

Complete Mitochondrial Genome of *Phoneutria Boliviensis* (Araneae, Ctenidae): New Insights Into the Phylogeny and Evolution of Spider

Carlos Fernando Prada (✉ cfprada@ut.edu.co)

Universidad del Tolima

Lida Marcela Franco

Universidad de Ibagué

Felipe Cabarcas

Centro Nacional de Secuenciación Genómica (CNSG), Universidad de Antioquia

Research Article

Keywords: Comparative genomics, Mitochondrial genomes, Gene order analyses, Spider evolution

Posted Date: September 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-936263/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Spiders are the most abundant land predators and megadiverse on earth. In recent years, the mitochondrial genome has been sequenced, mainly for ecological and commercial purposes, reporting some level of rearrangements in this genome. However, there is poor genetic information in several taxonomic families of spiders. The aim of this study was to obtain the sequence of the complete genome of *Phoneutria boliviensis* and, based on this, extract the mitogenomes of other species of the family Ctenidae from published transcriptomes to perform a comparative study among spider species to determine the relationship between the level of mitochondrial rearrangement and its possible relationship with molecular variability in spiders. Complete mitochondrial genomes of eighteen spiders (including nine Ctenidae species) were obtained by two different methodologies (sequencing and transcriptome extraction). Fifty-eight spider mitochondrial genomes were downloaded from the NCBI database for gene order analysis. After verifying the annotation of each mitochondrial gene, a phylogeny and gene order, analysis from 76 spider mitochondrial genomes was obtained. Our results show a high rate of annotation error in the mitochondrial genomes of spiders published in databases, which could lead to false phylogenetic relationships. Moreover, to provide new mitochondrial genomes in spiders by two different methodologies to obtain them, our analysis identifies six different mitochondrial architectures among all spiders. Translocation or tandem duplication random loss (TDRL) events in tRNA genes were identified to explain the evolution of the spider mitochondrial genome. In addition, our findings provide new insights into spider mitochondrial evolution.

Introduction

Spiders (Araneae) belong to the ancient class Arachnida, and predatory arthropods that have inhabited Earth for approximately 400 million years are among the largest animal groups in the world (approximately 49,700 described species) due to their broad diversity, worldwide distribution and conspicuous synapomorphies^{1–3}. The order Araneae, with 129 families, is classified into two infraorders, Mesothelae (primitive spiders) and Opisthothelae (modern spiders). The infraorder Opisthothelae is classified into two suborders Mygalomorphae and Araneomorphae. The family Ctenidae belongs to the suborder Araneomorphae of infraorder Opisthothelae, with 49 genera and 533 species^{2,4}. Despite their importance, the phylogenetic relationships of species belonging to the family Ctenidae remain under study. Spiders of the genus *Phoneutria*, with eight described species known as “Banana spiders”, are restricted to Central America and South America. These spiders are essentially wandering, nocturnal spiders and represent one of the main groups of medically important spiders in South America because of their defensive behavior, synanthropic habits, potent venom, and wide range of prey, including both invertebrates and vertebrates^{5,6}.

The animal mitochondrial genome (mitogenome) has been widely used for the reconstruction of phylogenetic relationships at several taxonomic levels due to its rich signals from sequence information and gene arrangement, from populations to phyla, and has been widely used for the resolution of

taxonomic controversies⁷⁻⁹. The arthropod mitochondrial genome is a compact circular molecule of usually 14–19 kb in size, with 37 genes (13 protein-coding genes, 22 tRNA genes, two rRNA genes), and a noncoding D-loop region. The arrangement of the genes within the arthropod mitogenome is also highly conserved; however, there are some invertebrate lineages with radically rearranged mitochondrial gene order, as in insects^{7,10}.

Despite the rapidly increasing mitochondrial genome sequences from diverse invertebrate species, spider mitochondrial genomes are poorly studied in all arthropods. To date, 58 mitochondrial genomes of 23 spider families have been accessible in the GenBank database (<https://www.ncbi.nlm.nih.gov/>), although genome rearrangements are rare in the majority of spider species^{4,11-15}. Due to the small number of spider mitogenomes sequenced, it is not clear both the phylogenetic relationships and the types of rearrangements in this genome.

In the present study, we sequenced the complete mitochondrial genome of *Phoneutria boliviensis* using next-generation sequencing and other spider mitogenomes from transcriptome extraction. These nine mitochondrial genomes were the first sequenced in the family Ctenidae and nine other spider mitochondrial genomes of different taxonomic families. Therefore, this study aims to compare these eighteen *de novo* mitochondrial genomes with 58 spider mitochondrial genomes to observe their phylogenetic relationships, gene features, gene arrangements, and D-loop region variation that occurred during spider evolution.

Materials And Methods

Sample collection and DNA extraction.

A female adult specimen of *Phoneutria boliviensis* was collected in September 2019 from Ibaguè (4°32'22.3"N, 75°05'37.1"W), Tolima, Colombia. The morphological identification of this specimen was performed by published taxonomic keys¹⁶ and stored in 96% ethyl alcohol at –80°C in the Molecular Biology Laboratory of the University of Ibaguè (Ibaguè, Colombia). The sample was homogenized, and DNA extraction of tissues was performed using the Qiagen DNEasy Tissue kit (Qiagen, Valencia, CA) under the manufacturer's conditions. Genomic DNA concentration was quantified by a dsDNA high-sensitivity kit (Thermo Fisher Scientific, MA, USA) using a Qubit fluorometer.

Mitochondrial genome sequencing and assembly of *Phoneutria boliviensis*

The AIM (Advanced Identification Methods GmbH) (<http://www.aimethods-lab.com/>) carried out sequencing and assembly following standard protocols. The whole genome library of genomic DNA was sequenced using the Illumina HiSeq 2500 platform (2×150 base paired-end reads) (Illumina, USA), which yielded ~1 million reads. The Illumina DNA Flex kit was used for the construction of the paired-end library with standard protocols. The trimming and filtering of the raw sequencing reads were performed by using NGS-Toolkit¹⁷ to remove adapter contamination and low-quality reads with a cutoff of Phred quality

scores of Q20, following previous protocols⁴. High-quality reads were assembled with Geneious 10¹⁸ using the *Pirata subpiraticus* mitochondrial genome (NC_025523) as a reference.

Spider mitochondrial extraction from transcriptome

Forty Araneae transcriptomes, classified by family, were downloaded from NCBI's SRA database (<https://www.ncbi.nlm.nih.gov/sra/?term=>). Homologous contigs in each transcriptome sequence were extracted using the Trinity assembly program from a reference mitogenome. For this analysis, the mitogenome of *Phoneutria boliviensis* (obtained in this work) from the transcriptomes of the Ctenidae family, *Loxosceles similis* (NC_042902) from the transcriptomes of the Leptonetidae family, *Telamonia vlijmi* (NC_024287) from the transcriptomes of the Salticidae and Desidae families, *Cyriopagopus hainanus* (MN877932) from the transcriptomes of the Theraphosidae family and *Pardosa laura* (NC_025223) from the transcriptomes of the Lycosidae family were used as a reference to extract each mitogenome. Mitogenome assembly was performed with Bowtie2 mapper using the Geneious Prime platform. Only mitogenomes obtained with a total coverage of 70% and 13 protein-coding genes (PCGs) coverage of 95% or more were taken into account for the analysis. The data on the transcriptomes used and the process of mitogenome extraction are summarized in Supplementary Table S1.

Mitochondrial genome Annotation and gene rearrangement confirmation

The annotation of PCGs and rRNAs for *P. boliviensis* and 17 other Araneae mitogenomes (obtained by transcriptome extraction) were initially identified via the MITOS web server¹⁹. The annotation of the tRNA genes was performed using tRNAscan-SE 1.21²⁰. The circular image of nine Ctenidae mitochondrial genomes was drawn by using the online server CGView²¹ (http://stothard.afns.ualberta.ca/cgview_server/). We retrieved the sequences and gene annotations of the 58 complete spider mitochondrial genomes that are available at the organelle genome resources database from NCBI (<https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/>) as of January 20, 2021.

Gene annotation verification analysis was based on three previously postulated methodologies²²: a) All gene annotations were refined using other spider mitochondrial sequences from GenBank using the Muscle program²³ with Geneious 10¹⁸ to confirm the locations and boundaries of each gene. Nucleotide identity greater than 80% was used to confirm the orientation of each gene with possible rearrangement. b) The NCBI-BLAST2 sequence comparison program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to confirm rearrangements in the spider mitochondrial genome by comparing nucleotide sequences of orthologous regions from evolutionarily close species. c) Putative tRNA inversions were verified by tRNAscan-SE 1.21²⁰. The orientation of tRNAs will be taken as positive when it coincides in two of the three methodologies used.

After confirming each genetic annotation, numerical gene order was made (from 1 to 38, considering 1 as the *trnM* and 2 as *nad2* gene; and so on) and their gene orientation (plus/plus as + plus/minus as -). A

list of spider species, sorted taxonomically and providing the GenBank IDs, reported gene rearrangements (color-coded to the results of these analyses), of each mitochondrial gene and genome, as annotated in these GenBank records, is summarized in Supplementary Table S2.

The D-loop regions were extracted, and their bp size and CG and AT contents were identified using Geneious Prime ¹⁸. To calculate the skewness, we used the formula $AT\ skew = (A - T) / (A + T)$ and $GC\ skew = (G - C) / (G + C)$ ²⁴. Multiple alignment was performed by the Clustal Omega program ²⁵. Inverted repeats or palindromes in the D-loop region were checked using Tandem Repeats Finder (<http://tandem.bu.edu/trf/trf.html>).

Phylogenetic analysis

A total of 58 Araneae mitogenomes available in GenBank and 18 obtained in this work were included in the phylogenetic analysis. The complete scorpion mitochondrial genome *Buthus occitanus* (NC_010765) was selected as an out-group (Table 1). The 13 PCGs of all 75 spiders and the out-group were concatenated (the gene order in plus/plus orientation is as follows: *atp6*, *atp 8*, *cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4 L*, *nad5* and *nad6*) and subsequently aligned with the MAFFT program ²⁶. PartitionFinder version 2.1.1 ²⁷ with the greedy algorithm was used to find the best substitution models for maximum likelihood (ML). ML analysis was made using IQ-tree ²⁸. The phylogenetic tree was visualized and edited using FigTree v1. 4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Table 1

Features of the eighteen spider mitochondrial genomes obtained in this work. *The total size in bp was calculated based on the reference genome and the missing gaps in the sequence were filled with N

Family	Species	Size (bp)*	%CG	Annotated genes (%CG)			
				PCGs	tRNA	rRNA	D-loop
Theraphosidae	<i>Aphonopelma hentzi</i>	12,991	31.3	13 (31.5)	22 (30.6)	2 (30.5)	0
Leptonetidae	<i>Appaleptoneta fiskei</i>	13,945	33.4	13 (34.9)	22 (30.1)	2 (27.1)	1 (30.2)
	<i>Calileptoneta cokendolpheri</i>	14,022	34.7	13 (36.3)	22 (28.1)	2 (29.5)	1 (34.5)
	<i>Ozarkia georgia</i>	13,654	30.3	13 (30.4)	22 (26.9)	2 (29.2)	0
Salticidae	<i>Bavia aericeps</i>	14,270	22.7	13 (23.4)	22 (22.4)	2 (19.1)	1 (18.6)
	<i>Maripanthus draconis</i>	14,371	24.1	13 (25.2)	22 (24.4)	2 (19.5)	1 (17.2)
	<i>Padillothorax badut</i>	14,058	23.0	13 (23.6)	22 (21.5)	2 (20.0)	1 (21.1)
	<i>Phidippus johnsoni</i>	14,749	23.6	13 (24.2)	22 (22.7)	2 (20.0)	1 (19.9)
	<i>Piranthus planolancis</i>	14,592	24.2	13 (25.2)	22 (22.4)	2 (18.7)	1 (21.2)
Ctenidae	<i>Anahita punctulata</i>	14,022	25.1	12 (25.9)	10 (21.8)	2 (21.6)	1 (24.4)
	<i>Ctenus hibernalis</i>	13,035	25.9	12 (26.6)	9 (27.1)	2 (21.8)	0
	<i>Ctenus captiosus</i>	13,110	24.9	13 (24.9)	14 (28.9)	2 (22.3)	0
	<i>Ctenus exlineae</i>	13,505	26.8	12 (27.6)	10 (30.4)	2 (21.5)	0
	<i>Cupiennius salei</i>	13,826	21.6	12 (22.3)	17 (21.8)	2 (18.9)	0
	<i>Isoctenus ordinario</i>	13,434	24.5	13 (25.1)	14 (24.6)	2 (20.3)	0
	<i>Leptoctenus byrrhus</i>	14,192	27.7	13 (29.3)	20 (24.8)	2 (21.0)	1 (25.7)

Family	Species	Size (bp)*	%CG	Annotated genes (%CG)			
				PCGs	tRNA	rRNA	D-loop
	<i>Phoneutria boliviensis</i>	14,724	28.4	13 (29.7)	22 (25.6)	2 (22.6)	1 (28.3)
	<i>Phoneutria nigriventer</i>	14,186	26.3	13 (27.5)	19 (24.5)	2 (21.3)	1 (24.8)

Gene arrangement analysis

Common interval analysis was performed using CREx²⁹ for pairwise comparisons for inference of ancestral genome states. Pairwise comparisons were conducted for each spider mitochondrial genome against the scorpion mitochondrial genome *Buthus occitanus* to determine the minimum number of genome rearrangement events separating each genic order from the ancestral state.

Results And Discussion

Mitogenome features

The complete mitogenome sequences of nine Ctenidae species range from 13,035 bp in *Ctenus hibernalis* to 14,724 bp in *Phoneutria boliviensis*. This variation in the size of base pairs of these mitogenomes was due to the process of extraction and assembly from transcriptomes available in databases performed in this work. For example, in Ctenidae species, our results show the nucleotide sequence absence of several tRNAs, identifying between 9 and 20 genes (out of a total of 22 tRNAs). In addition, the absence of the sequence corresponding to the *nad4 L* gene (in *Ctenus hibernalis*, *Ctenus exlineae* and *Cupiennius salei*) and the *nad3* gene (*Anahita punctulata*) was also detected.

According to Forni et al.³⁰, although the procedure to reconstruct complete mitochondrial genomes from RNA-Seq experiments is efficient, certain mitochondrial regions, such as tRNAs, D-loops and a fragment of the *atp6* gene, have low read representatively and therefore often generate small gaps during the assembly process. In this same work, the authors indicate that the percentage of reads reached during the process of extraction and assembly of mitogenomes from transcriptomes varies from 16 to 57% using both intrafamilial and congeneric starting references, filtering for possible contaminants and nuclear gene matches³⁰.

Previous studies showed that, in addition to the low coverage (4x) in the aforementioned mitochondrial regions, other characteristics of the RNA-Seq experiments (such as depth of coverage or platform type) or the use of reference mitogenomes (closest evolutionary) are determining factors in obtaining the most complete mitogenome possible^{30,31}. These characteristics would explain why of the 51 transcriptomes analyzed in this work, only 18 mitogenomes satisfied our inclusion criteria; in some of them, small gaps

in the previously mentioned mitochondrial regions were found. Because of this, the mitochondrial genomes of *Ctenus hibernalis* and *Cupiennius salei* were not taken into account in the phylogenetic analysis, as they did not have more than 80% of the coding region. The map of the nine mitochondrial genomes of Ctenidae family species is summarized in Figure 1. Therefore, only the genetic order of *P. boliviensis* and *Leptoctenus byrrhus* were considered for CREX analysis.

The other nine mitogenomes of Araneae species belonging to three different taxonomic families were obtained in this work, ranging from 12,991 bp (*Aphonopelma hentzi*, Family Theraphosidae) to 14,749 bp (*Phidippus johnsoni*, Family Salticidae) in size. The difference in size in base pairs between these mitogenomes is mainly due to the absence of the D-loop region sequence not only in those obtained in this work but also in those deposited in the databases. Despite this, the relatively small mitogenome sizes (<15 kb) are well within the observed range of other Araneae mitogenomes available from the GenBank database^{4,11,13-15} compared to those of other Arthropod species⁷. In addition, truncated tRNA genes may explain why spider mitogenomes are smaller than those of other arthropods³²

The GC content of nucleotides was between 21.6% (*Cupiennius salei*, family Ctenidae) and 34.7% (*Calileptoneta cokendolpheri*, family Leptonetidae). The highest GC content was observed in the PCG (22.3-36.3%), followed by tRNAs (21.8-30.6%), rRNAs (18.9-30.5%), and D-loops (17.2-34.5%). All eighteen Araneae mitochondrial genomes presented the structure of 37 genes (13 PCGs, 22 tRNAs and two rRNAs) and an A+T-rich D-loop region, where a total of 573 genes were annotated. However, some of these species did not have this complete gene set. The features of the eighteen spider mitochondrial genomes obtained in this work are summarized in Table 1 and Supplementary Table S2.

As in most Araneae mitogenomes, the nucleotide composition shows a low G+C, with contents of less than 35% and small variations between species^{13,14}, similar to those observed in Metazoan mitogenomes³³. In addition, the majority strand contains 22 genes and a minority contains 15 genes, similar features to those presented in other spider mitogenomes^{4,14}.

Variation in the D-loop region in spiders

The D-loop region was detected and annotated in 11 of the 18 mitochondrial genomes obtained in this work. Similarly, this region was observed in 66 of the 76 spiders, located between *trnM* and *trnQ* in all Opisthothelae spiders and between *rrnS* and *trnI* in Mesothelae spiders (Supplementary Table S2). The position of the D-loop region is consistent with that observed in previous work in several spider species^{4,14}. In addition, previous studies indicate that the D-loop region remained totally or partially assembled, mainly because it is composed of tandem repeats, which are very difficult to sequence and de novo assemble properly; likewise, their expression in RNA-Seq experiments is low or completely lacking³⁰. Moreover, we obtained complete or partial D-loop regions in 61% of them when mitochondrial genomes of congeneric species were used as references.

Our results show that the D-loop of the 66 spider region is characterized by an average size of 766.2 bp, an average AT % of 75.1% and a low nucleotide identity of 18.7%, showing high variation in this species.

For example, the two spiders of the suborder Mesothelae (family Liphistiidae) and the five species of the suborder Mygalomorphae (families Nemesiidae (1), Theraphosidae (3), Dipluridae (1)) have a D-loop region with an average of 388.5 bp, while the species of the suborder Araneomorphae have a D-loop region with an average of 800.3 bp. The largest D-loop region was observed in *Argyroneta aquatica* (family Cybaeidae) at 2047 bp, where both 5' and 3' AT tandem repeat expansions were detected. Likewise, the AT content was variable in all species, ranging from 66.3–82.6%. However, most of the D-loop regions belonging to the Liphistiidae, Theraphosidae, Dipluridae, Dysderidae, Hypochilidae, Sicariidae, Pholcidae, Oecobiidae, Araneidae, Nephilidae, Eutichuridae and Selenopidae families showed positive AT/GC skews, indicating an obvious bias toward the use of A and G, similar to those reported in *Lyroglyphus crotalus* of the Theraphosidae family⁴ and other spiders³². In contrast, our analyses show that mainly Entelegynae spiders have a negative AT/GC skew, indicating a bias toward the use of T and G.

Regarding the presence of repetitive regions, our results show three categories of D-loop regions: a) sequences with a size in pb of ≤ 500 bp (n= 14) that do not have tandem repeat elements; b) D-loop regions between 500 and 740 bp (n= 25), that do not have tandem repeats or that are of low complexity (AT-rich regions with less than 20 nucleotides and less than 2 copies); and c) D-loop regions above 740 bp (n= 27), that have tandem repeats of higher complexity (AT-rich regions with more than 20 nucleotides and more than 3 copies). The molecular characteristics of the D-loop region in the spider species analyzed are summarized in Supplementary Table S3.

Although these molecular features (AT/GC skew and presence of tandem repeat elements could be discriminant between Haplogynae and Entelegynae spiders), it is necessary to analyze a larger number of complete D-loop regions to confirm this postulate.

Gene annotation errors and rearrangement level in the Araneae mitochondrial genome

According to the annotation of the 58 Araneae mitogenome mitochondrial genomes found in the NCBI database, there were a total of 203 cases of 2,166 genes annotated to be in arrangements differing from the ancestral gene order of the scorpion mitochondrial genome *Buthus occitanus*. Of all these reorganizations, 17 were identified as gene annotation errors, which means an error rate of 8.4% (17/203), indicating that 21% (12/57) of mitochondrial genomes have at least one gene badly annotated (Supplementary Table S2). For example, the mitochondrial genome with the highest number of annotation errors (5) was that of *Amaurobius fenestralis* (family Amaurobiidae), in which the order *trnS*, *trnR*, *trnE*, *-trnL* and *-trnF*, without annotation of *trnA* and *trnN* genes, was reported. Our analysis indicates that the *-trnL2*, *trnN*, *trnA*, *trnS1*, *trnR*, *trnE* and *-trnF* genes are found in this region, confirming the annotation error. In addition to false inversions, our analyses indicate a high number (40) of unannotated genes, but our results confirm their presence position in the genome. Most of these unannotated genes were tRNAs (21), followed by the D-loop region (19) (Supplementary Table S2). Examples of false inversions and deletions in the spider mitochondrial genome are shown in Supplementary Figure S1.

Previously, it has been reported that thousands of mitogenomes deposited in the database have shown a high rate of annotation errors, most of which are easily detectable^{34,35}. Earlier studies indicate that such errors are frequent in the Hexapoda mitogenome, with an annotation error rate of 5.5% concentrated mainly in tRNA genes³⁶. Although our results confirm that annotation errors in the spider mitogenome are concentrated in tRNAs, the error rate is almost double that observed in Hexapoda, generating false gene orders in this taxonomic group.

Phylogenetic analysis

A phylogenetic tree from 74 Araneae species and an outgroup species based on the nucleotide sequence of 13 mitochondrial protein-coding genes using maximum likelihood (ML) was obtained (Figure 2). Our findings provide strong support for the monophyly of two suborders (Opisthothelae and Mesothelae) in Araneae with high supporting values, demonstrating that the two species of the Liphistiidae family (Mesothelae) are ancestral sequences with respect to the Opisthothelae species. In addition, the subdivision of Opisthothelae into two monophyletic groups, Mygalomorphae and Araneomorphae infraorders, was supported in the phylogenetic tree; however, Araneomorphae was recovered paraphyletically in this analysis. The estimated tree in this study also supports the monophyly of Haplogynae and Entelegynae taxa. Likewise, species of the Lycosidae, Oxyopidae and Ctenidae families formed a phylogenetically closely related clade (Figure 2). Although most of the data support a classical taxonomic classification, our phylogenetic analysis separates two species of the Pisauridae family (*Dolomedes angustivirgatus* and *Pisaura mirabilis*) into distinct monophyletic groups, showing that the phylogenetic position of the family Pisauridae is unstable and needs further study. The estimated ML tree from the mitochondrial genome is consistent with those observed in previous analyses of 13 PCG mitochondrial genes^{4,32} and multilocus phylogeny³⁷. However, more mitochondrial genome data of Mesothelae, Mygalomorphae and haplogynae taxa would cover the consistent sign of deep phylogenetic relationship between all Araneae species.

Spider mitochondrial genome architectures and CREX analysis

By comparing the mitochondrial gene order rearrangement scenario of the Araneae order with the putative ancestral scorpions (*Buthus occitanus*) inferred by CREX analysis, three tRNAs (*trnY*, *trnL2*, and *trnI*) and a D-loop region were found to translocate and one TDRL event in block *trnN-trnA-trnS1-trnR* (Figure 2).

Compared to the putative ancestral scorpion architecture of *Buthus occitanus* (A), spiders belonging to the infraorders Mesothelae (primitive spiders) present only one translocation of the *trnY* gene in the *trnW-trnY-trnC* region (architecture B). Mygalomorphae and Haplogynae spiders differ from Mesothelae by a tandem duplication and random loss (TDRL) event in the *trnN-trnA-trnS1-trnR* region, a *trnL2* gene translocation event in the same region, a *trnT* gene translocation event in the *cob-trnS2* region and a translocation of the D-loop region (architecture C). Only the three species of the Dysderidae family vary in Mesothelae spiders by translocation of the *trnI* gene in the *trnM-nad2* region (architecture D). In addition, all Entelegynae differ from most Haplogynae spiders (architecture C) by translocation of the *trnI* gene in

the *nad6-cob* region (architecture E and F). However, two of the three species of the Tetragnathidae family have a different architecture. *Tetragnatha maxillosa* varies in architecture F by translocation of the *trnG* gene in the *trnQ-D loop* region (architecture E1). In addition, *Tetragnatha nitens* (architecture E2) has a new translocation of the *trnW* gene in the *trnQ-Dloop* region, evidencing a generic order *trnQ-trnW-trnG-Dloop*. Our analysis indicates that *Leucauge celebesiana* has the gene order of architecture F (Supplementary table S2 and Figure 2).

Mitogenomes have been recently applied, particularly to studies regarding the phylogeny and evolution of spiders^{14,38}. The gene rearrangements and mechanisms inducing the different gene orders or architectures of the spider mitogenomes are available in public databases, displaying an accelerated gene rearrangement in Araneomorphae compared to Mygalomorphae^{4,13,32,38}. However, despite the low number of species analyzed, these gene rearrangements are overrepresented due to the large number of annotation errors, generating up to nineteen mitochondrial gene rearrangement types and 12 architectures among 44 spiders³⁸, while in our analysis of 76 spider species, nine mitochondrial gene rearrangement types (8 translocation and one TDRL event) and six architectures were confirmed (see Supplementary Table S2 and Figure 2).

Conclusion

The complete mitochondrial genome of *P. boliviensis* was characterized, allowing extraction from transcriptomes of eight other mitogenomes of the family Ctenidae, in addition to nine other species of different families and comparison with other members of the Araneae. The analysis of the spider mitochondrial genomes available in the database shows that a significant rate of gene annotation error in this taxonomic group significantly affected the phylogenetic relationships in terms of genetic order. In addition to expanding the number of spider mitogenomes available, our analyses report a phylogenetic position of the families Ctenidae and Leptonetidae, which provides a potentially more robust phylogeny and systematics of Araneae. Most importantly, the findings from this study provide new evidence of a low rate of rearrangement among spider mitochondrial genomes, with divergences observed at the phylogenetic and genetic levels between Mesothelae and Opisthothelae spiders.

Declarations

Funding

We thank the University of Tolima for supporting this research. The first author thanks Oficina de Investigaciones y Desarrollo Científico de la Universidad del Tolima for postdoctoral fellowships (4/2019). We are grateful to the Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias-currently Minciencias, Colombia) for financial support for this project (ID Project 130780864623).

Authors' Contributions

Carlos Fernando Prada Quiroga: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing Visualization, Supervision, Project administration and Funding acquisition

Lida Marcela Franco Pérez: Methodology, Validation, Writing - Original Draft, Writing - Review & Editing

Felipe Cabarcas Jaramillo: Methodology, Validation, Writing - Original Draft, Writing - Review & Editing

Conflict of interest

The authors have no conflict of interest.

References

1. Coddington, J. A. & Levi, H. W. Systematics and evolution of spiders (Araneae). *Annual review of ecology and systematics*, **22**, 565–592 (1991).
2. Catalog, W. S. World Spider Catalog. Version 22.5. *Natural History Museum Bern*, <https://doi.org/10.24436/2> (2021).
3. Lüddecke, T., Herzig, V., Vilcinskis, A. & von Reumont, B. M. The biology and evolution of spider venoms. *Biological reviews of the Cambridge Philosophical Society*, <https://doi.org/10.1111/brv.12793> (2021).
4. Kumar, V. *et al.* The Complete Mitochondrial Genome of endemic giant tarantula, *Lyrognaethus crotalus* (Araneae: Theraphosidae) and comparative analysis. *Scientific reports*, **10**, 74 <https://doi.org/10.1038/s41598-019-57065-8> (2020).
5. Hazzi, N. A. Natural history of *Phoneutria boliviensis* (Araneae: Ctenidae): habitats, reproductive behavior, postembryonic development and prey-wrapping. *The Journal of Arachnology*, **42**, 303–310 <https://doi.org/10.1111/j.1469-185X.1982.tb00363.x> (2014).
6. Sierra Ramírez, D. *et al.* Deciphering the diet of a wandering spider (*Phoneutria boliviensis*; Araneae: Ctenidae) by DNA metabarcoding of gut contents. *Ecology and evolution*, **11**, 5950–5965 <https://doi.org/10.1002/ece3.7320> (2021).
7. Cameron, S. L. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual review of entomology*, **59**, 95–117 <https://doi.org/10.1146/annurev-ento-011613-162007> (2014).
8. Gissi, C., Iannelli, F. & Pesole, G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species., **101**, 301–320 <https://doi.org/10.1038/hdy.2008.62> (2008).
9. Li, Y. T. *et al.* Comparative Mitochondrial Genome Analyses of Sesamid and Other Brachyuran Crabs Reveal Gene Rearrangements and Phylogeny. *Frontiers in genetics*, **11**, 536640 <https://doi.org/10.3389/fgene.2020.536640> (2020).
10. Boore, J. L. Animal mitochondrial genomes. *Nucleic acids research*, **27**, 1767–1780 (1999).
11. Fang, W. Y., Wang, Z. L., Li, C., Yang, X. Q. & Yu, X. P. The complete mitogenome of a jumping spider *Carrhotus xanthogramma* (Araneae: Salticidae) and comparative analysis in four salticid

- mitogenomes., **144**, 699–709 <https://doi.org/10.1007/s10709-016-9936-8> (2016).
12. Masta, S. E. & Boore, J. L. Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. *Molecular biology and evolution*, **25**, 949–959 <https://doi.org/10.1093/molbev/msn051> (2008).
 13. Wang, Z. L., Li, C., Fang, W. Y. & Yu, X. P. The Complete Mitochondrial Genome of two Tetragnatha Spiders (Araneae: Tetragnathidae): Severe Truncation of tRNAs and Novel Gene Rearrangements in Araneae. *International journal of biological sciences*, **12**, 109–119 <https://doi.org/10.7150/ijbs.12358> (2016).
 14. Yong, H. S. *et al.* Complete mitochondrial genomes and phylogenetic relationships of the genera Nephila and Trichonephila (Araneae, Araneoidea). *Scientific reports*, **11**, 10680 <https://doi.org/10.1038/s41598-021-90162-1> (2021).
 15. Zhu, X. L. & Zhang, Z. S. The complete mitochondrial genome of Agelena silvatica (Araneae: Agelenidae). *Mitochondrial DNA. Part B, Resources*, **2**, 58–59 <https://doi.org/10.1080/23802359.2017.1280702> (2017).
 16. Simó, M. & Brescovit, A. D. Revision and cladistic analysis of the Neotropical spider genus Phoneutria Perty, 1833 (Araneae, Ctenidae), with notes on related Cteninae. *Bulletin-british arachnological society*, **12**, 67–82 (2001).
 17. Patel, R. K. & Jain, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS one*, **7**, e30619 <https://doi.org/10.1371/journal.pone.0030619> (2012).
 18. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, **28**, 1647–1649 <https://doi.org/10.1093/bioinformatics/bts199> (2012).
 19. Bernt, M. *et al.* MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular phylogenetics and evolution*, **69**, 313–319 <https://doi.org/10.1016/j.ympev.2012.08.023> (2013).
 20. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research*, **25**, 955–964 <https://doi.org/10.1093/nar/25.5.955> (1997).
 21. Grant, J. R. & Stothard, P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic acids research*, **36**, W181–184 <https://doi.org/10.1093/nar/gkn179> (2008).
 22. Prada, C. F. & Boore, J. L. Gene annotation errors are common in the mammalian mitochondrial genomes database. *BMC genomics*, **20**, 73 <https://doi.org/10.1186/s12864-019-5447-1> (2019).
 23. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792–1797 <https://doi.org/10.1093/nar/gkh340> (2004).
 24. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of molecular evolution*, **41**, 353–358 <https://doi.org/10.1007/bf00186547> (1995).
 25. Sievers, F. & Higgins, D. G. The Clustal Omega Multiple Alignment Package. *Methods in molecular biology. (Clifton, N.J.)*, **2231**, 3–16 https://doi.org/10.1007/978-1-0716-1036-7_1 (2021).

26. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**, 772–780 <https://doi.org/10.1093/molbev/mst010> (2013).
27. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular biology and evolution*, **34**, 772–773 <https://doi.org/10.1093/molbev/msw260> (2017).
28. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, **32**, 268–274 <https://doi.org/10.1093/molbev/msu300> (2015).
29. Bernt, M. *et al.* CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics (Oxford, England)*, **23**, 2957–2958 <https://doi.org/10.1093/bioinformatics/btm468> (2007).
30. Forni, G. *et al.* Complete mitochondrial genomes from transcriptomes: assessing pros and cons of data mining for assembling new mitogenomes. *Scientific reports*, **9**, 14806 <https://doi.org/10.1038/s41598-019-51313-7> (2019).
31. Moreira, D. A., Furtado, C. & Parente, T. E. The use of transcriptomic next-generation sequencing data to assemble mitochondrial genomes of *Ancistrus* spp. (Loricariidae)., **573**, 171–175 <https://doi.org/10.1016/j.gene.2015.08.059> (2015).
32. Pons, J., Bover, P., Bidegaray-Batista, L. & Arnedo, M. A. Arm-less mitochondrial tRNAs conserved for over 30 millions of years in spiders. *BMC genomics*, **20**, 665 <https://doi.org/10.1186/s12864-019-6026-1> (2019).
33. Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Systematic biology*, **54**, 277–298 <https://doi.org/10.1080/10635150590947843> (2005).
34. Boore, J. L. Requirements and standards for organelle genome databases. *Omicron: a journal of integrative biology*, **10**, 119–126 <https://doi.org/10.1089/omi.2006.10.119> (2006).
35. Donath, A. *et al.* Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic acids research*, **47**, 10543–10552 <https://doi.org/10.1093/nar/gkz833> (2019).
36. Moreno-Carmona, M., Cameron, S. L. & Prada Quiroga, C. F. How are the mitochondrial genomes reorganized in Hexapoda? Differential evolution and the first report of convergences within Hexapoda., **791**, 145719 <https://doi.org/10.1016/j.gene.2021.145719> (2021).
37. Wheeler, W. C. *et al.* The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling., **33**, 574–616 <https://doi.org/10.1111/cla.1218> (2017).
38. Tyagi, K. *et al.* The gene arrangement and phylogeny using mitochondrial genomes in spiders (Arachnida: Araneae). *International journal of biological macromolecules*, **146**, 488–496 <https://doi.org/10.1016/j.ijbiomac.2020.01.014> (2020).

Figures

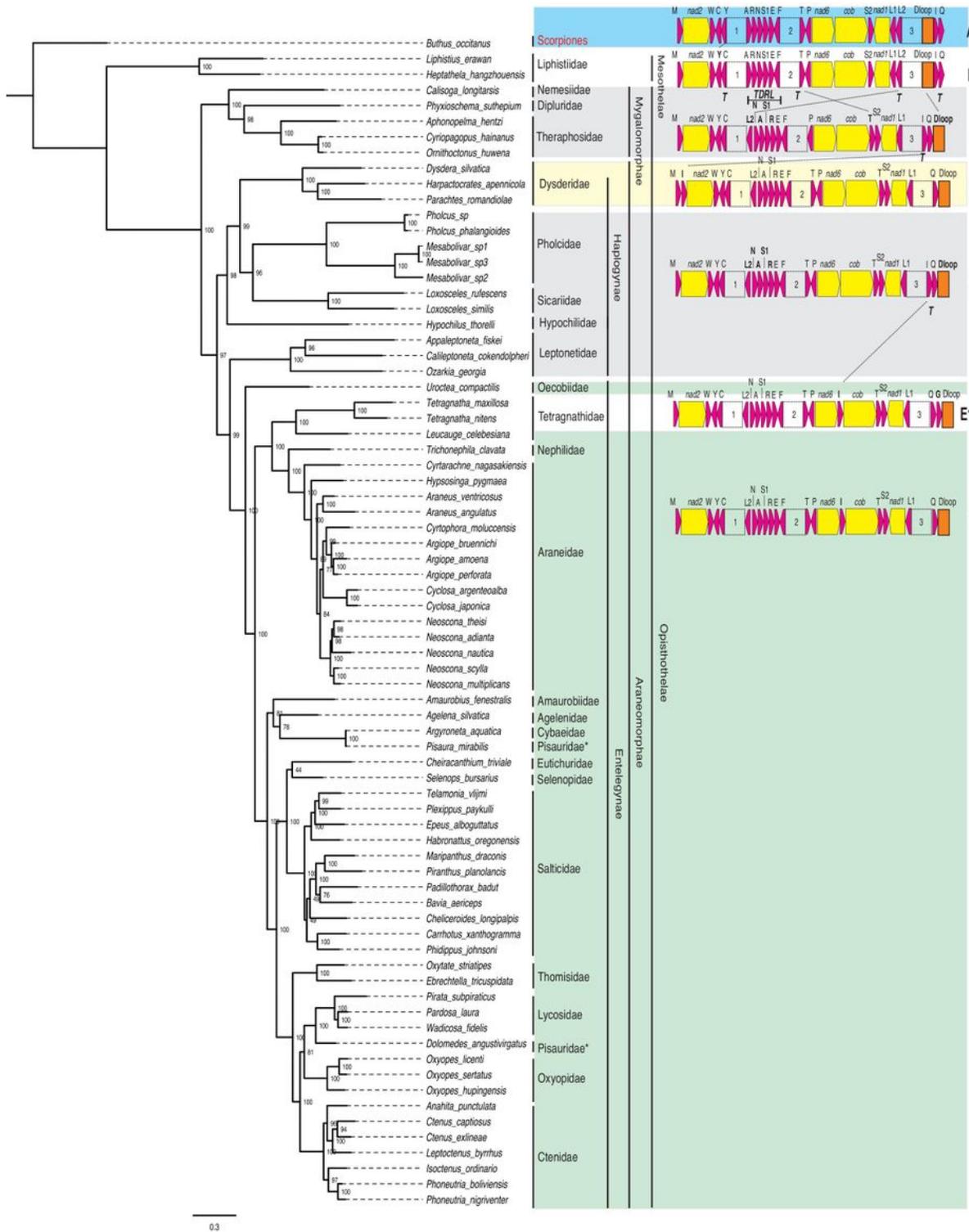


Figure 2

Maximum likelihood phylogenetic tree inferred by 13 PCGs and CREX analyses of the 74 Araneae mitochondrial genomes. The complete scorpion mitochondrial genome *Buthus occitanus* (NC_010765) was selected as the out-group. Numeric values at the nodes are ML posterior probabilities. The phylogenetic tree was visualized and edited using FigTree v1. 4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>). In the CREX analyses (right), PCGs are shown as yellow

arrows, tRNA genes as purple color arrows, and D-loop regions as orange rectangles. The dotted boxes with numbers represent conserved blocks of genes: 1 (cox1-cox2-trnK-trnD-atp8-atp6-cox3-trnG-nad3), 2 (nad5-trnH-nad4-nad4 L) and 3 (rrnL-trnV-rrnS). The letters T are shown as translocation events, and the TDRL is shown as tandem duplication and random loss events.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.pdf](#)
- [SupplementaryMateriallegend.docx](#)
- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS3.xlsx](#)