829, respectively). Translated amino acid sequences were aligned manually in MacClade 4 (Maddison and Maddison 2000). The large ribosomal subunit RNA gene (*16S* or *rrnL*) sequence in *H. oregonensis* had previously been determined (Masta 2000) and annotated. The boundaries of the small ribosomal subunit RNA (*12S* or *rrnS*) were determined by comparing alignments and secondary structure with other known arthropod *rrnS* sequences.

The tRNA genes were the most difficult genes to annotate, and a series of steps was required to determine their locations and approximate gene boundaries. After all the protein-coding and ribosomal gene sequences had been determined, the remaining regions were searched for tRNAs with the use of the program tRNAscan-SE 1.21 (Lowe and Eddy 1997). Because no tRNAs were found using this program's default parameters for mitochondrial DNA, the search was modified to find secondary structures that had very low Cove scores (> 0.1) and to find tRNAs that lacked the T arm, as in nematode mtDNA. Few potential tRNAs were identified even with these low-stringency searches. Therefore, most tRNAs were determined by aligning all chelicerate tRNA sequences in GenBank for each tRNA and looking for anticodon arm motifs that were conserved among Chelicerata. We then searched for conserved motifs (or sometimes only the 3-bp anticodon sequence) in *H. oregonensis* using MacVector, searching both strands in both directions, for all regions not previously assigned as protein-coding or ribosomal-encoding DNA. The sequence around each potential anticodon region that shared sequence similarity was then examined to determine if the potential existed for it to form stem-and-loop regions typical of tRNAs. Searches for potential tRNAs included sequence from the ends of genes that had already been assigned, to allow for potential overlap of gene-coding regions.

Comparisons among the protein-coding genes of *Ixodes hexagonus* (GenBank accession number NC002010), *Rhipicephalus sanguineus* (GenBank accession number NC002074), *Limulus polyphemus* (GenBank accession number NC003057) and *H. oregonensis* were made for each protein-coding gene by aligning the inferred amino acids. We inferred initiation codons to be those codons that most closely corresponded with the inferred initiation site of other mtDNA genes. The termini of these genes were inferred to be the first in-frame stop codon (when present and not overlapping the downstream gene), or an abbreviated stop codon corresponding well to the termini of other chelicerate genes when necessary to avoid overlap with the downstream gene. The nucleotide content of the coding regions of each strand of DNA was inferred with the aid of MacVector.

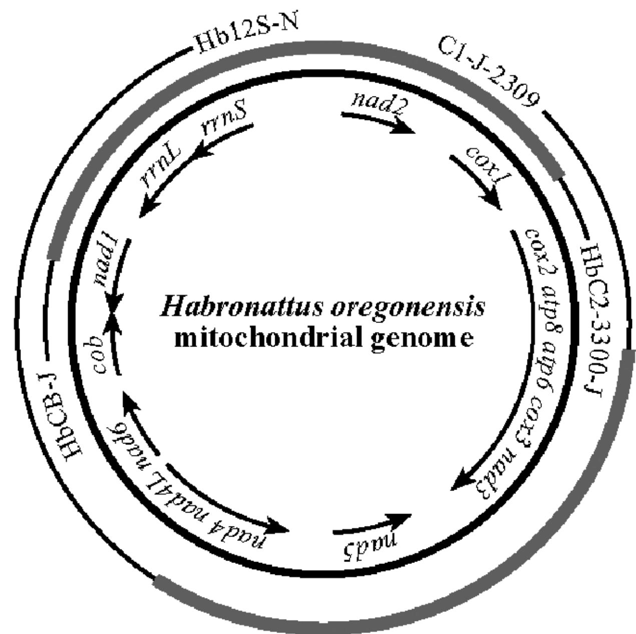


FIG. 1.—Mitochondrial gene map of *Habronattus oregonensis* with the location and names of primers used for amplification. The approximate location of the ribosomal and protein-coding genes is shown, with arrows indicating direction of transcription. tRNA genes are omitted. The thick semicircular lines shaded gray indicate regions that had extensive sequence coverage, whereas the thinner lines are regions that were obtained by primer walking on long-PCR products, and demarcate the limits of long-PCR amplification products. The areas of very high depth include the majority of the tRNA genes.

Results

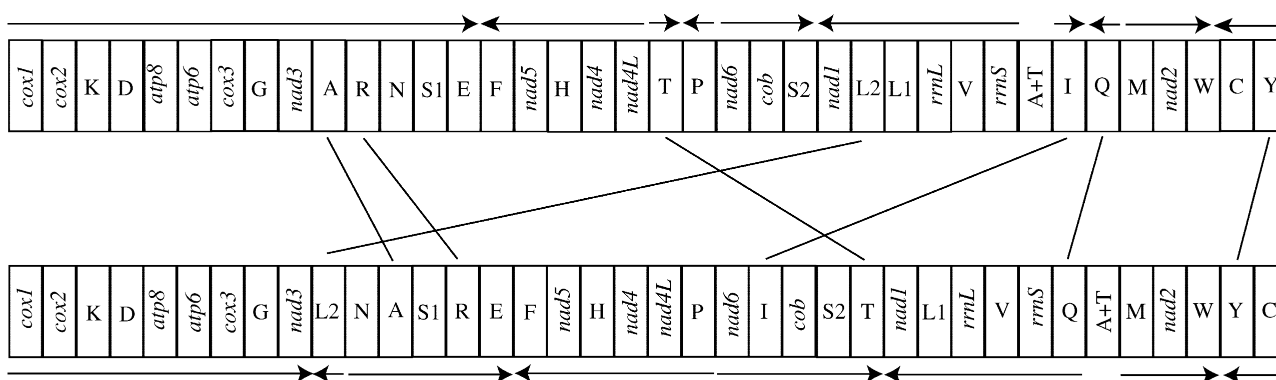
The mitochondrial genome of *H. oregonensis* (GenBank accession # A4571145) encodes the 37 genes typically found in metazoan mitochondria. The gene order of the protein-coding and ribosomal RNA genes is identical to that of *L. polyphemus*, but seven tRNA genes have a different arrangement (fig. 2). Most of the genes are shorter relative to other chelicerates, as detailed below, such that the total size of the genome is 14,381 base pairs.

Transfer RNAs

It was not possible to infer the exact gene boundaries for all the tRNA genes, because many lack potential for forming complete aminoacyl stems and, in some cases, even lack the adjacent T arm and have sequences that are not well conserved with respect to other chelicerate tRNA gene sequences. Instead, these tRNA genes appear to be

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

Limulus polyphemus



Habronattus oregonensis

FIG. 2.—Mitochondrial gene arrangement of *Limulus polyphemus* and *H. oregonensis*. These circular genomes have been graphically linearized at the arbitrarily chosen *cox1* to more clearly show gene rearrangements. Protein and rRNA genes are abbreviated as in the text; tRNA genes are abbreviated using the one-letter amino acid code. The two leucine and two serine tRNA genes are further differentiated by numeral, where L1 = *trnL*(*uag*), L2 = *trnL*(*uaa*), S1 = *trnS*(*ucu*), and S2 = *trnS*(*uga*). Gene sizes are not drawn to scale.

missing the sequence to encode parts of typical tRNAs (see fig. 3). Because we do not know the exact extent of the tRNA genes at their 3′ end, we have annotated these to end at the boundary of the adjacent downstream gene. Most of the inferred tRNA gene sequences are considerably shorter with respect to those of other chelicerates. In particular, the majority of the inferred tRNA secondary structures appear to be truncated at their 3′ end, such that they possess no potential to form a normal aminoacyl acceptor stem (fig. 3). In at least nine tRNAs, the downstream gene on the same strand begins at a position that would otherwise be expected to be part of the tRNA gene. Even allowing for the possibility of gene overlap, those nucleotides that would be shared between the two genes could not form all the required bonds with the adjacent nucleotides in the 5′ portion of the acceptor stem. Poorly or unpaired acceptor stems are also found for tRNA genes with extensive overlap with genes encoded on the opposite DNA strand. An extreme example of this is *trnL2*, in which only one possible nucleotide base pairing can occur among the seven nucleotides of the 5′ portion of the acceptor stem and the 3′ nucleotides. Although the exact gene boundaries cannot be ascertained with DNA sequence data alone, it appears that none of the 22 tRNA sequences have the potential to form a fully paired, seven-member aminoacyl acceptor stem. It seems unlikely that these unpaired nucleotides are cases of gene overlap with the downstream gene on the same strand; rather, it is more parsimonious to conclude that most of the tRNAs do not encode all or part of the 3′ portion of the aminoacyl acceptor stem. Perhaps this structure is formed posttranscriptionally, as it is in centipede mt tRNAs (see below). Some of these tRNAs are even further truncated; only 13 tRNAs can be inferred to possess a base-paired T arm. At least nine tRNAs contain a TV-replacement loop, similar to those first found in nematodes (Wolstenholme et al. 1987). Even among some of the tRNAs that can potentially fold into a T arm, the inferred stems and loops are small (*trnL2* and *trnM*) or poorly paired (*trnY*, *trnH*, and *trnV*), casting doubt on their identity as T arms.

The amount of truncation or gene overlap found in each tRNA gene varies. The most extreme example is *trnC*, whose sequence overlaps the adjacent gene on the opposite strand (*cox1*) for 7 nt at the 5′ end and appears to overlap a gene on the same strand (*trnY*) for 15 nt on the 3′ end. This amount of sequence overlap of genes on the same strand has not been found in mtDNA genes to date. Furthermore, if the *trnC* gene does overlap *trnY* by 15 nt, it forms a very poorly paired acceptor stem, with only 2 of 7 nt capable of forming pair bonds. Instead, it seems more parsimonious to conclude that the 3′ end of *trnC* is not entirely encoded by the *trnY* sequence. Only 36 nt of the *trnC* gene encode solely *trnC*. However, the nucleotides present in the D and anticodon arm are highly conserved among *trnC* genes present in other chelicerates. Additionally, the D and anticodon stems are of the typical length found in tRNAs (3 to 4 bp and 5 bp, respectively [Dirheimer et al. 1995]), the length found in all 22 tRNAs of *H. oregonensis*. Therefore, because of both the sequence conservation and the secondary structure conservation in the D and anticodon arms, the *trnC* gene appears not to be a pseudogene. Five other tRNA genes (*trnK*, *trnD*, *trnR*, *trnP*, and *trnL1*) lack the potential to both form a T arm and to have their TV-replacement loops immediately followed by downstream sequences encoding adjacent genes on the same strand. The sequences encoding solely these tRNAs range in size from 45 (*trnR*) to 50 (*trnD*) nt. Only *trnQ* has what can be considered a normal full-length (66 bp) tRNA gene sequence, in that it does not overlap that of an adjacent gene and can be folded into a cloverleaf secondary structure.

Several candidate regions were identified for the tRNA$^{Ser(UCN)}$ gene (*trnS2*). However, only one has a perfectly paired anticodon stem (see fig. 3), and this sequence also shares the greatest similarity with the *trnS2* genes of other chelicerates. This *trnS2* gene is downstream

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

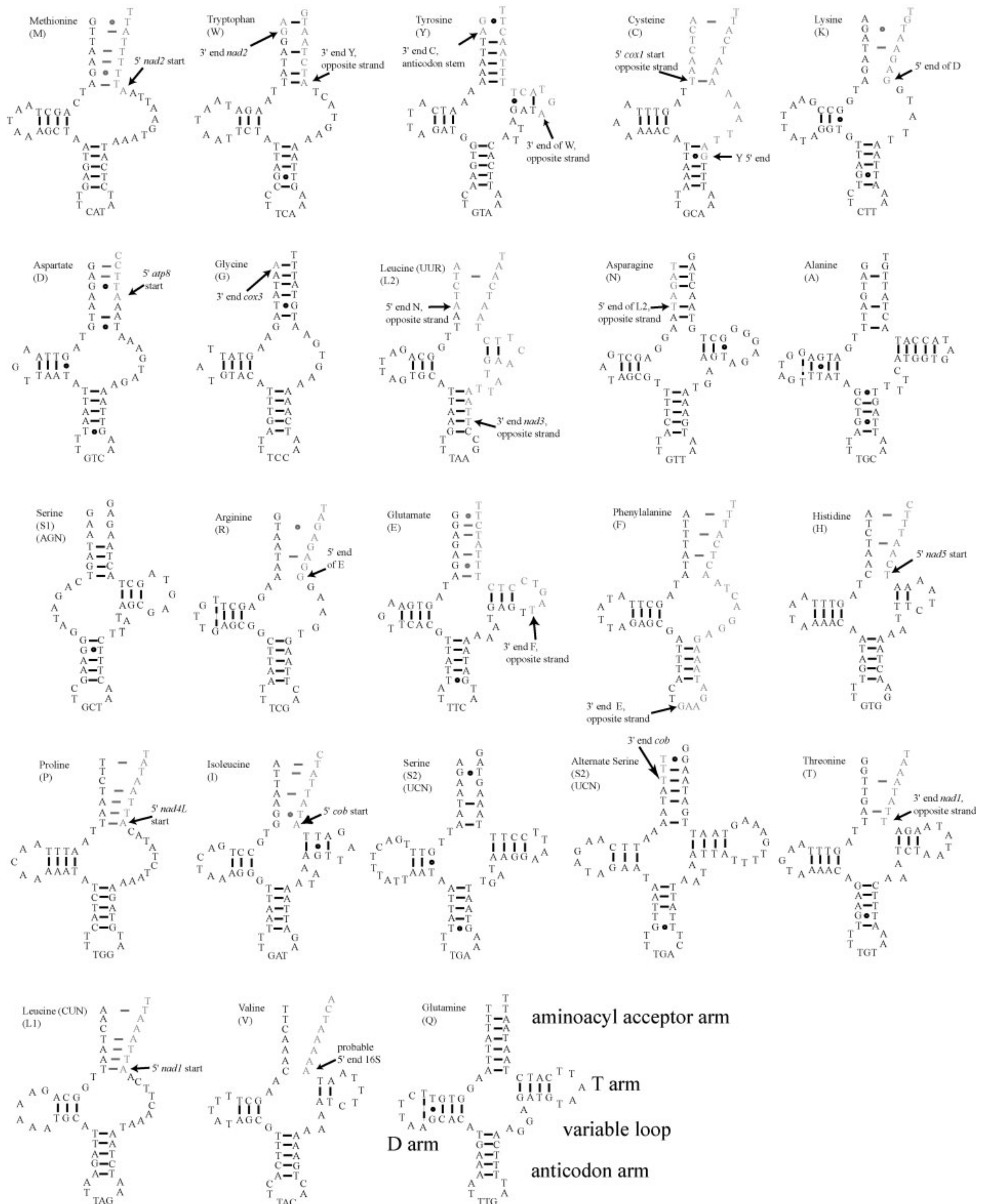


FIG. 3.—Inferred secondary structures of the 22 mitochondrial tRNAs from *H. oregonensis*, shown in the order they occur in the genome, beginning with *trnM*. The lack of sequence conservation and the probable RNA editing at the 3′ end makes it impossible to ascertain with certainty the location of the ends of the tRNA genes. In these structures, we have sought to minimize overlap with adjacent genes on the same strand, while inferring a T arm whenever possible. However, alternative structures may be possible for some of the tRNAs depicted. The location of adjacent gene boundaries are indicated with arrows, and gray font indicates that the sequence is part of an adjacent gene. Nucleotides in the 3′ acceptor stem region are angled away from the 5′ stem when we infer that they are not actually part of the tRNA gene. Inferred Watson-Crick nucleotide bonds are illustrated by lines, and GT bonds are illustrated by dots. Dashed lines indicate that a nucleotide bond could potentially form, but that such a bond is not usually found in those regions in typical tRNAs. Gray lines and dots indicate potential bonds, if these sequences actually encode 3′ acceptor stems.

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

Table 1
Properties of *Habronattus oregonensis* mtDNA Protein-Coding Genes and Comparisons with Other Chelicerates

	Number of Amino Acids					
Gene	*H. oregonensis*	*I. hexagonus*	*R. sanguineus*	*L. polyphemus*	Inferred Initiation Codon	Inferred Termination Codon
atp6	220	220	221	224	ATT	TAA
atp8	51	51	52	51	ATT	TAG
cob	374	366	358	377	ATA	T
cox1	513	512	512	511	ATT	TAA
cox2	222	225	225	228	TTG	TAG
cox3	261	261	257	261	TTG	TAA
nad1	306	313	313	310	ATT	TAA
nad2	319	318	313	338	ATT	TAG
nad3	113	111	114	114	ATA	TAA
nad4	429	436	433	445	ATA	T
nad4L	89	91	91	99	ATT	T
nad5	547	554	552	571	ATC	T
nad6	143	141	149	153	ATG	TAA

of *cob*, as is the case for horseshoe crab mtDNA, but just upstream from this region is an alternative potential *trnS2* sequence (fig. 3), although it shares less sequence similarity with other chelicerates and also has a mispairing in the anticodon stem. The potential for a 3′ acceptor stem exists for this sequence, although it too has a nucleotide mispairing with the adjacent 5′ stem.

Ribosomal RNA Genes

The entire sequence and the structure of the 3′ half of the large ribosomal subunit RNA has previously been reported (Masta 2000). The small subunit RNA (*rrnS*) has a sequence length of about 693 bp, similar to the size in the ticks *R. hexagonus* and *I. sanguineus* (Black and Roehrdanz 1998) but almost 100 bp shorter than that of *L. polyphemus* (799 bp [Lavrov, Boore, and Brown 2000]) and the insects *Drosophila yakuba* (789 bp [Clary and Wolstenholme 1984]) and *Anopheles gambiae* (800 bp [Beard, Hamm, and Collins 1993]).

Protein-Coding Genes

Most of the protein-coding gene sequences encode fewer amino acids than those of *L. polyphemus* and slightly fewer than those of ticks (table 1). In general, genes that are reduced in size in ticks, relative to *L. polyphemus*, are further reduced in *H. oregonensis*. The only potential exception is *cob*, which appears to be longer than that found in either ticks or the horseshoe crab.

The inferred start codons for 11 of the 13 protein-coding genes are ATN, which is typical for metazoan mitochondria (Wolstenholme 1992). Less typical is the inferred use of TTG as an initiation codon, which we find in the *cox2* and *cox3* genes of *H. oregonensis*.

The termination codons TAG or TAA appear at the end of nine of the protein-coding genes. These termination codons are not present in *nad4, nad4L, nad5*, and *cob*. We inferred these genes to be terminated by truncated stop codons (T), which are presumably polyadenylated after transcription to form complete TAA stop codons (Ojala, Montoya, and Attardi 1981). If, alternatively, these genes were to extend to the first in-frame stop codon, it would require overlap of the downstream genes by 26, 117, 20, and 24 bp for *nad4, nad4L, nad5*, and *cob*, respectively.

Noncoding Regions

There are nine regions of the mtDNA that do not appear to encode genes, and most of these consist of 2 to 4 nt between adjacent genes. The largest unknown region (716 bp) occurs in a similar region to where the A+T-rich region (control region) occurs in Limulus (see fig. 2). This region has an A+T composition of 77.1% and is, therefore, not strikingly A+T rich, compared with the 74% A+T composition of the remainder of the genome.

There are few or no noncoding nucleotides between adjacent protein-coding genes, with at most 4 bp between such genes. The largest unassigned region is 65 bp, between the *trnN* and *trnA* genes. This region contains a sequence that can be folded into a secondary structure similar to that predicted for $\mathrm{tRNA}^{\mathrm{Ser(UCN)}}$, although except for the 3-nt anticodon, the sequence is very divergent from other chelicerate *trnS2* genes, as mentioned above.

Strand Bias

The strand encoding the majority of proteins we will refer to as the α-strand, following the convention used for the mitochondrial genome sequence of the chelicerate *L. polyphemus* (Lavrov, Boore, and Brown 2000). This strand includes nine protein-coding genes and 13 tRNAs. The other strand, referred to here as the β-strand, includes four protein-coding genes (*nad1, nad4L, nad4,* and *nad5*), both ribosomal RNA genes, and nine tRNAs (figs. 2 and 4). The genes encoded on the α-strand have an overall nucleotide frequency of $T = 0.507$, $C = 0.105$, $A = 0.335$, and $G = 0.211$. The regions encoding the genes on the β-strand have an overall nucleotide frequency of $T = 0.375$, $C = 0.152$, $A = 0.386$, and $G = 0.087$. Thus, the genes on the α-strand are more rich in G and T. Perner and Kocher (1995) developed a measure of nucleotide composition skewness that we used to determine the skew on the α-strand, where AT skew = [A-T]/[A+T] and GC skew = [G-C]/[G+C]. We find an asymmetrical nucleotide bias for the two strands, with the α-strand AT skew = −0.112

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

and GC skew = 0.301. This strand asymmetry is consistent with a mitochondrial replication–induced mutation bias (Francino and Ochman 1997).

The tRNAs encoded by both the α-strands and β-strands share similar secondary structures and degrees of overlap with adjacent genes (see figs. 2 and 3). Thus, it does not appear that the strand on which a tRNA is encoded influences whether it maintains a T arm and 3′ acceptor stem sequence.

Discussion

The mitochondrial genome of *H. oregonensis* is the smallest reported from any chelicerate. In general, the mitochondrial genomes of chelicerates are smaller than those of other arthropods. This compact size results from most of their genes being shorter than their counterparts in other arthropods. *Habronattus oregonensis* has achieved an even smaller size through further reduction in the size of its genes, in particular its ribosomal and transfer RNA genes. The tRNA genes are drastically reduced at their 3′ ends so that many no longer encode a T arm, and, likely, most do not even encode a 3′ acceptor stem. Therefore, these tRNA genes are even smaller than nematode mt tRNA genes, including *trnS1*, which has been described as the minimal tRNA (Okimoto and Wolstenholme 1990; Ohtsuki, Kawai, and Watanabe 2002). While the possibility may exist that these genes instead overlap with adjacent genes, there are a number of reasons this seems unlikely. First, this would require more extensive overlap than has been previously found in any mitochondrial genes. Second, because the mtDNA is transcribed as a single polycistronic RNA (in those few animals in which it has been studied), it is not obvious how the overlapping genes would be resolved to generate two complete RNAs. Third, the sequences do not yield the predicted stable base pairs with the 5′ acceptor stem, and some exhibit no or very little potential to form any base pairs in this region (*trnV, trnL2, trnC, trnR, trnH,* and *trnT*). Without the possibility of forming stable base pairs and without sequence conservation with tRNA genes from other organisms, these regions cannot be annotated as acceptor stem regions. However, the 3′ acceptor stem is not necessary if these sequences are posttranscriptionally modified by the RNA-editing mechanism, such as the one described below. Last, if we postulate that the genes are not truncated, but instead overlapping, we would have to infer both that an alternative splicing mechanism exists to allow the two transcripts to exist and that these transcripts also are modified by RNA editing to replace the unpaired nucleotides in the acceptor stem. It is more parsimonious to simply infer that the 3′ acceptor stem sequence is missing and they are probably posttranscriptionally edited.

While we cannot entirely rule out the possibility that the mitochondrial tRNAs could be nonfunctional and that the mitochondrion instead utilizes tRNAs imported from the cytoplasm, this does not seem a plausible explanation for these aberrant tRNA structures. The stem regions both in the anticodon arms and in the D arms exhibit typical Watson-Crick pairing of nucleotides, and the anticodon sequence itself is conserved, so it does not appear that these have become pseudogenes. Additionally, depending on the identity of the tRNA, there are high degrees of sequence conservation with other chelicerate tRNAs for the D arms and anticodon arms (yet not for any other region). Pseudogene tRNAs would be expected to mutate at the same rate along the length of the gene, so that all regions should have lost their stem pairings to the same degree that the T and acceptor stems have.

RNA Editing

Our finding that sequence is lacking for both the T arm and most likely for the 3′ acceptor stems in many of this genome's tRNAs is unprecedented among metazoan mitochondrial tRNAs. Nematode mitochondrial tRNAs are characterized by TV-replacement loops, but they possess well-paired acceptor stem sequences (Wolstenholme et al. 1987). Gastropods contain a few mitochondrial tRNAs that lack the T arm and also have mispaired acceptor stems (Yamazaki et al. 1997; Kurabayashi and Ueshima 2000) but not nearly to the extent found in *H. oregonensis*. Because it was previously demonstrated that tRNA editing occurs in the acceptor stems of a gastropod (Yokobori and Pääbo 1995), the authors suggested this type of RNA editing may be widespread among gastropods (Yamazaki et al. 1997).

The centipede *Lithobius forficatus* has been found to possess mitochondrial tRNAs with acceptor stem sequences that are not capable of forming normal bonds and with varying amounts of overlap with downstream genes (Lavrov, Boore, and Brown 2000) in a pattern that is similar to that of *H. oregonensis* mt tRNAs. These authors showed that a novel type of posttranscriptional RNA editing occurs whereby the 5′ end of the aminoacyl acceptor stem appears to serve as a template for building the 3′ end of the acceptor stem. We postulate that a similar, but more extensive, type of editing occurs in *Habronattus*, because without such a mechanism, its tRNAs would presumably be nonfunctional. If so, the same or a similar type of RNA editing must have arisen independently in the lineage leading to these spiders, as the other chelicerates whose mitochondria have been sequenced have paired nucleotides in their acceptor stem sequences.

If RNA editing does indeed occur through the use of the 5′ portion of the acceptor stem as a template, we would predict there may be little sequence conservation of the 5′ stem, because any sequence would be sufficient to form a template. This is because these 5′ stem nucleotides would be released from the constraint of needing compensatory mutations to maintain pairing of the arm. This seems to be the case, as the only regions of high sequence similarity among chelicerates are the D arms and anticodon arms. Likewise, lack of selection pressure on the acceptor stem sequences may mean that a mutation in that region behaves essentially as a neutral mutation. If this is so, we may expect to find standing genetic variation in the acceptor stem sequences within and between populations of species that have the capacity to form stable tRNA acceptor stems via RNA editing with the 5′ acceptor stem serving as a template. This is consistent with the observation of intraspecific sequence variation in the 5′

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

and 3′ regions of the *trnL1* and *trnV* genes in two species of *Habronattus* (Masta 2000). These lines of evidence provide strong support for the role of RNA editing in *Habronattus* tRNAs. Positive confirmation of its occurrence will require analysis of the processed tRNA transcripts.

Phylogenetic Utility of Information from Mitochondrial Genomes

Relationships among chelicerates are contentious and poorly resolved. However, there is agreement among both morphological (Shultz 1990) and molecular (Wheeler and Hayashi 1998) data that Araneae is one of the more recently derived orders within Arachnida. Therefore, we infer that the tRNA gene rearrangements we observe in *H. oregonensis* occurred after the divergence of Acari, because both prostriate ticks and *Limulus polyphemus* share an identical gene arrangement (Black and Roehrdanz 1998; Roehrdanz, Degrugillier, and Black 2002). Although it is not possible to determine from this gene order alone exactly what the pattern of gene rearrangement has been over the course of evolutionary time, the sampling of additional arachnid taxa may make this possible. Currently, no information is available on the mitochondrial gene arrangements of other spiders or orders of arachnids believed to be more closely related to Araneae than is Acari. Therefore, we do not know which, if any, of the tRNA gene rearrangements we find in *H. oregonensis* may be shared, derived characters, and hence phylogenetically informative.

We can also infer that the extreme tRNA truncation and likely RNA editing found in spiders is evolutionarily derived, as these unusual features have not been described in Acari. The evolutionary loss of an arm in a tRNA gene is probably irreversible, and there are no known cases in which a normal cloverleaf-shaped tRNA has evolved secondarily from a tRNA that has lost an arm. Because losses of arms or helices of RNA are such rare evolutionary events, it is possible that they can be used as reliable phylogenetic characters for delimiting lineages. One such study has already shown that tRNA secondary structure delimits a clade of lizards (Macey, Schulte, and Larson 2000). Therefore, it is likely that tRNA secondary structure will also be able to provide characters for reconstructing relationships among spiders or other highly derived arachnids. However, all phylogenetic inferences based on tRNA arrangement or inferred secondary structure will have to await further sequencing of other arachnid mitochondrial genomes.

Mitochondrial Genome Size Evolution

The reduction in size of tRNA genes, ribosomal genes, and at least half of the protein-coding genes in the mitochondrial genome of *H. oregonensis* suggests there has been an overall trend toward minimization of mitochondrial genome size in spiders. *Habronattus oregonensis* possesses the smallest known tRNA genes, such that these may represent the minimal size necessary for proper tRNA functioning.

Lynch (1996) has shown that deleterious mutations are accumulating at a faster rate in metazoan mitochondrial tRNAs than in their nuclear counterparts. He suggested that this mutation accumulation may ultimately lead to a loss in overall fitness in major lineages of organisms (Lynch 1996, 1997). The extreme truncation of mitochondrial tRNA genes found in *H. oregonensis* could, by this reasoning, suggest that spider mitochondria have very low fitness. Could this mean, then, that the old and species-rich order of arachnids may suffer an irreversible “mutational meltdown” (Gabriel, Lynch, and Bürger 1993)?

Although mutations that appear to be deleterious have accumulated in the tRNAs of *Habronattus*, it is possible that posttranscriptional RNA editing may serve as a mechanism whereby such substitutions or losses of nucleotides could accrue, while still allowing the genome to maintain normally functioning tRNAs. If the ability to correct acceptor stem nucleotide mispairing arose via RNA editing, the selection pressure to maintain normal base pairings, or indeed to maintain any 3′ acceptor stem sequence, would be lost. If the occurrence of an RNA editing mechanism is coupled with a selective replicative advantage for small genome size, then we would predict that mitochondrial genomes that had lost 3′ acceptor stem regions would have an advantage. Therefore, the evolution of a type of RNA editing that allows loss of these regions could serve both to counteract deleterious mutations that disrupt nucleotide stem pairings and to lead to reduction in gene size. As with all evolutionary novelties, it is simply a matter of chance whether they arise within a given lineage, and it has been shown that different types of RNA-editing mechanisms have arisen multiple times in different lineages over the course of evolution (Covello and Gray 1993; Gray 2001). However, once this type of RNA editing arises, we would expect over evolutionary time that all tRNAs within a genome would eventually lose their 3′ acceptor stem sequences. The time scale over which this would occur would depend on the strength of the selective advantage of smaller genome size. The stronger the selective pressure, the more rapid the modification would be and the less variation should be found among tRNAs within the mitochondrial genome.

Following the same logic, once RNA editing using the 5′ stem as a template arises in a lineage, loss of sequence conservation in acceptor stem regions should occur in any taxonomic groups that share this trait. Strong selection for small genome size should result in rapid truncation of all tRNAs, whereas weak selection may result in a random pattern of gradual and/or punctuated loss. Truncation should be irreversible and should occur at random to the tRNAs present in the mitochondrion.

Habronattus oregonensis appears to lack all or part of the 3′ acceptor stem in at least 17 of its 22 tRNAs, and it is possible that editing corrects the mispairings present in all 22 of the tRNAs. The only other mitochondrial tRNA sequences published for spiders (all Araneomorphs) are of $tRNA^{Leu(CUN)}$. We inferred the secondary structures for this tRNA from spiders from the families Nesticidae, Eresidae, Araneidae, and Linyphiidae (Hedin 1997*b*; Johannesen and Veith 2001; Johannesen and Toft 2002; Hormiga, Arnedo, and Gillespie 2003). Our examination

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

of these sequences (GenBank accession numbers AF004599, AF374185, AY043262, and AY078743) reveals that they possess mispaired nucleotides in the region that should encode the acceptor stem and that they most likely lack a T arm. Hence, all spiders sampled to date appear to share the same type of secondary structure for $tRNA^{Leu(CUN)}$ as is found in many of the mitochondrial tRNAs of *H. oregonensis*. Therefore, the loss of the T arm and the potential gain of an RNA-editing mechanism to modify the tRNA acceptor stem appear to be shared, derived characters of at least the Araneomorph taxon of spiders. Understanding whether and when RNA editing may have arisen and its possibly effect on tRNA evolution and genome size will require thorough sampling of arachnid lineages.

The greatest reduction in the size of mitochondrial genes in *Habronattus* has occurred in the ribosomal genes. As previously found (Masta 2000), *rrnL* is almost 300 bp shorter than in *Limulus polyphemus*, and in the present study, we report that *rrnS* is almost 100 bp shorter than in *L. polyphemus*. The reduced sizes of these spider ribosomal genes parallels the small sizes of nematode mitochondrial ribosomal genes. It seems possible that the concomitant changes that must have occurred in ribosomal structure in each of these lineages once these ribosomal genes became truncated also facilitated the loss of the T arms of the tRNAs (Wolstenholme et al. 1987; Masta 2000). Changes in ribosomal structure and the evolution of RNA editing may help both to overcome the effects of deleterious mutations and to allow mitochondrial genomes to evolve to a minimal size.

Acknowledgments

We wish to thank Matt Fourcade for valuable computer assistance. This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Berkeley. National Laboratory under contract No. DE-AC03-76SF00098.

Literature Cited

Beard, C. B., D. M. Hamm, and F. H. Collins. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. **2**:103–124.

Black, W. C., IV, and R. L. Roehrdanz. 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. Mol. Biol. Evol. **15**:1772–1785.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. **27**:1767–1780.

Boore, J. L., and W. M. Brown. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. **8**:668–674.

Boore, J. L., T. M. Collins, D. Stanton, L. L. Daehler, and W. M. Brown. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature **376**:163–167.

Boore, J. L., D. V. Lavrov, and W. M. Brown. 1998. Gene translocation links insects and crustaceans. Nature **392**: 667–668.

Börner, G. V., S.-I. Yokobori, M. Mörl, M. Dörner, and S. Pääbo. 1997. RNA editing in metazoan mitochondria: staying fit without sex. FEBS Lett. **409**:320–324.

Clary, D. O., and D. R. Wolstenholme. 1985. The ribosomal RNA genes of *Drosophila* mitochondrial DNA. Nucleic Acids Res. **13**:4029–4045.

Clayton, D. A. 1982. Replication of animal mitochondrial DNA. Cell **28**:693–705.

Covello, P. S., and M. W. Gray. 1993. On the evolution of RNA editing. TIG **9**:265–268.

Dirheimer, G., G. Keith, P. Dumas, and E. Westhof. 1995. Primary, secondary, and tertiary structures of tRNAs. Pp. 93–126 *in* D. Söll and U. L. RajBhandary, eds. tRNA: Structure, biosynthesis, and function. ASM Press, Washington, DC.

Dowton, M. 1999. Relationships among the cyclostome braconid (Hymenoptera: Braconidae) subfamilies inferred from a mitochondrial tRNA gene rearrangement. Mol. Phylogenet. Evol. **11**:283–287.

Francino, M. P., and H. Ochman. 1997. Strand asymmetries in DNA evolution. TIG **13**:240–245.

Gabriel, W., M. Lynch, and R. Bürger. 1993. Muller's ratchet and mutational meltdowns. Evolution **47**:1744–1757.

Gray, M. R. 2001. Speculations on the origin and evolution of editing. Pp. 160–184 *in* B. L. Bass, ed. RNA editing. Oxford University Press, Oxford, UK.

Hedin, M. C. 1997*a*. Molecular phylogenetics at the populations/species interface in cave spiders of the southern Appalachians (Araneae: Nesticidae: *Nesticus*). Mol. Biol. Evol. **14**: 309–324.

———. 1997*b*. Speciational history in a diverse clade of habitat-specialized spiders (Araneae: Nesticidae: *Nesticus*): inferences from geographic-based sampling. Evolution **51**:1929–1945.

Helfenbein, K. G., W. M. Brown, and J. L. Boore. 2001. The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. Mol. Biol. Evol. **18**:1734–1744.

Hormiga, G., M. Arnedo, and R. G. Gillespie. 2003. Speciation on a conveyor belt: sequential colonization of the Hawaiian islands by Orsonwelles spiders (Araneae, Linyphiidae). Syst. Biol. **52**:70–88.

Johannesen, J., and S. Toft. 2002. A test for reproductive separation of alternate generations in a biennial spider, Araneus diadematus (Araneae, Araneidae). J. Arachnol. **30**:65–69.

Johannesen, J., and M. Veith. 2001. Population history of *Eresus cinnaberinus* (Araneae: Eresidae) colour variants at a putative species transition. Heredity **87**:114–124.

Kurabayashi, A., and R. Ueshima. 2000. Complete sequence of the mitochondrial DNA of the primitive opistobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. Mol. Biol. Evol. **17**:266–277.

Lavrov, D. V., J. L. Boore, and W. M. Brown. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol. Biol. Evol. **17**:813–824.

Lavrov, D. V., W. M. Brown, and J. L. Boore. 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forticatus*. Proc. Natl. Acad. Sci. USA **97**:13738–13742.

Lonergan, K. M., and M. W. Gray. 1993. Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. Science **259**:812–816.

Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024

Lynch, M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol. Biol. Evol. **13**:209–220.

———. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. Mol. Biol. Evol. **14**: 914–925.

Macey, J. R., A. Larson, N. B. Ananjeva, and T. J. Papenfuss. 1997. Evolutionary shifts in three major structural features of the mitochondrial genome among iguanian lizards. J. Mol. Evol. **44**:660–674.

Macey, J. R., J. A. Schulte II, and A. Larson. 2000. Evolution and phylogenetic information content of mitochondrial genomic structural features illustrated with acrodont lizards. Syst. Biol. **49**:257–277.

Maddison, D. R., and W. P. Maddison. 2000. MacClade 4: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Mass.

Masta, S. E. 2000. Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TΨC arm. Mol. Biol. Evol. **17**:1091–1100.

Morrison, C. L., A. W. Harvey, S. Lavery, K. Tieu, Y. Huang, and C. W. Cunningham. 2002. Mitochondrial gene rearrangements confirm the parallel evolution of the crab-like form. Proc. R. Soc. Lond. B Biol. Sci. **269**:345–350.

Navajas, M., Y. Le Conte, M. Solignac, S. Cros-Arteil, and J. M. Cornuet. 2002. The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite *Varroa destructor* (Acari: Mesostigmata). Mol. Biol. Evol. **19**:2313–2317.

Ohtsuki, T., G. Kawai, and K. Watanabe. 2002. The minimal tRNA: unique structure of *Ascaris suum* mitochondrial $tRNA^{Ser}_{UCU}$ having a short T arm and lacking the entire D arm. FEBS Lett. **514**:37–43.

Ojala, D., J. Montoya, and G. Attardi. 1981. tRNA punctuation model of RNA processing in human mitochondria. Nature **290**:470–474.

Okimoto, R., and D. R. Wolstenholme. 1990. A set of tRNAs that lack either the TψC arm or the dihydrouridine arm: towards a minimal tRNA adaptor. EMBO J. **9**:3405–3411.

Perna, N. T., and T. D. Kocher. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. **41**:353–358.

Roehrdanz, R. L., M. E. Degrugillier, and W. C. Black, IV. 2002. Novel rearrangements of arthropod mitochondrial DNA detected with long-PCR: applications to arthropod phylogeny and estimation. Mol. Biol. Evol. **19**:841–849.

Sambrook, J., and D. W. Russell. 2001. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York.

Shultz, J. W. 1990. Evolutionary morphology and phylogeny of arachnida. Cladistics **6**:1–38.

Smith, M. J., A. Arndt, S. Gorski, and E. Fajber. 1993. The phylogeny of echinoderm classes based on mitochondrial gene arrangements. J. Mol. Evol. **36**:545–554.

Terrett, J. A., S. Miles, and R. H. Thomas. 1996. Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). J. Mol. Evol. **42**:160–168.

Weygoldt, P. 1998. Evolution and systematics of the Chelicerata. Exp. Appl. Acarol. **22**:63–79.

Wheeler, W. C., and C. Y. Hayashi. 1998. The phylogeny of the extant chelicerate orders. Cladistics **14**:173–192.

Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Pp. 173–216 *in* D. R. Wolstenholme and K. W. Jeon, eds. Mitochondrial genomes. Academic Press, New York.

Wolstenholme, D. R., J. L. Macfarlane, R. Okimoto, D. O. Clary, and J. A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA. **84**:1324–1328.

Yamazaki, N., R. Ueshima, J. A. Terrett et al. (12 co-authors). 1997. Evolution of pulmonate gastropod mitochondrial genomes: comparisons of gene organizations of Euhadra, Cepaea and Albinaria and implications of unusual tRNA secondary structures. Genetics **145**:749–758.

Yokobori, S., and S. Pääbo. 1995. Transfer RNA editing in land snail mitochondria. Proc. Natl. Acad. Sci. USA. **92**:10432–10435.

William Martin; Associate Editor

Accepted January 6, 2004

Downloaded from https://academic.oup.com/mbe/article/21/5/893/1014089 by guest on 23 April 2024