



OPEN New mitochondrial genomes of three whip spider species from the Amazon (Arachnida, Amblypygi) with phylogenetic relationships and comparative analysis

Acácio Freitas Nogueira[✉], Eder S. Pires, Guilherme Oliveira, Leonardo Carreira Trevelin & Santelmo Vasconcelos[✉]

The complete mitochondrial genomes of the whip spiders *Charinus carajas*, *C. ferreus*, and *Heterophrynus longicornis* were sequenced, annotated, and compared with other mitogenomes of whip spiders and arachnids. The three new mitogenomes have the 37 genes usually observed in Metazoa: 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs), and two ribosomal RNAs (rRNAs), plus a non-coding control region (CR). Most PCGs presented an ATN start codon, except *cox1* in both *Charinus* species, initiating with TTA. Most PCGs terminated with stop codons TAA or TAG, except *nad5* of *C. carajas* and *cox3* of *H. longicornis*, which presented an incomplete stop codon (T). The Ka/Ks ratios were less than one for all the PCGs, indicating these genes are under purifying selection. All the tRNAs, except for serine 1 (*trnS1*), had the typical cloverleaf-shaped secondary structure. All the phylogenetic analyses resolved *Charinus carajas* and *C. ferreus* as monophyletic groups. Nonetheless, we did not recover the monophyly of *Heterophrynus longicornis*. The phylogenies under partitioned models did not recover suprageneric taxonomic groups as clades, but the Bayesian inference under the CAT infinite mixture model recovered the family Phrynidae and the superfamily Phrynoidea as monophyletic groups.

Keywords *Charinus*, *Heterophrynus*, Ka/Ks ratio, Cave species

The arachnid order Amblypygi, the whip spiders, are unique organisms with dorsoventrally compressed bodies, long and raptorial pedipalps, and an extremely elongated antenniform first pair of legs that are not used for locomotion but as specialized sensory structures¹. Despite being a relatively small order with approximately 278 living species², whip spiders are found in tropical and subtropical regions worldwide^{3,4}, with the most diversity hosted in South America⁵. Systematics and phylogenetic relationships in Amblypygi have been investigated mostly using morphologic characters^{6–9}, however, recent studies have also used molecular data to infer phylogenetic relationships in the order^{4,10}. In early cladistics studies and recent molecular phylogenetic analyses, Amblypygi has been resolved as a monophyletic group^{6,10–13}. Another undisputed relationship is Amblypygi as the sister group to Uropygi, a clade formed by the orders Thelyphonida (whip scorpion) and Schizomida (short-tailed whip scorpions), with these three orders forming the clade referred as Pedipalpi^{12–15}.

Molecular data have also been employed in studies of cryptic diversification and phylogeography of whip spiders^{16–20}. In this trend, complete mitochondrial genomes (mitogenomes) have been used to infer phylogenetic relationships in Arachnida^{21,22}. However, complete mitogenomes of only two whip spiders, *Damon diadema* and *Phrynus* sp., have been published so far²¹, which hamper the use of most of these organisms in mitogenomic studies of biodiversity.

Therefore, given the lack of available mitochondrial genomes for other Amblypygi taxa, we sequenced and annotated the complete mitochondrial genomes of three species occurring in the Eastern Amazon, *Charinus carajas*, *C. ferreus* and *Heterophrynus longicornis*. *Charinus* was the type genus of the pantropical family Charinidae⁸. However, recent changes in the classification of Amblypygi based on a global molecular phylogeny using ultraconserved elements (UCEs), as well as the combination of UCEs with six Sanger loci, have synonymized

Instituto Tecnológico Vale, Rua Boaventura da Silva 955, Belém, Pará CEP 66055-090, Brazil. ✉email: acacio.nogueira@pq.itv.org; santelmo.vasconcelos@itv.org

Charinidae with Charontidae. *Charinus* now belongs to Charontidae, the most diverse family in the order, with around 150 species¹⁰. The two sequenced *Charinus* are narrow-range species endemic to ferruginous caves in the Serra dos Carajás⁵, an iron ore mining region, while *H. longicornis* is widely distributed in northern and central South America¹⁶. This study aims to determine the features of these *de novo* assembled mitochondrial genomes, to compare the data of the three species with those of other related arachnids (including other Amblypygi taxa), and to elucidate the phylogenetic relationships of the herein sequenced species with other whip spiders based on mitochondrial sequences.

Furthermore, mtDNA has been one of the most popular markers for the study of molecular biodiversity in animals²³. The reasons for that are specific biological properties, such as the high variability of mitochondrial sequences in natural populations because of its elevated mutation rate²⁴, which can generate some signals about population and species history over short time frames²³. In addition, the clonal inheritance of its whole genome, which behaves as a single non-recombining locus, makes it an outstanding marker for tracing the evolution of genealogies²⁵. These properties combine elevated polymorphism and traceability, indispensable features for classical phylogeography²⁵, species identification²⁶, and modern biodiversity monitoring via eDNA metabarcoding²⁷.

Results and discussion

Overall genomes features

The complete mitochondrial genomes presented lengths of 14,785 bp for *Heterophrynus longicornis*, 14,828 bp for *Charinus carajas*, and 14,824 bp (two samples) or 14,822 bp (one sample) for *Charinus ferreus* (Fig. 1). The pairwise identity among samples of *C. carajas* was 99.8%, and 99.9% among samples of *C. ferreus*. Nucleotide composition was approximately 35% of A, 28.9% of C, 8.9% of G, and 27.2% of T for *C. carajas*; 32.9% of A, 34.4% of C, 9.5% of G, and 23.2% of T for *C. ferreus*; and 32.6% of A, 28% of C, 9.7% of G, and 29.7% of T for *H. longicornis*. All the mitogenomes were positively skewed for the AT proportion (0.125 for *C. carajas*; 0.173 for *C. ferreus*, and 0.047 for *H. longicornis*), and negatively skewed for the GC content (− 0.529 for *C. carajas*; − 0.567 for *C. ferreus* and − 0.485 for *H. longicornis*). These considerably high negative values of GC skew were also observed in the mitogenomes of the two previously sequenced Amblypygi (− 0.473 for *Damon diadema* and − 0.453 for *Phrynus* sp.), being yet negative in the mitogenomes of the whip scorpion *Mastigoproctus giganteus* (− 0.297)²¹ and the Schizomida *Schizomus zhensis* (− 0.393)²⁸. However, the known Araneae mitogenomes reported positively skewed GC content, as in the cases of *Nephila pilipes* (0.240), *Trichonephila antipodiana* (0.271), *T. vitiana* (0.266)²⁹, and *Cheiracanthium triviale* (0.194)³⁰. Considering the phylogenetic relationships of the cited taxa, with Uropygi being the sister group to Amblypygi and both forming Pedipalpi, the clade sister to Araneae, these results may indicate a bias toward the excessive use of Cs over Gs from Araneae to Amblypygi or vice versa. AT skewness presents a less notable pattern because all the amblypygids herein sequenced plus *Damon diadema* (0.101) exhibited positive values, whereas *Phrynus* sp. was the only exception with a value of − 0.03²¹. On the other arachnids, all the Araneae and Thelyphonida species analyzed presented negative values of AT skew^{21,29,30}, but *Schizomus zhensis* presented positive (0.084) AT skew.

The mitogenomes assembled here contained the typical 37 genes of bilaterians, being 13 protein-coding genes, 22 transfer RNAs, and two ribosomal RNAs (*rrnL* and *rrnS*), as well as a putative control region characterized by a non-coding A-T rich segment (Fig. 1; Table 1). The gene arrangement in the three mitogenomes is identical to those of *Damon diadema* and *Phrynus* sp. Such gene order is almost equal to that of Pycnogonida mitogenome (from which differs by a single tRNA translocation event), which was inferred to represent the putative ancestral gene order of Chelicerata based on an analysis of 464 complete arthropod mitochondrial genomes³¹. Twenty-three genes (nine PCGs and 14 tRNAs) are encoded on the majority strand (J-strand) and the other 14 (4 PCGs, 8 tRNAs, and 2 rRNAs) are encoded on the minority strand (N-strand). The mitogenome of *C. carajas* contained 14 genic overlaps, ranging from one to 45 bp, while *C. ferreus* presented 15 overlaps, ranging from one to 20 bp, and *H. longicornis* with nine overlaps, ranging from one to 17 bp. The number of intergenic spacers, excluding the control region, was 13 in *C. carajas* and *C. ferreus*, and 16 in *H. longicornis*. The longest intergenic spacer in all species was between *rrnL* and *trnV*. Although this interval had reached a few over 200 bp in both species of *Charinus*, it was just 35 bp in *Heterophrynus* (Table 1).

Protein-coding genes and codon usage

The complete extension of the coding regions was 10,913 bp for *C. carajas*, 10,897 bp for *C. ferreus*, and 11,012 bp for *H. longicornis*. The total length of each PCG in the mitogenomes of each species, as well as their initial and terminal codons, are listed in Table 1. All the PCGs initiate by the usual start codon ATN (N = A/T/G/C), except for *cox1* of *Charinus*, which initiate with the unusual codon TTA, as also observed for *Phrynus* sp. and other arachnids (namely, *Nephila pilipes*, *Trichonephila antipodiana* and *T. vitiana*)^{21,29}. Nevertheless, *Cheiracanthium triviale*, another Araneomorphae species, has unusual start codons (TTA, TTG and TTT) in other PCGs but not in *cox1*³⁰. Most PCGs terminate with one of the complete stop codons TAA or TAG, except *cox2* of both *Charinus* species, *cox3* of *C. carajas* and *H. longicornis*, *nad3* of *C. carajas*, and *nad5* of *C. carajas* and *H. longicornis*, which terminate with an incomplete stop codon T or TA. Truncated stop codons consisting only of a T or a TA were also observed in *Phrynus* (three genes) and *Damon* (five genes) mitogenomes, and such incomplete stop codons are presumed to be completed by post-transcriptional polyadenylation³².

The average A + T content was 61.0% for *H. longicornis* and *C. carajas*, and 52.5% for *C. ferreus*, ranging from 59.8% (*cox3*) to 64.0% (*nad3*), from 57.3% (*cox1*) to 67.9% (*atp8*), and from 51% (*nad3*) to 57.4% (*nad4l*), respectively. Leucine1 was the most frequently used amino acid (AA) in both *Charinus* mitogenomes, while Isoleucine was the most observed in *H. longicornis*. Cysteine was the least used AA in the three mitogenomes. The RSCU for all AAs in the three species are shown in Supplementary Figure S1.

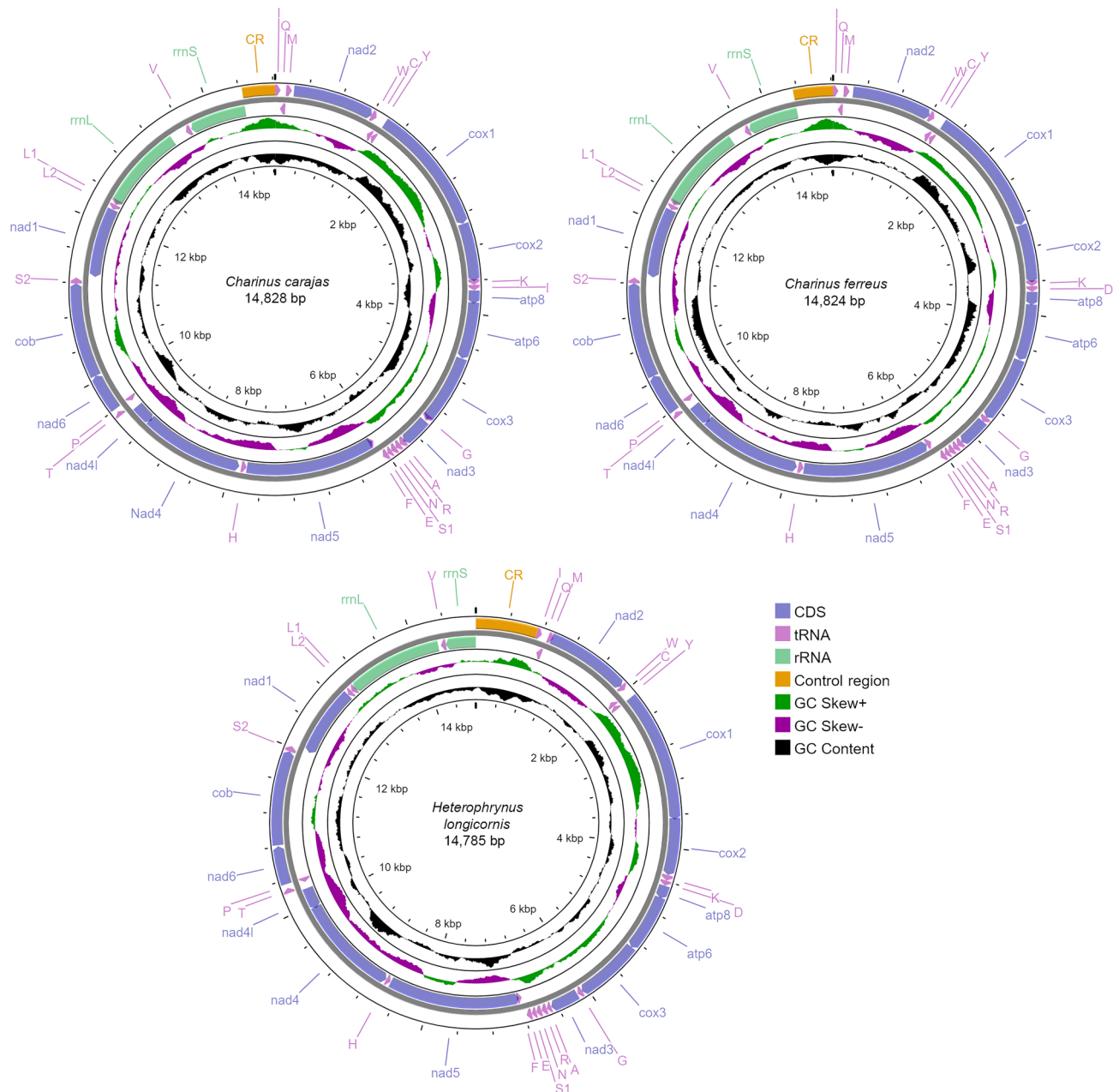


Fig. 1. Circular maps of the complete mitochondrial genomes of *Charinus carajas*, *C. ferreus*, and *Heterophrynus longicornis*, showing the PCGs (CDS), rRNAs, tRNAs, and the control region. The images were generated using the CGView/Proksee online server (<https://proksee.ca/>)

The Ka/Ks ratios were less than one for the 13 PCGs (Fig. 2), indicating that all these genes are under purifying selection. The *cox1* and *atp8* genes had the lowest and the highest Ka/Ks ratios, respectively, as previously reported for spiders^{29,33}. Following an increasing order we found: *cox1* < *cox3* < *cox2* < *cob* < *atp6* < *nad3* < *nad5* < *nad2* < *nad4* < *nad4l* < *nad1* < *nad6* < *atp8*. In addition, the fact that *cox1* presented the lowest Ka/Ks ratio in the mitogenomes of whip spiders, representing fewer changes in amino acids, gives support to its use as a suitable molecular marker for DNA-based species delimitation within the order, as well as reported in other arachnids^{34,35}.

Transfer and ribosomal RNAs

The size of the tRNAs ranged between 56 and 68 bp in *H. longicornis*, 55–70 bp in *C. carajas*, and 55–74 bp in *C. ferreus* (Table 1). The sum of all the sequences of tRNAs had lengths of 1387 bp in *H. longicornis*, 1449 bp in *C. carajas*, and 1,446 bp in *C. ferreus*. In the *Charinus* species, all the transfer RNAs except the tRNA of serine1 (*trnS1*), have the typical cloverleaf-shaped secondary structure. The predicted secondary structures of all tRNAs of the three species are shown in Supplementary Figures S2–S4. The *trnS1* lacks the DHU arm, which

Gene	Strand	Length			gen. ovl./inter. spac.			Start/stop codons		
		Cc	Cf	HI	Cc	Cf	HI	Cc	Cf	HI
<i>trnI(tgt)</i>	+	60	63	63	0	0	0			
<i>trnP(tgg)</i>	−	63	65	62	5	3	− 15			
<i>nad6</i>	+	432	432	459	− 1	− 1	13	ATT/TAA	ATC/TAA	ATA/TAA
<i>cob</i>	+	1131	1137	1128	− 2	− 2	− 2	ATG/TAA	ATG/TAA	ATG/TAG
<i>trnS2(tga)</i>	+	70	66	68	21	19	6			
<i>nad1</i>	−	915	915	915	4	4	2	ATG/TAA	ATG/TAA	ATG/TAG
<i>trnL2(taa)</i>	−	68	66	63	2	3	1			
<i>trnL1(tag)</i>	−	66	65	63	− 45	0	0			
<i>rrnL</i>	−	1118	1083	1226	202	201	35			
<i>trnV(tac)</i>	−	70	67	66	9	1	− 3			
<i>rrnS</i>	−	749	665	395	0	0	0			
CR		389	469	731	0	0	0			
<i>trnI(gat)</i>	+	66	64	65	− 3	− 3	9			
<i>trnQ(ttg)</i>	−	69	66	67	− 1	− 1	3			
<i>trnM(cat)</i>	+	70	70	65	0	0	0			
<i>nad2</i>	+	996	996	984	− 2	− 2	− 2	ATC/TAA	ATA/TAA	ATC/TAA
<i>trnW(tca)</i>	+	67	67	62	− 1	− 1	0			
<i>trnC(gca)</i>	−	62	65	63	1	3	0			
<i>trnY(gta)</i>	−	61	63	60	− 5	− 5	− 8			
<i>cox1</i>	+	1536	1536	1560	3	3	− 17	TTA/TAA	TTA/TAA	ATC/TAA
<i>cox2</i>	+	670	670	681	0	0	1	ATG/T(AA)	ATG/T(AA)	ATG/TAA
<i>trnK(ctt)</i>	+	68	69	68	− 1	− 1	0			
<i>trnD(gtc)</i>	+	66	68	62	0	0	1			
<i>atp8</i>	+	156	156	156	− 7	− 7	− 7	ATC/TAA	ATC/TAA	ATG/TAA
<i>atp6</i>	+	672	672	681	3	5	− 1	ATG/TAA	ATG/TAA	ATG/TAA
<i>cox3</i>	+	785	786	784	− 1	− 1	0	ATG/TA(A)	ATG/TAA	ATG/T(AA)
<i>trnG(tcc)</i>	+	67	68	63	0	0	6			
<i>nad3</i>	+	346	348	342	0	− 2	0	ATT/T(AA)	ATC/TAG	ATT/TAG
<i>trnA(tgc)</i>	+	70	67	56	0	− 1	1			
<i>trnR(tcg)</i>	+	64	63	64	− 1	− 1	2			
<i>trnN(gtt)</i>	+	66	64	64	0	0	0			
<i>trnS1(gct)</i>	+	55	55	56	1	1	1			
<i>trnE(ttc)</i>	+	64	63	62	− 4	− 4	4			
<i>trnF(gaa)</i>	−	67	68	63	0	− 1	0			
<i>nad5</i>	−	1660	1668	1699	4	4	0	ATT/T(AA)	ATT/TAG	ATG/T(AA)
<i>trnH(gtg)</i>	−	70	74	62	26	16	3			
<i>nad4</i>	−	1332	1329	1338	− 20	− 20	− 20	ATG/TAA	ATG/TAA	ATG/TAG
<i>nad4l</i>	−	282	282	285	23	23	21	ATG/TAG	ATG/TAG	ATG/TAG

Table 1. List of annotated genes and their features in the mitochondrial genomes of *Charinus carajas* (C), *Charinus ferreus* (Cf), and *Heterophrynus longicornis* (HI) CR = control region; length in bp; minus sign indicates genic overlaps (gen. ovl); positive numbers indicate the number of bases as intergenic spacers (inter. Spac.). 3' a residues added to the mRNA to complete the stop codons are put between parentheses

is not uncommon since this seems to be a feature broadly shared among metazoans, with the metazoan ancestor losing the DHU arm structure before its diversification³⁶. The mitogenome of *H. longicornis* differs from those of *Damon* and *Phrynus*, the more related genera, by lacking DHU and TΨC arms in the *trnA*, while in *Damon diadema* and *Phrynus* sp., this tRNA only lacks the DHU arm²¹. In contrast, many mitochondrial tRNAs of spiders have atypical or aberrant cloverleaf secondary structures, lacking DHU and TΨC arms^{29,33,37}.

The location of the two ribosomal RNA genes is the same in *Charinus* and *Heterophrynus*: *rrnS* is located between *trnV* and the control region and *rrnL* between *trnL1* and *trnV*. The position of these rRNA genes, relative to the annotated regions of the genome, is the same found in the other whip spiders, *Damon* and *Phrynus*, and in Thelyphonida and Schizomida, in accordance with Amblypygi and Uropygi as closely related taxa forming the clade Pedipaldi¹². In contrast, in araneid spiders, *rrnL* is located between the same genes, but *rrnS* is positioned between *trnV* and *trnQ*²⁹. This difference is probably due to changes in the position of *trnQ* since tRNAs are more prone to rearrangement than the PCGs and rRNAs²².

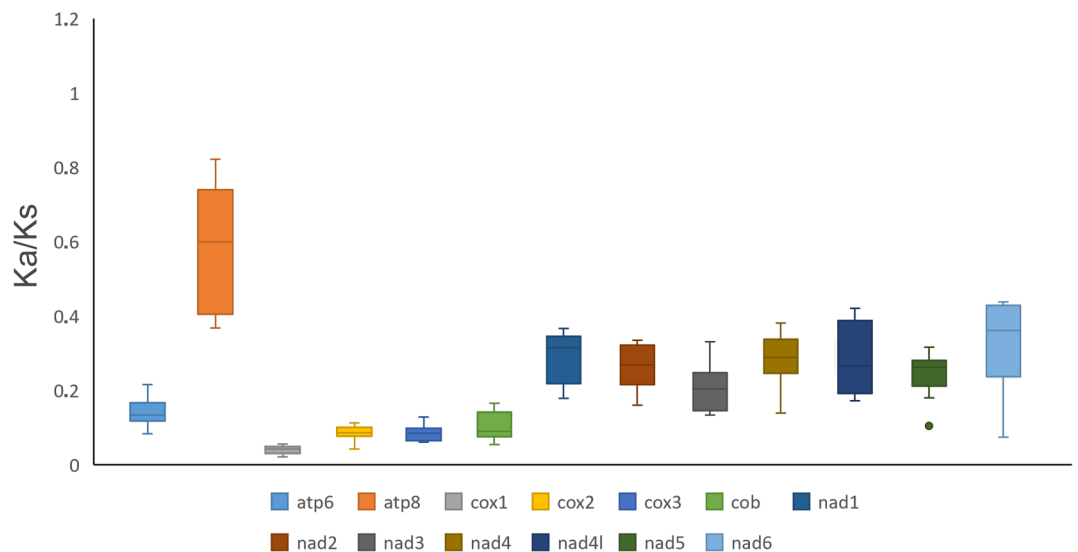


Fig. 2. Box plot for pairwise divergence of Ka/Ks ratio of the 13 PCGs in the mitogenomes of Amblypygi ($n = 5$: *Charinus carajas*, *C. ferreus*; *Damon diadema*, *Heterophrynus longicornis* and *Phrynus* sp.). The analysis was computed in DnaSP v6

Control region

In all mitogenomes, the non-coding control region (CR) was located between *rrnS* and *trnI*, as found in the other known mitogenomes of whip spiders. The base content is A + T rich in the three species, being 69.6% in *H. longicornis*, 68.9% in *C. carajas*, and 65.2% in *C. ferreus*. AT skewness was slightly negative for *H. longicornis* (-0.037) and *C. carajas* (-0.016), contrasting with the positive values discovered for the other regions and the whole mitogenomes, and positive for *C. ferreus* (0.098). GC skewness was negative for all the species (-0.395 in *H. longicornis*, -0.209 in *C. carajas*, and -0.239 in *C. ferreus*).

Phylogenetic analyses

In the taxonomically more comprehensive phylogenetic trees of the three-gene matrix (Fig. 3a, d), *Charinus carajas* and *C. ferreus* only grouped with conspecifics, recovering both as monophyletic groups. The two species were also recovered as sister clades. These relationships were recovered in Bayesian and maximum likelihood (ML) trees with high supports (see Fig. 3a, d), indicating these species are well delimited from each other. Our sequence of *Heterophrynus longicornis* clustered with the other congeneric species, assigning it to this genus with maximum clade support (UFBoot = 100, and percentages of posterior probabilities (PP) = 100). However, the three specimens ascribed to *H. longicornis* did not form a clade. Our sample (ITV2673), from the eastern Amazon (Brazil), recovered as sister to the exemplar Lo55 (HI-B-CdP1-Lo55, BIN: ADC2132), from northeastern Brazil in the Caatinga-Cerrado biome (A Savana-kind ecosystem), whereas the exemplar LP 3830 (AMCC LP 3830), from French Guiana (northern Amazon), retrieved as sister to *H. alces* (Ha-FG-Jose-A7), also from the French Guiana. The exemplar of *H. batesii* came from the southwestern Amazon and was recovered as the sister to the two *H. longicornis* from Brazil. Therefore, all *Heterophrynus* samples from Brazil, irrespective of the taxon assigned, formed a monophyletic group, and the same happened with the specimens from French Guiana. Species of *Heterophrynus* are widely distributed throughout central South America and may occur in sympatry^{16,38}. They can exhibit a high level of intraspecific mtDNA genetic divergence, which accounts for the cryptic diversity in *H. batesii* and *H. longicornis*¹⁶. Thus, the finding of a phylogenetic relationship shaped by geography, unexpected according to taxonomy, indicates that cryptic diversity in the genus needs to be disentangled before any accurate systematic conclusion can be reached. Such a proposal would demand sampling more species and a systematic revision, which is beyond the scope of this paper.

In the maximum likelihood (ML) and Bayesian trees of the nucleotide matrix of 14 genes (Fig. 3b, e), the two *Charinus* species formed a clade with maximum node support (UFBoot = 100, PP = 100). *Heterophrynus longicornis* was sister to the *Charinus* clade in the ML tree and to a clade formed by *Damon diadema* and *Charinus* in the Bayesian tree. However, these relationships were not significant (UFBoot and PP < 95) in both trees. Those topologies did not recover Phrynidae (*Heterophrynus* and *Phrynus*) or Phrynoidea (*Damon*, *Heterophrynus* and *Phrynus*) as monophyletic groups.

In the ML tree of the amino acid (AA) matrix (Fig. 3c), *H. longicornis* was sister to *Damon*, and the two formed a clade with *Phrynus*, thus gathering only all members of Phrynoidea into a monophyletic group, even though Phrynidae has not yet been recovered. Again, the statistical support for those relationships was not significant. Only in the Bayesian tree of the AA matrix (Fig. 3f) did *Heterophrynus* and *Phrynus* arise as sister taxa. *Damon* was recovered as the sister to the clade formed by those two species. Moreover, these relationships were highly supported. Thus, the monophyly of the family Phrynidae and the superfamily Phrynoidea were only recovered in the Bayesian inference from amino acids, in which we implement the CAT mixture model.

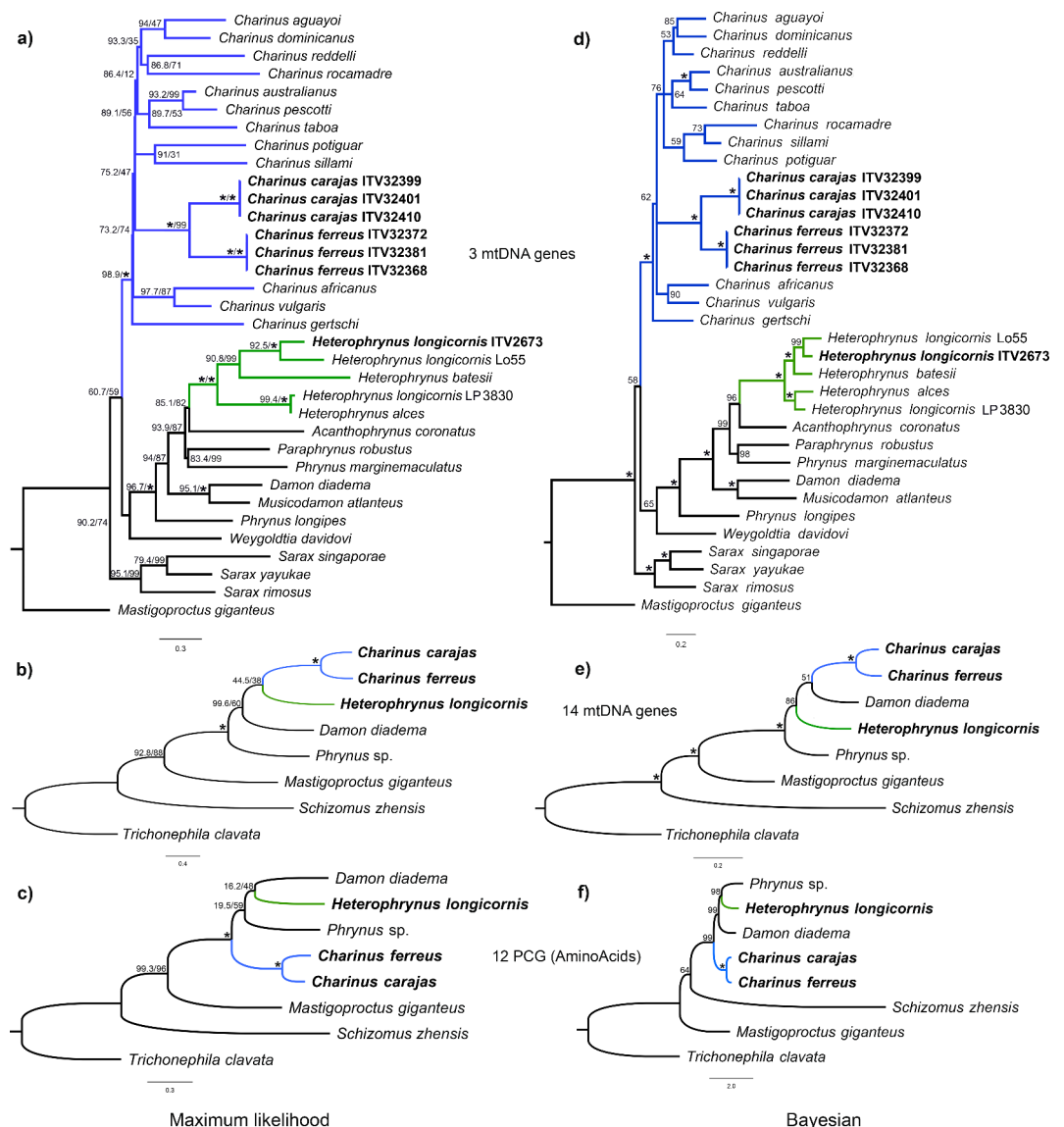


Fig. 3. Phylogenetic trees showing the relationship among the species herein sequenced with other arachnids using maximum likelihood (ML) (a–c) and Bayesian analyses (d–f). Branches in blue represent the genus *Charinus*; branches in green represent the genus *Heterophrynus*. Numbers on nodes are SH-aLRT/UFBoot supports in ML analyses and posterior probabilities (PP) in percentage in Bayesian analyses; asterisks indicate full node supports (i.e., SH-aLRT, UFBoot and PP = 100). Boldface terminals indicate novel mitogenomic data from this study. The three-gene matrix (a, d) with only *cox1* and rRNA sequences has a more extensive taxonomic sampling. The complete datasets, with the mitogenomes of *Damon diadema*, *Phrynus* sp., *Mastigoproctus giganteus*, *Schizomus zhensis*, and *Trichonephila clavata*, used nucleotide sequences of 14 genes (12 PCGs and rRNAs) (b, e) or amino acid sequences of the 12 PCGs (c, f). The monophyly of Phrynidae and Phrynoidea was only obtained in the PhyloBayes tree under the CAT-GTR model on amino acids (f).

In the phylogenomics of Amblypygi, all datasets (UCE matrices and their combination with Sanger loci) recovered the monophyly of Phrynidae and Phrynoidea with full support¹⁰. The success in retrieving these taxa only by the Bayesian CAT-GTR model may have been because site-heterogeneous mixture models better fit highly variable sequences and are less prone to systematic errors than classical empirical matrices^{39–42}. In addition, PhyloBayes has been largely used in mitogenomics and phylogenomics of Arachnida to accommodate full across-site heterogeneity and account for possible sources of phylogenetic errors^{13,22,43–45}. Moreover, the erroneous grouping of unrelated taxa in phylogenetic analyses is more prevalent when the taxon sampling is low and the taxa sampled are distantly related^{13,41,46}, as in our data sets. Nevertheless, as previously mentioned, there were only two available mitogenomes of whip spiders, and even with our contribution of three new species, more than doubling that amount, we still have a tiny fraction of an order having around 270 species. Therefore, a global mitogenomic phylogenetic analysis of whip spiders is far from being reached with current datasets, and it is important to clarify that such an attempt falls outside the scope of this paper. Our focus is on providing new

mitochondrial genomes of individual species and exploring the phylogenetic relationship of these taxa with relatives based on these new data.

Conclusions

The mitogenomes of *Charinus* and *Heterophrynus* herein provided were compared with those of *Damon*, *Phrynus*, and other related arachnid orders, allowing us to find general features shared only or mainly by Amblypygi taxa. Some of these attributes are: (1) biases in nucleotide usage (Cs over Gs), (2) preservation of the presumed ancestral order of genes in arthropods, and (3) most of the tRNAs presenting the typical cloverleaf-shaped secondary structure. These shared features revealed the evolutionary trends followed by the most recent common ancestors of modern whip spiders. Other results were already expected, considering the nature of the mtDNA, as we observed $Ka/Ks < 1$ for all PCGs, a feature that is widespread in the mitogenomes of other arachnids. The phylogenetic analyses assigned the three species mitogenomes to their respective genera and confirmed *Charinus carajas* and *C. ferreus* as sister and well-delimited species. We highlight the supposed advantages of using the Bayesian CAT model to better account for rate heterogeneity among amino acid sites. In general, the results in this paper filled a knowledge gap in the genetics of whip spiders and provided data for future studies on the molecular biodiversity of Amblypygi, particularly those that benefit from the biological properties of the mtDNA, such as demographic histories and phylogeography. Such evolutionary studies will demand a denser sampling of individuals per species and would be especially important for *Charinus carajas* and *C. ferreus*, two short-range endemics and cave-dwelling species in the Amazon⁵, for which there is no genetic data assessing the status of their populations.

Methods

Specimen sample and DNA extraction

The amblypygids herein sequenced were collected in the caves located within the limits of the *Floresta Nacional de Carajás*, in the southeast state of Pará, Brazilian Amazon. All specimens were fixed in 95% ethanol and stored at -20°C until used for DNA extraction. Specimen collections were authorized by the environmental agency ICMBio/MMA under license number 49994-12. Genomic DNA of the samples was isolated using the DNeasy Blood & Tissue Kit (Qiagen), following the manufacturer's protocol for insects, and stored at -20°C . Individual isolates from which the sequences were obtained are ITV2673 for *H. longicornis*; ITV32368, ITV32372, and ITV32381 for *C. ferreus*; and ITV32399, ITV32401, and ITV32410 for *C. carajas*.

Genome sequencing, assembly, annotation, and analyses

Paired-end libraries were constructed using approximately 50 ng of DNA with the QXT SureSelect (Agilent Technologies) and sequenced in an Illumina platform with a NextSeq 500 v2.5 kit high-output (300 cycles), following the manufacturer's instructions. Mitochondrial genomes were *de novo* assembled with NOVOPlasty v4.2⁴⁷ using the *cox1* nucleotide sequence of *Charinus neocaledonicus* (GenBank: MT040916) as seed input, and annotations were performed with MITOS2⁴⁸. MEGA X⁴⁹ was used to calculate nucleotide composition, amino acid frequency, and relative synonymous codon usage (RSCU). The ratios of non-synonymous substitutions (Ka) per synonymous (Ks) substitutions of each protein-coding gene (PCGs) were estimated in the program DnaSP v6⁵⁰ using all the available mitogenomes of whip spiders (Supplementary Table S1). Strand asymmetry was calculated by the following formulas: AT skew = $(A\% - T\%)/(A\% + T\%)$ and GC skew = $(G\% - C\%)/(G\% + C\%)$ ⁵¹.

Model selection and phylogenetic analyses

We conducted two distinct phylogenetic inferences. In one, we used 12 PCGs (all except *nad6*) and the two ribosomal RNAs (rRNAs) of our samples and from the mitogenomes of *Phrynus* sp., *Damon diadema*, *Mastigoproctus giganteus*, *Schizomus zhensis*, and *Trichonephila clavata*, available in the GenBank (Supplementary Table S1), to construct nucleotide (PCGs and rRNAs) and amino acid (AA alignments of PCGs) matrices. These matrices were narrow in terms of taxa sampled but broad in the number of characters used, with *T. clavata* used to root the trees. In the other inference, to include more amblypygid taxa, particularly those more closely related to *Charinus* and *Heterophrynus*, the nucleotide matrix of 14 genes was reduced to comprise only *cox1* and the two rRNAs genes (Supplementary Table S1). Resulting in a DNA matrix broad in the number of taxa used but narrow in terms of characters sampled, in which *M. giganteus* was used to root the trees.

Individual genes were independently aligned using MAFFT v7⁵² with the algorithm *Auto*. The alignments were trimmed to avoid a large amount of missing data at the ends and then concatenated into a unique matrix in Geneious Prime 2022.0.2 (<https://www.geneious.com>)⁵³. For partitioning analyses, we defined data blocks based on genes and codon positions, one for each rRNA and one for each codon position of protein-coding genes, for the DNA alignments, while for the amino acid (AA) data set, we assigned data blocks based on genes. Hence, for the DNA data sets, we set five data blocks for the three-gene matrix and 35 for the 14-gene matrix (we joined *atp6* and *atp8* due to their contiguity in the mitogenomes), whereas for the AA data set, we set 12 data blocks.

We used ModelFinder⁵⁴ to find best-fit partitioning schemes⁵⁵ and model selection for the maximum-likelihood (ML) analyses according to BIC (Supplementary Table S2). We employed the edge-unlinked partition model that accounts for heterotachy, i.e., rate variation across lineages, whilst rate heterogeneity across sites (RHAS) was accommodated by the FreeRate model⁵⁶ (-Q partitions.nex -m MF + MERGE). We used IQ-TREE 2 with hill-climbing and stochastic NNI as the tree search strategy⁵⁷. Nodal support was estimated via 1000 ultrafast bootstrap (UFBoot) replicates and Shimodaira–Hasegawa aLRT (SH-aLRT) with 1000 replicates (-B 1000 -alrt 1000)⁵⁸.

For the Bayesian analyses of the DNA data sets, we used MrBayes 3.2⁵⁹ for the phylogenetic reconstructions and PartitionFinder2⁶⁰ to find best-fit partitioning schemes and model selection as this program provides partition definitions in the specific MrBayes block format (Supplementary Table S3). In PartitionFinder2, branch lengths were set to 'unlinked' for all analyses and the partitioning scheme search was set to 'all' for the reduced three-gene matrix and 'greedy' for the complete data sets. For the trees search, we executed two independent runs of four chains, sampling generations until we reached good convergence and stationarity results (i.e., ESS > 200 in combined runs). Nonetheless, for the Bayesian phylogenetic reconstructions using the AA alignments, we utilized PhyloBayes MPI⁶¹. This software implements the CAT infinite mixture model for modeling among-site variation in the rate of substitution⁶². We initially implemented the CAT-GTR model (virtually always the better-fitted model among those implemented in PhyloBayes), where exchange rates are always considered free parameters and inferred from the data, and checked the runs for satisfactory convergence (i.e., maxdiff < 0.1)^{61,62}.

Data availability

The complete mitogenomes herein sequenced were deposited in GenBank under the accession numbers OQ758255 (ITV32399), OQ758256 (ITV32401) and OQ758257 (ITV32410) for *Charinus carajas*; OQ108440 (ITV32368), OQ108441 (ITV32372) and OQ108442 (ITV32381) for *C. ferreus*; and OP573470 for *Heterophrynus longicornis*.

Received: 30 May 2024; Accepted: 23 October 2024

Published online: 01 November 2024

References

- Weygoldt, P. *Whip Spiders (Chelicerata: Amblypygi). Their Biology, Morphology and Systematics* (Apollo Brooks, 2000).
- World Amblypygi Catalog. World Amblypygi Catalog. Natural History Museum Bern. (2022). <https://wac.nmbe.ch>
- Chapin, K. J. & Hebets, E. A. The behavioral ecology of amblypygids. *J. Arachnol.* **44**, 1–14 (2016).
- De Miranda, G. S., Giupponi, A. P. L., Scharff, N. & Prendini, L. Phylogeny and biogeography of the pantropical whip spider family Charinidae (Arachnida: Amblypygi). *Zool. J. Linn. Soc.* **194**, 136–180 (2022).
- de Giupponi, A. P., de Miranda, G. S. & L. & Eight New species of *Charinus* Simon, 1892 (Arachnida: Amblypygi: Charinidae) endemic for the Brazilian Amazon, with notes on their conservational status. *PLoS ONE*. **11**, e0148277 (2016).
- Weygoldt, P. Evolutionary morphology of whip spiders: towards a phylogenetic system (Chelicerata: Arachnida: Amblypygi)*. *J. Zoological Syst. Evolutionary Res.* **34**, 185–202 (1996).
- De Miranda, G. S., Kury, A. B. & Giupponi, A. P. D. L. Review of *Trichodamon* Mello-Leitão 1935 and phylogenetic placement of the genus in Phrynichidae (Arachnida, Amblypygi). *Zool. Anzeiger*. **273**, 33–55 (2018).
- de Miranda, G. S., Giupponi, A. P. L., Prendini, L. & Scharff, N. Systematic revision of the pantropical whip spider family Charinidae Quintero, 1986 (Arachnida, Amblypygi). *EJT*. **772**, 1–409 (2021).
- Prendini, L. Systematics of the group of African whip spiders (Chelicerata: Amblypygi): evidence from behaviour, morphology and DNA. *Org. Divers. Evol.* **5**, 203–236 (2005).
- De Miranda, G. S. et al. The rediscovery of a Relict unlocks the First Global phylogeny of whip spiders (Amblypygi). *Syst. Biol.* **syae021** <https://doi.org/10.1093/sysbio/syae021> (2024).
- Garwood, R. J., Dunlop, J. A., Knecht, B. J. & Hegna, T. A. The phylogeny of fossil whip spiders. *BMC Evol. Biol.* **17**, 105 (2017).
- Shultz, J. W. A phylogenetic analysis of the arachnid orders based on morphological characters. *Zool. J. Linn. Soc.* **150**, 221–265 (2007).
- Ballesteros, J. A. et al. Comprehensive species sampling and sophisticated algorithmic approaches refute the Monophyly of Arachnida. *Mol. Biol. Evol.* **39**, msac021 (2022).
- Shear, W. A., Selden, P. A., Rolfé, W. D. I., Bonamo, P. M. & Grierson, J. D. New Terrestrial Arachnids from the Devonian of Gilboa, New York (Arachnida, Trigonotarbita). *American Museum Novitates* 1–74 (1987).
- Howard, R. J., Puttick, M. N., Edgecombe, G. D. & Lozano-Fernandez, J. Arachnid monophyly: morphological, palaeontological and molecular support for a single terrestrialization within Chelicerata. *Arthropod Struct. Dev.* **59**, 100997 (2020).
- Reveillon, F., Wattier, R., Montuire, S., Carvalho, L. S. & Bollache, L. Cryptic diversity within three south American whip spider species (Arachnida, Amblypygi). *Zool. Res.* **41**, 595–598 (2020).
- Seiter, M. et al. Cryptic diversity in the whip spider genus *Paraphrynus* (Amblypygi: Phrynidae): integrating morphology, karyotype and DNA (2020). <https://doi.org/10.26049/ASP78-2-2020-04>
- Agnarsson, I., Coddington, J. A., Caicedo-Quiroga, L., May-Collado, L. J. & Pálsson, S. Deep mtDNA sequence divergences and possible species Radiation of Whip spiders (Arachnida, Amblypygi, Phrynidae, Phrynus/Paraphrynus) among Caribbean Oceanic and Cave Islands. *Taxonomy*. **3**, 133–147 (2023).
- Espósito, L. A. et al. Islands within islands: diversification of tailless whip spiders (Amblypygi, Phrynus) in Caribbean caves. *Mol. Phylogenet. Evol.* **93**, 107–117 (2015).
- Schramm, F. D., Valdez-Mondragón, A. & Prendini, L. Volcanism and palaeoclimate change drive diversification of the world's largest whip spider (Amblypygi). *Mol. Ecol.* **30**, 2872–2890 (2021).
- Fahren, K., Masta, S. E. & Podsiadlowski, L. The first complete mitochondrial genome sequences of Amblypygi (Chelicerata: Arachnida) reveal conservation of the ancestral arthropod gene order. *Genome*. **52**, 456–466 (2009).
- Ban, X., Shao, Z., Wu, L., Sun, J. & Xue, X. Highly diversified mitochondrial genomes provide new evidence for interordinal relationships in the Arachnida. *Cladistics*. **38**, 452–464 (2022).
- Galtier, N., Nabholz, B., Glémin, S. & Hurst, G.D.D. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* **18**, 4541–4550 (2009).
- Bernt, M., Braband, A., Schierwater, B. & Stadler, P. F. Genetic aspects of mitochondrial genome evolution. *Mol. Phylogenet. Evol.* **69**, 328–338 (2013).
- Avise, J. C. et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**, 489–522 (1987).
- Ratnasingham, S., & Hebert, P. D. N. bold: the barcode of life data system. *Mol. Ecol. Notes*. **7**, 355–364 (2007). <http://www.barcodinglife.org>
- Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA. *Global Ecol. Conserv.* **17**, e00547 (2019).
- Bu, Y., Gao, Y., & Nunes Godeiro, N. The first complete mitochondrial genome of the micro-whip scorpion *Schizomus Zhensis* (Arachnida: Schizomida) and phylogenetic analysis. *Mitochond. DNA Part. B*. **7**, 709–711 (2022).

29. Yong, H. S. et al. Complete mitochondrial genomes and phylogenetic relationships of the genera *Nephila* and *Trichonephila* (Araneae, Araneoidea). *Sci. Rep.* **11**, 10680 (2021).
30. Tyagi, K. et al. The gene arrangement and phylogeny using mitochondrial genomes in spiders (Arachnida: Araneae). *Int. J. Biol. Macromol.* **146**, 488–496 (2020).
31. Sterling-Montealegre, R. A. & Prada, C. F. Variability and evolution of gene order rearrangement in mitochondrial genomes of arthropods (except Hexapoda). *Gene*. **892**, 147906 (2024).
32. Ojala, D., Montoya, J. & Attardi, G. tRNA punctuation model of RNA processing in human mitochondria. *Nature*. **290**, 470–474 (1981).
33. Kumar, V. et al. The complete mitochondrial genome of endemic giant tarantula, *Lyrognathus crotalus* (Araneae: Theraphosidae) and comparative analysis. *Sci. Rep.* **10**, 74 (2020).
34. Young, M. R., deWaard, J. R. & Hebert, P. D. N. DNA barcodes enable higher taxonomic assignments in the Acari. *Sci. Rep.* **11**, 15922 (2021).
35. Astrin, J. J. et al. Towards a DNA barcode reference database for spiders and harvestmen of Germany. *PLOS ONE*. **11**, e0162624 (2016).
36. Wolstenholme, D. R. Animal Mitochondrial DNA: Structure and Evolution. in *International Review of Cytology* (eds. Wolstenholme, D. R. & Jeon, K. W.) vol. 141 173–216 (Academic Press, 1992).
37. Masta, S. E. & Boore, J. L. Parallel evolution of truncated transfer RNA genes in Arachnid mitochondrial genomes. *Mol. Biol. Evol.* **25**, 949–959 (2008).
38. Carvalho, L. S., Oliveira-Marques, F. N., Silva, P. R. R. & Arachnida Amblypygi, *Heterophrynus longicornis* (Butler, 1873): distribution extension for the state of Piauí northeastern Brazil. *cl* **7**, 267 (2011).
39. Philippe, H. et al. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).
40. Kapli, P., Flouri, T. & Telford, M. J. Systematic errors in phylogenetic trees. *Curr. Biol.* **31**, R59–R64 (2021).
41. Xu, S. et al. Comparative mitogenomics and phylogenetic analyses of Pentatomoidea (Hemiptera: Heteroptera). *Genes*. **12**, 1306 (2021).
42. Cai, C. Ant backbone phylogeny resolved by modelling compositional heterogeneity among sites in genomic data. *Commun. Biol.* **7**, 106 (2024).
43. Ballesteros, J. A. & Sharma, P. P. A critical appraisal of the placement of Xiphosura (Chelicerata) with account of known sources of phylogenetic error. *Syst. Biol.* **68**, 896–917 (2019).
44. Sharma, P. P. et al. Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Mol. Biol. Evol.* **31**, 2963–2984 (2014).
45. Benavides, L. R., Pinto-da-Rocha, R. & Giribet, G. The phylogeny and evolution of the flashiest of the Armored Harvestmen (Arachnida: Opiliones). *Syst. Biol.* **70**, 648–659 (2021).
46. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7**, S4 (2007).
47. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucl. Acids Res.* **gkw955** <https://doi.org/10.1093/nar/gkw955> (2016).
48. Bernt, M. et al. Improved de novo metazoan mitochondrial genome annotation. *MITOS Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
49. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
50. Rozas, J. et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
51. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
52. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
53. Kearse, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. **28**, 1647–1649 (2012).
54. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. E., Von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods*. **14**, 587–589 (2017).
55. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).
56. Soubrier, J. et al. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **29**, 3345–3358 (2012).
57. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
58. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
59. Ronquist, F. et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
60. 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. *Mol. Biol. Evol.* **34**, 772–773 (2017).
61. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
62. Lartillot, N. & Philippe, H. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).

Acknowledgements

This research was funded by Vale S.A. (Projeto Cavidades, RBRS000603.84).

Author contributions

GO provided funding. LCT collected the samples. ESP and SV prepared samples for DNA extractions and implemented the DNA sequencing. AFN executed the bioinformatics analyses (assembly, annotation and phylogenetics) and wrote the manuscript. GO, LCT and SV revised the manuscript. SV supervised the work.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-77525-0>.

Correspondence and requests for materials should be addressed to A.F.N. or S.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024