Mitochondrial genome sequence of Unionicola foili (Acari: Unionicolidae): a unique gene order with implications for phylogenetic inference

Brian R. Ernsting · Dale D. Edwards · Katie J. Aldred · Jeffrey S. Fites · Caroline R. Neff

Received: 31 December 2008 / Accepted: 22 March 2009 / Published online: 8 April 2009 Springer Science+Business Media B.V. 2009

Abstract The mitochondrial genome of *Unionicola foili* is circular, 14,738 bp in length, and contains several notable features. The sequence and annotation revealed a unique gene order, continuing a pattern of highly rearranged mitochondrial genomes in the Trombidiformes. U. foili mitochondrial tRNA sequences predict non-canonical secondary structures for these molecules, and our annotation suggests an in-frame fusion between the *nad4L* and nad5 genes in this genome. The unique gene order and unusual tRNA structures could serve as idiosyncratic characters and have the potential to be phylogenetically informative.

Keywords *Unionicola* \cdot Mitochondrial genome \cdot Gene rearrangement \cdot $tRNA$ secondary structure \cdot Molecular phylogeny \cdot Water mites

Introduction

Water mites of the genus *Unionicola* (Acariformes: Prostigmata) are common symbionts of freshwater sponges and mollusks. More than half of the known species of Unionicola are symbiotic with freshwater mussels, living on the gills or mantle and foot of their hosts and using these tissues as sites of oviposition (Vidrine [1996b\)](#page-11-0). Some species of mussel-mites depend upon their hosts only for sites for oviposition and post-larval resting stages, while others are obligate symbionts of their hosts throughout post-larval development (Mitchell [1955\)](#page-10-0). The evolution of unionicolid mussel-mites appears to be closely tied to the evolutionary history of their hosts, given that major clades of mussels harbor unique assemblages of *Unionicola* subgenera (Vidrine [1996a](#page-11-0)).

Although the identification and classification of Unionicola have been reasonably well documented, the phylogenetic systematics of the group is less well understood. In a recent study (Ernsting et al. [2006](#page-10-0)), we used heterogeneity in sequence data of the mitochondrial cytochrome oxidase subunit I $(coxI)$ gene to construct a hypothesis of evolutionary relationships among a number of closely related species of mites within the subgenus

B. R. Ernsting (\boxtimes) \cdot D. D. Edwards \cdot K. J. Aldred \cdot J. S. Fites \cdot C. R. Neff Department of Biology, University of Evansville, Evansville, IN 47722, USA e-mail: be6@evansville.edu

Unionicola (formerly *Parasitatax*). However, as this analysis was expanded to include a more diverse assemblage of mites from several additional *Unionicola* mussel-mite subgenera of North America, we were unable to generate a well-supported phylogeny (D. D. Edwards et al., manuscript submitted). The relatively rapid rate of evolution of the $\cos l$ sequence led us to take two complementary approaches to generating molecular data that could result in better phylogenetic resolution. First, we amplified and sequenced nuclear rRNA sequences from a number of *Unionicola* mites but found that, in contrast to the *cox1* data, these sequences were not sufficiently substituted to be informative at low to intermediate taxonomic levels.

The second approach involves sequencing the complete mitochondrial genome of a representative species of Unionicola. This genome sequence would be expected to be informative on two counts. The complete genome would allow us to choose candidate mitochondrial loci that are likely to be phylogenetically informative for Unionicola species, and the sequence will provide baseline data helpful for designing PCR primers to amplify the chosen loci from other *Unionicola* species. Perhaps more importantly, the complete genome might be expected to yield idiosyncratic markers: molecular character states that, unlike sequence data, represent unique events, and can be used as synapomorphies in phylogenetic reconstruction (Murrell et al. [2003](#page-10-0)). As the number of complete mitochondrial genomes increases rapidly, these small genomes are excellent candidates for improving phylogenetic reconstruction in cases where homoplasy and mutational saturation in sequence data can lead to unresolved phylogenies (Boore [2006\)](#page-10-0).

Within the chelicerates, genome structure synapomorphies are relatively rare in some clades, including soft ticks (Shao et al. [2004](#page-11-0)), but abundant in others, including spiders (Masta and Boore [2008\)](#page-10-0) and the Trombidiformes mites (Van Leeuwen et al. [2006;](#page-11-0) Shao et al. [2005\)](#page-11-0). Of the six species of Trombidiformes whose completed mitochondrial genomes have been submitted to GenBank, there are five unique gene orders. Additionally, given a correlation between rates of nucleotide substitution and genome rearrangement in animal mitochondria (Xu et al. [2006](#page-11-0); Shao et al. [2003](#page-11-0)) the highly derived $\cos I$ sequence data among *Unionicola* subgenera suggest that unionicolid mites may contain informative genome rearrangements.

The current work presents the sequence and annotation of the complete mitochondrial genome of Unionicola foili. The highly rearranged genome and non-canonical tRNA structures seen in this species are presented in comparison to other Trombidiformes mites, and are proposed as potentially informative in reconstructing the phylogeny of Unionicola mites.

Materials and methods

Samples and DNA extraction

Unionicola foili individuals used in this study were removed from host mussels (Utterbackia imbecillis) collected from ponds at various locations in Vanderburgh County, Indiana. Specimen processing and DNA extraction were carried out as previously described (Ernsting et al. [2006](#page-10-0)). Briefly, mites were removed from host mussels, washed, and frozen. Individual mites were ground and total DNA was prepared using a DNeasy Tissue Kit (Qiagen).

PCR amplification

We used a two-stage approach to amplify the mitochondrial genome of U . *foili* from total genomic DNA. In the first stage, fragments of the genome were amplified using regular PCR with primers designed based on the published sequences of related mites. The amplified fragments were then sequenced, and these sequence data were used to design a second set of primers specific for U. foili. The U. foili-specific primers were then used in long-range PCR and standard PCR reactions that amplified the mitochondrial genome in three overlapping fragments.

For the cob gene, primers COB-F and COB-R (Table 1) were used in 50 *l*l PCR reactions with Promega PCR Master Mix (Promega). PCR conditions were an initial denaturation at 94 \degree C for 5 min, followed by 30 cycles of denaturation at 94 \degree C for 1 min, annealing at 52° C for 1 min, and extension at 72° C for 2 min, with a 5 min final extension at 72 $\rm{°C}$. PCR products were visualized on 1% agarose gels and purified using a QIAquick PCR Purification Kit (Qiagen). For the $\cos l$ gene, PCR amplification and DNA sequencing were as described (Ernsting et al. [2006\)](#page-10-0).

Long PCR was performed using a GeneAmp XL PCR Kit (Applied Biosystems). For the region from cox1 to nad4 (the CN template), primers were CN-C1 and CN-N1 (Table 1), with the CN-C1 primer being based on U. foili cox1 sequence, and the CN-N1 primer a conserved PCR primer originally described by Simon and coworkers (Simon et al. [1994\)](#page-11-0), and subsequently used to amplify a fragment of the mitochondrial genome of *Varroa destructor* (Navajas et al. [2002\)](#page-11-0). The PCR conditions were an initial 2 min 92° C denaturation, followed by 10 cycles of denaturation at 92°C for 10 s, annealing at 48°C for 30 s, and extension at 68° C for 5 min. The next 20 cycles allowed additional time for extension: denaturation at 92°C for 10 s, annealing at 48°C for 30 s, and extension at 68°C for 5 min plus 20 s per cycle. The PCR included a final extension of 7 min at 68°C. For the region from $\cos l$ to \cosh (the CB template), primers were CB-C1 and CB-B1 (Table 1), both based on U. foili sequence for these genes, and long PCR conditions were identical to those used for the amplification of the CN template.

The final fragment, the BN template, consisted of the region from *cob* to *nad4*, and was amplified using standard PCR as described above. Primers were BN-B1 and BN-N1 (Table 1), and PCR conditions were an initial denaturation at 94° C for 5 min, followed by

Primer name	Sequence $(5'–3')$	Reference		
$COB-F$	TTTTGAGGTGCAACAGTAATTAC	This work		
$COB-R$	AAATATCATTCAGGTTGAATATG	This work		
FB.	GCACCAGATATAGCTTTTCCACG	Ernsting et al. (2006)		
$RC-2$	GGATAATCTGAATAACGACGAGG	Ernsting et al. (2006)		
$CN-C1$	CCTGTTTTAGCTGGGGCTATTAC	This work		
$CN-N1$	AAAGCTCATGTTGAAGCTCC	Simon et al. (1994)		
$CB-C1$	GTTCATCCTGTTCCTGCTCC	This work		
$CB-B1$	CCCAGCTAATCCTATATCAACTCCC	This work		
$BN-N1$	AAGACCCCCAAATCCCCCAAAC	This work		
$BN-B1$	TCCTACCGAAAACCCTCCTC	This work		

Table 1 PCR primers used for the amplification of the *Unionicola foili* mitochondrial genome

30 cycles of denaturation at 94 $\rm ^{o}C$ for 1 min, annealing at 52 $\rm ^{o}C$ for 1 min, and extension at 72° C for 2 min and 30 s with a 5 min 72° C final extension.

DNA sequencing

Each of the three overlapping fragments was sequenced using a primer-walking strategy where initial sequence reads were carried out with the primers used in the PCR amplification, and subsequent primers were designed based on returned sequence data. PCR products were visualized on 0.8 or 1% agarose gels, and fragment size was estimated using R_f plots. PCR products were excised from the gel, purified using a QIAquick Gel Extraction Kit (Qiagen), and sequenced commercially by SeqWright DNA Technology Services (Houston, TX).

Sequence assembly and annotation

In our primer-walking approach, we emphasized long overlaps between steps such that the average base was sequenced 3.4 times. Sequence assembly was carried out using CodonCode Aligner software (CodonCode), and contigs were exported as FASTA files. Protein and rRNA gene boundaries were determined using the BLAST (Altschul et al. [1990,](#page-9-0) [1997\)](#page-9-0) and ORF-evaluation tools in the MacVector software suite (MacVector, Inc.). tRNA genes were annotated using tRNAscan-SE (Lowe and Eddy [1997\)](#page-10-0) on sequence regions not previously assigned to protein or rRNA genes. For some tRNA identification, the source parameter on tRNAscan-SE was set to ''Nematode Mitochondrion''. tRNAs that were not identified by the tRNAscan software were assigned using a combination of by-eye and Clustal W version 2.0 (Larkin et al. [2007\)](#page-10-0) sequence alignments involving putative U. foili tRNAs and the corresponding tRNA sequences reported for other Trombidiformes mites.

Results

General features

The circular mitochondrial genome of U. foili (GenBank accession EU856396) is 14,738 bp in size, 73% $A + T$, and contains the 37 genes typical for animal mitochondria: 13 protein-coding genes, 22 tRNA genes, and 2 rRNA genes. Twenty-two genes (9 proteincoding and 13 tRNA) are located on the majority (J) strand, with the remaining 4 proteincoding, 2 rRNA and 9 tRNA genes on the minority (N) strand. In addition, the U. foili genome contains two large noncoding regions of 645 and 387 bp, each of which contains regions of predicted secondary structure (Mathews et al. [1999](#page-10-0)) which could serve as replication origins (Zhang et al. [1995](#page-11-0)).

The U. foili mitochondrial genome is the seventh representative from the order Trombidiformes and shares general characteristics with these relatives (Table [2\)](#page-4-0). First, the $A + T$ content is typical for a Trombidiformes mite. Second, the overall genome size continues a pattern of compact genomes in this group. Finally, both protein-coding genes and tRNA genes are reduced in size when compared to the homologous sequences in the mitochondrial genome of Drosophila yakuba, a well-studied arthropod (Clary and Wolstenholme [1985;](#page-10-0) Crease [1999;](#page-10-0) Shao et al. [2005\)](#page-11-0).

Species	Accession	(bp)	Length $A + T$ Protein (%)	length $(\%)^{\rm a}$	Average tRNA size (bp)	Reference
Ascoschoengastia $TATW-1$	NC 010596 16,067 71			-5.8	57	Mitani et al. (2008)
Leptotrombidium akamushi	NC 007601	13,698 68		-8.2	53	Shao et al. (2006)
L. deliense	NC 007600	13.731	-70	-8.2	54	Shao et al. (2006)
L. pallidum	NC 007177 16,779		71	-8.0	54	Shao et al. (2005)
Tetranychus urticae	NC 010526 13,103 85			-8.8	52	Van Leeuwen et al. (2006)
Walchia hayashii	NC 010595 14,857 73			-5.7	54	Mitani et al. (2008)
Trombidiformes average		14,706 73		-7.5	53	
U. foili	NC 011036 14,738		73	-4.7	55	This work
D. yakuba	NC 001322 16,019 79			N/A	67	Clary and Wolstenholme (1985)

Table 2 Comparison of *Unionicola foili* mitochondrial genome features with those of other Trombidiformes and the fruit fly Drosophila yakuba

Variation in average protein length compared to D. yakuba = 100% × (total encoded amino acids $-$ total encoded amino acids $(D. yakuba)$ /total encoded amino acids $(D. yakuba)$

Gene arrangement

Like the mitochondrial genomes of other Trombidiformes mites, the U. foili mitochondrial genome (Fig. [1](#page-5-0)) is highly rearranged when compared to that of the hypothetical ancestor of the arthropods, whose gene order is represented by the horseshoe crab Limulus polyphemus (Boore et al. [1995](#page-10-0); Shao et al. [2005](#page-11-0)). Additionally, the U. foili gene arrangement is highly derived even when compared to other Trombidiformes, continuing a pattern of extensive genome rearrangement among species in this taxon. Specifically, U. foili contains nine gene boundaries that are unique, eight of which involve tRNA translocations. Although the nad4L and nad5 genes are annotated individually in the GenBank accession for this sequence, the nineth unique gene boundary involves an in-frame fusion between these genes. Because our initial annotation did not indicate a stop codon for the *nad4L* gene, this region was amplified and resequenced independently from six additional individuals, confirming our original annotation.

In addition to the apparent *nad4L-nad5* gene fusion, we noted seven pairs of overlapping genes in the U. foili mitochondrial genome: for genes on the same strand, cox1 and $\cos 2$ overlap by one base, $\sin 6$ and $\sin 8$ overlap by seven bases, $\sin 6$ and $\cos 3$ overlap by one base, and *nad6* and *cob* overlap by one base. Similar overlaps between these gene pairs have been annotated in other Trombidiformes species with completed genomes (Shao et al. [2006,](#page-11-0) [2005\)](#page-11-0) and GenBank accession AB300500. For genes on opposite strands, tRNA-Thr and tRNA-Tyr overlap by one base, tRNA-Gln and tRNA-Ile overlap by one base, and tRNA-Arg and tRNA-Val overlap by three bases. The tRNA-Thr tRNA-Tyr gene boundary is unique to U. foili, and therefore the overlap has not been annotated in other species. The three-base overlap between tRNA-Arg and tRNA-Val, however, is similar to an annotated overlap in Walchia hayashii, and the tRNA-Gln tRNA-Ile overlap is similar to an overlap

Fig. 1 Comparison of the mitochondrial genome structures of Trombidiformes mites and the horseshoe crab Limulus polyphemus. Light gray protein-coding genes, unshaded tRNA genes, dark gray rRNA genes. Circular genome sequences were linearized at the $5'$ end of the *cox1* gene. Genes transcribed in the same direction as $\cos 1$ (left to right) are shown below the line, and genes transcribed in the opposite direction are shown above the line. For protein-coding and rRNA genes, the gene names are shown either in the rectangle or above or below the line. For tRNA genes, gene names are abbreviated with the single-letter abbreviation for the amino acid specified. Serine and Leucine each have two tRNA genes in these mitochondrial genomes, which are distinguished by their anticodons $(5'-3')$ S1: GCU, S2: UGA, L1: UAG, L2: UAA. Protein-coding and rRNA genes are drawn to approximate scale, but tRNA genes are exaggerated for clarity. Formatting for this figure is adapted from the OGRe database (Jameson et al. [2003\)](#page-10-0)

between these genes in L. polyphemus (Lavrov et al. [2000\)](#page-10-0). Since all three pairs of overlapping tRNA genes are on opposite strands, they would be processed from different transcripts, and these overlaps would not be expected to result in truncated tRNAs.

Protein-coding genes

Of the thirteen protein-coding genes annotated, twelve begin with one of the common start codons for mitochondrial genes (ATA, ATG, ATT). Only one gene, *nad5*, is annotated as having a TTG start codon. In addition to *nad4L*, four other genes, $\cos 2$, $\cos 3$, *nad5*, and cob were found to have incomplete stop codons. In each of these cases, the gene is followed by a tRNA gene, and stop codons for these genes are likely to be generated by cleavage and polyadenylation of the primary transcript (Ojala et al. [1981](#page-11-0)). Interestingly, although the *nad5* stop codon is generated by $3'$ processing, the tRNA is located on the opposite strand, suggesting that the secondary structure of the strand complementary to that which specifies the functional tRNA is also able to be recognized by the posttranscriptional processing machinery. This type of processing has also been reported for the mitochondrial nad3 of the spider *Ornithoctonus huwena* (Oiu et al. [2005](#page-11-0)).

rRNA genes

The genes for the small (rrnS) and large (rrnL) ribosomal RNAs were annotated to extend to the boundaries of neighboring genes. This leaves open the possibility that the rRNA genes may be somewhat shorter than the annotated sequence indicates, but if these differences exist, they are likely to be small, as the sizes of both rRNAs are similar to those annotated for the other Trombidiformes mites shown in Fig. [1](#page-5-0). $rrnS$ is annotated at 649 bp in U. foili with a range of sizes in other Trombidiformes mites of 601-679 bp, while rrnL is annotated at 1,016 bp with a range in the same group of 992–1,046 bp. In U. foili, as in most mitochondrial genomes of metazoans, the rRNA genes are located near each other on the minority (N) strand. Unusually, these genes are separated by the protein-coding nad1 gene.

Secondary structure of tRNAs

Twenty-two transfer RNA genes were annotated using a combination of tRNAscan-SE, alignments, and by-eye inspection. In addition to being reduced in size (Table [2\)](#page-4-0) U. foili tRNAs share structural features with the tRNAs of other Trombidiformes mites. Figure [2](#page-7-0) presents the inferred secondary structures of tRNA molecules from U. foili and two other Trombidiformes mites: Leptotrombidium akamushi (Shao et al. [2006\)](#page-11-0) and the prostigmatid mite Walchia hayashii (Mitani et al. [2008](#page-10-0)). In general, sequence conservation is high only in the anticodon loop. In base-paired stems, sequence conservation is often moderate to non-existent, although the base-pairing itself is strongly conserved. The exception to this pattern of limited sequence conservation is the D-arm. Most tRNAs with D-arms show moderate to strong sequence conservation for this region. Like the tRNAs of other Trombidiformes species (Shao et al. [2005\)](#page-11-0), the U. foili mitochondrial tRNAs show structures unusual for arthropods. All but seven of the U. foili tRNAs in Fig. [2](#page-7-0) lack either a T-arm or a D-arm, and are thus unable to fold into the canonical tRNA cloverleaf structure. For most tRNAs in Fig. [2](#page-7-0), the loss of D- or T-arms is conserved: with few exceptions, all three species share a similar predicted secondary structure.

Discussion

Overall, the results of this study suggest that the mitochondrial genome structure of U. foili shares important similarities to those of other Trombidiformes mites. The emerging

Gene		Species AA arm		D-arm		AC-arm	V-loop	T-arm	AA-arm
Α	Uf La Wh	AGATAAA AAAT AGAAAAG TCT AGAGAAA ATA				TTTGAATTGCACTCAAA TAGAGTTTGCACCTCTT ATAACTTTGCAAGTTTT	AAAA TAAA AAAA	GTTTCCC-AC GCATTTTAGC GCCAACAAGC	TTTATCT CTTTTCT TTTCTCT
Ċ	Uf La wh	ACAAGTA AA TGATCAT CA ATGAGTC T		GTTTTTTG---AAT GTCTTTTGAAAGAT <u>GTATTTTTTATAAT</u>	G A TT	TTAAATTGCAAATTTAA TTAAATTGCAAATTTAA TAAGGTTGCAAACCTTG	AGGAAATAA AGAAAA ATACTTAA		AGCTTGT AT <u>TGATC</u> GACTTAT
$\mathbf D$	Uf La Wh	AAAGTGT TA ACTTAAA CA <u>ATCTTAC</u> AA		ATAAAAATATTAT GTTATT----TAC GTATATAATATAC	A A m,	TTTCACTGTCAGTGAAA TTAAACTGTCGGTTTAA AGTTACCGTCAGTAACC	AATT AGAG CAAA	GTTG------AATC ACACTTT GCTTTATACAAAGC GCAAAGTTT-AAGC	TTTAAGT GTAAGAT
E	Uf T.a Wh	TCTTCTT TA CTCTCTT TA ATCCCTT AA		GTATAAAAAGTAC <u>GTCT</u> TAAA <u>A-GAT</u> GTAG-AAAA-TAC	A T. G	TTTATTTTTCATATAAA TTCGAGTTTCATTCGAA CTTGAGTTTCAATCAAG	AGGAAA AGAA AAAA	CCA--AAATGG TCATTTAAAGA	AAGAAGA AAGAGAG AAGGGAA
$\overline{\mathbf{F}}$	Uf La Wh	GCATTTT GA TT <u>TTGTA</u> TA TTTCGAG TA		GCTTGATTGAAAGT GCTTATAT-AAAGT GCTTTTATTAAAGT	G G G	TAATTTTGAAGAAATTA TIGTTTTGAAGAAACAA TTGCTTTGAAGAAGCAG	AGTAGTTT AGAAAAA AGAT		TGTATGC TTTCTAA
G	Uf La Wh	AAACAGA TAAA AGGTGTG ACAAG AAACGAA AAA				ATTAACTTCCAATTAAT <u>AATAATTTCCAATTAAT</u> GATAACTTCCAATTATC	AACCA CATT AACA	GCC-----AAAAGC GATTCA--CCCATC GGACTAAACCAACC	TCTGTTT CGCACTT TTCGTTA
H	Uf La Wh	TTATTCA CA CTTCTTT TA TTTTAGT TT		GTTTAGGAAAT GTTTAT-AAAT ATTT-GAAAAT	A A Ť	TTAAGTTGTGATCTTAA TCAAATTGTGATTTTGA TTAAGTTGTGGTCTTAG	AGAGAAT GGAATTT AGA		TGAGTAA TAGGAAG ATAAAAA
I	Uf La Wh	AGTGGAA TTAT AATGATA AAAAT	AAGTATA ATACATAA			ATATTTCGATACAATAT TTCTTTTGAT GGAAGAA TCTTCTTGATATGAAGA	TTAA AACT AAAA	GG--AAGTCC GAC-ATACTC GGCAAAAACC	TATACTT TTTCACT: TATCATT:
K	Uf La Wh	CTTTTAG TG GTTTTGA AA TTCTTAG AA		GCTGACAAAAGC GCT-AC-TCAGC ACTAAAAATAGT	A A A	TTGGTCTCTTAAACCAA CTGAACTTTTAATTCAG TTGGTCTCTTAAACCAA	AAAAT AGAA ATTAA	GATCTTA--TC GGGAAATCCCC GAGCTTGACTC	CTAAAAG TCAAAAT CTAAGAA
L1	Uf La Wh	AGCTAAA AA ACTTAAG AT	AATCTTG AATTTTAT	GTATATATTATAT TTTAAATA	A T	TTGAATTTAGAATTCAA TAGGCCTTAGGAGTCTA TCAGTTTTAGGAATTGA	GTAA TTTT	GAGT-GTTTT GGCT-TCGCC GTTGAAAGAC	ATTAGCT CTTAAGT CGAGATT
L2	Uf La Wh	ATCGATG TG AAGTGAT GA AAACAGG AA		GCAGAT--GAATTGT G AACATTTTAGATGTA A TGCGTTTT---TGCA	T	TAGAGTT TAA GTCTCTT TAGGACT TAA GTTCCTT TAGAATTTAAGTTTCTA	TTAT TAAA TTTTG	-GTTTT-AC- CTTTTTTAAG AGTGTTCTCT	CTTCGAT CTCACTT CTTGTTT:
Μ	Uf La Wh	TATAAGA TA AGAAAGG TA GAAGAAA TA		AGCTATTT-AAGCT AGCTATATCAAGCT AGCTATAT-AAGCT	A Α T.	ATGGGTTCATACCTCAT GCGGATTCATAATCCTT CAGGGCTCATAACTCTG	TGAA TGAA TGAA	AGGCC---ACCT ACATTTTTATGT ACAA----AAGT	TCTTATA CCTTTCC TTTCTTC
N	Uf La Wh	TTGATAG AA TTTAGGA AAATT TTAGAAT TATA		GCCCAAGGC	Α	TTTTACTGTTAATAAAA <u>GAGTA</u> TT GTT AA <u>TACTT</u> AGATATTGTTAATTTCT	AGAA AGAT TAAA	GAG---AAATC GGCTATT--CC GGCAAATAACC	CTATCAA TCCTAAA ATTCTAA
P	Uf La Wh	CAAAAGT TA CAGAAAA TA CAAGAAT TA		GTTTATA-AAGAAT GTTTATTTAAGAAC GTTTAAT--AGAAC	G Α A	TTGGCTTTGGGGGCTTA TCTACTTTGGGGGTGGG ATTACTTTGGGTGTAAT	GGGAATTA AGG-AAAA AGG-TTAA		TCTTTTG TTTTCTG TTTCTTG
Q	Uf La Wh	TTACTTT TG TATCAAA TG AAGTAGA CG		GTGTATTGTCAT GTGTA-AAGCAT TTCT----GGAA	G C A	TCAAAATTTGATTTTGA TCCCCTTTTGGGGGGGA GAGTGTCTTGTTAATTT	AGGAAAA TAA	AGGTTTTT ATCTTAT	ATAGTAA ATTGATA TTTGTTT
R	Uf La Wh	AATTGAA AAAG GTCGAAG AAA ATTGAGA ATT				AAGGATTTCGAATCCTA AAAGACTTCGGATCTTT AAGAATTTCGGATTCTT	AAAA AAAG TTCT	GCTT-AGGC GGAAAATCC GAGGTAATC	TTCAGTT CTTCAAC! TCTCATT
S1	Uf La Wh	GAATGAA AA GAAAAGA AAAA GAGAACT ACAA				AAGAACTGCTAATTCTT GTAGGTTTCTAGCCTTC AGTTGCTGCTAACTACT	TAAATA AGTTT TAATA	-GTT--TATC- -GCTAATTGC- AGCTAAAAGCT	TACATTC! TCTTTTC <u>AATTCTC</u>
S ₂	Uf T.a Wh	AACTTTC TTTTTA ATTTCCT CTAACT ACTACTT TAAA				CTTATCTTGAAAATAAG CCCGTCTTGAAAACAGG TCTGTTT TGA AA <u>ACAGA</u>	TTT ATCT AAAA	TAGAAAAAATCTA $-\underline{AGA}TCATTTTCT-$ -AGAG-AAATCT-	GAAAGTT <u>AGGAAAT</u> AAGTAGT.
T	Uf T.a Wh	GACAATA AT <u>GAAAGTT</u> TA GAAAGAC TA		AGTTTATAA---AAAAT A -GTTTAAAATTTA <u>AAAT</u> A -GTTTATAA--GAAAAC G		TGAATTTTGTAATTTCA TIGATTTTGTAAATCAA TTGATTTTGTAAATCAA	AGAAGAA AGATAAAA AAAAAGAA		AAATGTC <u>АТСТТТС</u> AACTTTC
V	Uf La Wh	TTAACAG TTTA TGATATG TTTTA CAATCAA AAT				TTTGACTTACAATCAAA <u>AAAAATTTACAATTTTT</u> TTTAGCTTACAACTAAT	AAAA TCTC TATT	GCT--TTTGC CCGAATTTGG CCTAA--AGG	A CTGTTAA CATATCA [®] TTGATTG
W	TTF La Wh	AAAGCAA TA AGGACTC TA AAGATTA TA		GGTTATA-AAAACT GGTTAAATTAAACT AGTTATA-TAAACT	A T A	TCATCCTTCAAAGATGA TTGGCTTTCAAATCCAA TTAACCTTCAAAGTTAA	AGAAAA AGATCGG AAAAA		AAGCTTT CAGTTCA TAAGTCT.
Υ	Uf La wh	GTCTAAA AA GTAAGAG CAAA GTTAGAA TTTA		GTTTTGTGAAT	G	GAGAATTGTAAATTCTT AATGATTGTAAATCATT TTGGATTGTAAATCCTA	AGGGAAA AAAA TTTA	GAATAG-TC GATTTGTTC	TTTGGAT CTCTTTCI TTCTATC

Fig. 2 Inferred secondary structures of tRNAs from Unionicola foili (Uf), Leptotrombidium akamushi (La), and Walchia hayashii (Wh). Acceptor arm, D-arm, anticodon arm, and T-arm are regions involved in basepairing, with putatively-paired bases underlined. Anticodon bases are bold text. Shaded bases are conserved across these three species

characteristics of this group include compact, extensively rearranged genomes and tRNA structures that are highly derived when compared to those of other arthropods. Although these characteristics are not universally shared across the arachnids, there are a growing number of examples of rearranged genomes and unusual tRNA structures in non-Trombidiformes Arachnida (Jeyaprakash and Hoy [2007](#page-10-0); Fahrein et al. [2007;](#page-10-0) Masta and Boore [2004;](#page-10-0) Domes et al. [2008\)](#page-10-0). The wealth of unusual features in these groups has generated intriguing hypotheses on a range of topics including genome recombination (Shao et al. [2005\)](#page-11-0), RNA editing (Masta and Boore [2004](#page-10-0)), loss of tRNA genes (Domes et al. [2008\)](#page-10-0), and the possibility of non-standard translation of some ORFs (Jeyaprakash and Hoy [2007](#page-10-0)). In the U. foili mitochondrial genome, the most unusual aspects are the non-canonical structures of most tRNAs and the apparent gene fusion between nad4L and nad5.

Non-canonical mitochondrial tRNA structures like those seen in the U. foili mitochondrial genome are well documented in only a few groups. Loss of the cloverleaf structure has been most extensively described in representatives of the phylum Nematoda (Wolstenholme et al. [1987;](#page-11-0) Montiel et al. [2006](#page-10-0)) but is now emerging as common among several groups of arachnids, including spiders and mites (Jeyaprakash and Hoy [2007;](#page-10-0) Masta and Boore [2004;](#page-10-0) Shao et al. [2005](#page-11-0)). Recently, an analysis of tRNA structure across several arachnid groups indicated that the propensity to evolve away from the canonical tRNA structure may have arisen independently in diverse lineages (Masta and Boore [2008\)](#page-10-0), and suggested that as-yet unexplored evolutionary pressures drive this striking propensity.

In the U. foili genome, the apparent gene fusion between $nad4L$ and $nad5$ is unprecedented among sequenced mitochondrial genomes. There are at least four possible scenarios for the expression of functional protein(s) from the fused genes. First, RNA editing may create a stop codon after the region is transcribed. mRNA editing is widespread in the mitochondria of trypanosomes (Ochsenreiter et al. [2008\)](#page-11-0), and dinoflagellates (Lin et al. [2002\)](#page-10-0), and in the nuclear transcriptomes of arthropods (Stapleton et al. [2006\)](#page-11-0). Mitochondrial tRNA editing has been demonstrated as a mechanism for correcting mismatched tRNA acceptor stems in land snails (Yokobori and Päabo [1995](#page-11-0)) and proposed as a solution to similar mismatches in the spider *Habronattus oregonensis* (Masta [2000\)](#page-10-0). Yokobori and Päabo [\(1995](#page-11-0)) have proposed possible mechanisms for tRNA editing that involve a basepaired region serving as the template for the edited product. Analysis of the region near the potential nad4L-nad5 editing site did not show significant predicted secondary structure, but this does not rule out template-directed editing that might involve sequences outside of this region. Second, a modified genetic code could lead to the usage of a non-standard stop codon for the nad4L ORF. A modified genetic code has been proposed as a possible explanation for unusual gene structure in the related mite Metaseiulus occidentalis (Jeyaprakash and Hoy [2007](#page-10-0)). Third, it is possible that the proteins are translated as a fusion protein and subsequently processed. Finally, it is also possible that the proteins are expressed as a fusion protein and function without processing, a situation that is possible given the localization of the ND4L and ND5 proteins to the hydrophobic membrane-bound arm of the NADH:ubiquinone oxidoreductase (complex I); (Chomyn et al. [1985;](#page-10-0) Wang et al. [1991](#page-11-0)) of the mitochondrial electron transport chain. As annotated, the downstream nad5 gene would use TTG as its start codon. TTG is a rare start codon in mitochondrial genomes (Bae et al. [2004](#page-10-0)), strengthening the likelihood of either a functional fusion protein or editing of the primary transcript.

Previous work in our laboratory has used the DNA sequence of the mitochondrial $\cos l$ gene to examine phylogenetic relationships among Unionicola mites both within and among subgenera. We were able to use the *cox1* sequence to reconstruct the phylogeny of closely related species within a subgenus (Ernsting et al. [2006\)](#page-10-0). When we extended this approach to examine the phylogeny of the species across subgenera, we encountered the loss of phylogenetic support associated with applying multiparameter statistical corrections to highly derived sequence data (D. D. Edwards et al., manuscript submitted), underscoring the need for additional loci or other informative characters. Since the rate of molecular evolution is well-correlated with the rate of genome rearrangement in arthropods (Shao et al. [2003;](#page-11-0) Xu et al. [2006\)](#page-11-0), and given the pattern set by previously-sequenced arachnid genomes, we predicted that the U. foili mitochondrial genome would also be extensively modified. Not only is the U. foili mitochondrial genome highly rearranged compared to other members of the Trombidiformes (Fig. [1](#page-5-0)), but preliminary data indicate that the mitochondrial genome of *Unionicola parkeri* (a closely related mantle mite from a different subgenus) has yet another unique gene order that has high affinity for the U. foili arrangement but that also has several unique gene-order characters. The rearrangements and other idiosyncratic markers, including tRNA secondary-structural characters (Murrell et al. [2003](#page-10-0)) seen in these groups have potential to be used as synapomorphic character states, allowing us to improve the reliability of phylogenetic inference (Boore [2006](#page-10-0); Hypsa [2006\)](#page-10-0). Additionally, when combined with DNA sequence data and other characters, genome-level characters can contribute to a total-evidence approach to phylogenetic reconstruction (Shao and Barker [2007](#page-11-0); Wortley and Scotland [2006](#page-11-0)).

The wealth of genome-level characters among Trombidiformes mites suggests the potential for these types of characters to contribute to phylogenetic inference in the group. Although a maximum parsimony analysis based on 19 gene-order characters for the taxa in Fig. [1](#page-5-0) did not result in a well-supported phylogeny of Trombidiformes species (data not shown), some features of the genome structures may be phylogenetically informative. In particular, species of *Leptotrombidium* share what would appear to be genome structure synapomorphies in the *nad4L-nad1* region that support their close relationship. Similarly, the genomes of Walchia hayashii and Ascoschoengastia sp. TATW-1 have a unique insertion of the *nad1* gene between the *rrnS* and *rrnL* genes. U. foili shares this synapomorphy, making it possible that this region will be useful in future phylogenetic reconstruction. Finally, some genome structure characters are similar or identical for most members of the group. For example, the gene order in the *nad4-nad6-cob* region is very similar in all Trombidiformes except Tetranychus urticae, which on the basis of its genome structure appears to be quite distantly related to both other Trombidiformes and their ancestors (Limulus). As more mitochondrial genome sequences are reported, the utility of genome-level characters for resolving phylogeny will become increasingly important. Additionally, the sequence of the U. foili mitochondrial genome suggests that the *cob* and rrnS genes may contain regions that are more conserved within the Unionicola than is cox1, raising the possibility that these sequences could contribute to phylogenetic inference. Future studies in our laboratory will be focused on assessing the usefulness of both sequence data and idiosyncratic characters within the *Unionicola*, with the ultimate goal of using both types of molecular data to generate a phylogeny of the genus.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Altschul SF, Madden TL, Schaffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. doi[:10.1093/nar/25.17.3389](http://dx.doi.org/10.1093/nar/25.17.3389)

- Bae JS, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. Mol Phylogenet Evol 32:978–985. doi:[10.1016/j.ympev.2004.03.009](http://dx.doi.org/10.1016/j.ympev.2004.03.009)
- Boore JL (2006) The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol 21:439–446. doi:[10.1016/j.tree.2006.05.009](http://dx.doi.org/10.1016/j.tree.2006.05.009)
- Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM (1995) Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature 376:163–165. doi:[10.1038/376163a0](http://dx.doi.org/10.1038/376163a0)
- Chomyn A, Mariottini P, Cleeter MW et al (1985) Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. Nature 314:592–597. doi: [10.1038/314592a0](http://dx.doi.org/10.1038/314592a0)
- Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of Drosophila yakuba: nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271. doi[:10.1007/BF02099755](http://dx.doi.org/10.1007/BF02099755)
- Crease TJ (1999) The complete sequence of the mitochondrial genome of Daphnia pulex (Cladocera: Crustacea). Gene 233:89–99. doi[:10.1016/S0378-1119\(99\)00151-1](http://dx.doi.org/10.1016/S0378-1119(99)00151-1)
- Domes K, Maraun M, Scheu S, Cameron SL (2008) The complete mitochondrial genome of the sexual oribatid mite Steganacarus magnus: genome rearrangements and loss of tRNAs. BMC Genomics 9:532. doi[:10.1186/1471-2164-9-532](http://dx.doi.org/10.1186/1471-2164-9-532)
- Ernsting BR, Edwards DD, Vidrine MF, Myers KS, Harmon CM (2006) Phylogenetic relationships among species of the subgenus *Parasitatax* (Acari: Unionicolidae: *Unionicola*) based on DNA sequence of the mitochondrial cytochrome oxidase I gene. Int J Acarol 32:195–202
- Fahrein K, Talarico G, Braband A, Podsiadlowski L (2007) The complete mitochondrial genome of Pseudocellus pearsei (Chelicerata: Ricinulei) and a comparison of mitochondrial gene rearrangements in Arachnida. BMC Genomics 8:386. doi:[10.1186/1471-2164-8-386](http://dx.doi.org/10.1186/1471-2164-8-386)
- Hypsa V (2006) Parasite histories and novel phylogenetic tools: alternative approaches to inferring parasite evolution from molecular markers. Int J Parasitol 36:141–155. doi[:10.1016/j.ijpara.2005.10.010](http://dx.doi.org/10.1016/j.ijpara.2005.10.010)
- Jameson D, Gibson AP, Hudelot C, Higgs PG (2003) OGRe: a relational database for comparative analysis of mitochondrial genomes. Nucleic Acids Res 31:202–206. doi:[10.1093/nar/gkg077](http://dx.doi.org/10.1093/nar/gkg077)
- Jeyaprakash A, Hoy MA (2007) The mitochondrial genome of the predatory mite Metaseiulus occidentalis (Arthropoda: Chelicerata: Acari: Phytoseiidae) is unexpectedly large and contains several novel features. Gene 391:264–274. doi[:10.1016/j.gene.2007.01.012](http://dx.doi.org/10.1016/j.gene.2007.01.012)
- Larkin MA, Blackshields G, Brown NP et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. doi:[10.1093/bioinformatics/btm404](http://dx.doi.org/10.1093/bioinformatics/btm404)
- Lavrov DV, Boore JL, Brown WM (2000) The complete mitochondrial DNA sequence of the horseshoe crab Limulus polyphemus. Mol Biol Evol 17:813–824
- Lin S, Zhang H, Spencer DF, Norman JE, Gray MW (2002) Widespread and extensive editing of mitochondrial mRNAS in dinoflagellates. J Mol Biol 320:727–739. doi:[10.1016/S0022-2836\(02\)00468-0](http://dx.doi.org/10.1016/S0022-2836(02)00468-0)
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. doi[:10.1093/nar/25.5.955](http://dx.doi.org/10.1093/nar/25.5.955)
- Masta SE (2000) Mitochondrial sequence evolution in spiders: intraspecific variation in tRNAs lacking the TPsiC Arm. Mol Biol Evol 17:1091–1100
- Masta SE, Boore JL (2004) The complete mitochondrial genome sequence of the spider Habronattus oregonensis reveals rearranged and extremely truncated tRNAs. Mol Biol Evol 21:893–902. doi: [10.1093/molbev/msh096](http://dx.doi.org/10.1093/molbev/msh096)
- Masta SE, Boore JL (2008) Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. Mol Biol Evol 25:949–959. doi:[10.1093/molbev/msn051](http://dx.doi.org/10.1093/molbev/msn051)
- Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288:911–940. doi:[10.1006/](http://dx.doi.org/10.1006/jmbi.1999.2700) [jmbi.1999.2700](http://dx.doi.org/10.1006/jmbi.1999.2700)
- Mitani H, Yuasa S, Takahashi M, Fukunaga M (2008) Mitochondrial gene order and molecular phylogenetic analyses indicate that the Leptotrombidium mite is paraphyletic. Unpublished GenBank Accession
- Mitchell RD (1955) Anatomy, life history, and evolution of the mites parasitizing fresh-water mussels. Miscell Publ_Mus Zool. Univ Mich 89:1–28
- Montiel R, Lucena MA, Medeiros J, Simoes N (2006) The complete mitochondrial genome of the entomopathogenic nematode Steinernema carpocapsae: insights into nematode mitochondrial DNA evolution and phylogeny. J Mol Evol 62:211–225. doi[:10.1007/s00239-005-0072-9](http://dx.doi.org/10.1007/s00239-005-0072-9)
- Murrell A, Campbell NJ, Barker SC (2003) The value of idiosyncratic markers and changes to conserved tRNA sequences from the mitochondrial genome of hard ticks (Acari: Ixodida: Ixodidae) for phylogenetic inference. Syst Biol 52:296–310
- Navajas M, Le Conte Y, Solignac M, Cros-Arteil S, Cornuet JM (2002) The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite Varroa destructor (Acari: Mesostigmata). Mol Biol Evol 19:2313–2317
- Ochsenreiter T, Cipriano M, Hajduk SL (2008) Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. PLoS ONE 3:e1566. doi:[10.1371/journal.pone.0001](http://dx.doi.org/10.1371/journal.pone.0001566) [566](http://dx.doi.org/10.1371/journal.pone.0001566)
- Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474. doi[:10.1038/290470a0](http://dx.doi.org/10.1038/290470a0)
- Qiu Y, Song D, Zhou K, Sun H (2005) The mitochondrial sequences of Heptathela hangzhouensis and Ornithoctonus huwena reveal unique gene arrangements and atypical tRNAs. J Mol Evol 60:57–71. doi[:10.1007/s00239-004-0010-2](http://dx.doi.org/10.1007/s00239-004-0010-2)
- Shao R, Barker SC (2007) Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. Parasitology 134:153–167. doi:[10.1017/S0031182006001429](http://dx.doi.org/10.1017/S0031182006001429)
- Shao R, Dowton M, Murrell A, Barker SC (2003) Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol Biol Evol 20:1612–1619. doi:[10.1093/](http://dx.doi.org/10.1093/molbev/msg176) [molbev/msg176](http://dx.doi.org/10.1093/molbev/msg176)
- Shao R, Aoki Y, Mitani H, Tabuchi N, Barker SC, Fukunaga M (2004) The mitochondrial genomes of soft ticks have an arrangement of genes that has remained unchanged for over 400 million years. Insect Mol Biol 13:219–224. doi[:10.1111/j.0962-1075.2004.00447.x](http://dx.doi.org/10.1111/j.0962-1075.2004.00447.x)
- Shao R, Mitani H, Barker SC, Takahashi M, Fukunaga M (2005) Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, Leptotrombidium pallidum. J Mol Evol 60:764–773. doi:[10.1007/s00239-004-0226-1](http://dx.doi.org/10.1007/s00239-004-0226-1)
- Shao R, Barker SC, Mitani H, Takahashi M, Fukunaga M (2006) Molecular mechanisms for the variation of mitochondrial gene content and gene arrangement among chigger mites of the genus Leptotrombidium (Acari: Acariformes). J Mol Evol 63:251–261. doi:[10.1007/s00239-005-0196-y](http://dx.doi.org/10.1007/s00239-005-0196-y)
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am 87:651–701
- Stapleton M, Carlson JW, Celniker SE (2006) RNA editing in *Drosophila melanogaster*: new targets and functional consequences. RNA 12:1922–1932. doi:[10.1261/rna.254306](http://dx.doi.org/10.1261/rna.254306)
- Van Leeuwen T, Tirry L, Nauen R (2006) Complete maternal inheritance of bifenazate resistance in Tetranychus urticae Koch (Acari: Tetranychidae) and its implications in mode of action considerations. Insect Biochem Mol Biol 36:869–877. doi:[10.1016/j.ibmb.2006.08.005](http://dx.doi.org/10.1016/j.ibmb.2006.08.005)
- Vidrine MF (1996a) North American Najadicola and Unionicola: diagnoses and distributions. Gail Vidrine Collectibles, Eunice
- Vidrine MF (1996b) North American Najadicola and Unionicola: systematics and coevolution. Gail Vidrine Collectibles, Eunice
- Wang DC, Meinhardt SW, Sackmann U, Weiss H, Ohnishi T (1991) The iron-sulfur clusters in the two related forms of mitochondrial NADH: ubiquinone oxidoreductase made by Neurospora crassa. Eur J Biochem 197:257–264. doi[:10.1111/j.1432-1033.1991.tb15906.x](http://dx.doi.org/10.1111/j.1432-1033.1991.tb15906.x)
- Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, Wahleithner JA (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc Natl Acad Sci USA 84: 1324–1328. doi:[10.1073/pnas.84.5.1324](http://dx.doi.org/10.1073/pnas.84.5.1324)
- Wortley AH, Scotland RW (2006) The effect of combining molecular and morphological data in published phylogenetic analyses. Syst Biol 55:677–685. doi[:10.1080/10635150600899798](http://dx.doi.org/10.1080/10635150600899798)
- Xu W, Jameson D, Tang B, Higgs PG (2006) The relationship between the rate of molecular evolution and the rate of genome rearrangement in animal mitochondrial genomes. J Mol Evol 63:375–392. doi: [10.1007/s00239-005-0246-5](http://dx.doi.org/10.1007/s00239-005-0246-5)
- Yokobori S, Päabo S (1995) Transfer RNA editing in land snail mitochondria. Proc Natl Acad Sci USA 92: 10432–10435. doi:[10.1073/pnas.92.22.10432](http://dx.doi.org/10.1073/pnas.92.22.10432)
- Zhang DX, Szymura JM, Hewitt GM (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. J Mol Evol 40:382–391. doi:[10.1007/BF00164024](http://dx.doi.org/10.1007/BF00164024)