



Complete Mitochondrial Genome of a Tongue Worm *Armillifer agkistrodontis*

Jian Li¹, Fu-Nan He¹, Hong-Xiang Zheng¹, Rui-Xiang Zhang¹, Yi-Jing Ren^{1,2}, Wei Hu^{1,3,*}

¹Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China; ²College of Animal Science and Technology, Guangxi University, Nanning, Guangxi 530005, China; ³National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, China

Abstract: *Armillifer agkistrodontis* (Ichthyostroaca: Pentastomida) is a parasitic pathogen, only reported in China, which can cause a zoonotic disease, pentastomiasis. A complete mitochondrial (mt) genome was 16,521 bp comprising 13 protein-coding genes (PCGs), 22 tRNA genes, 2 rRNA genes, and 1 non-coding region (NCR). A phylogenetic tree drawn with the concatenated amino acid sequences of the 6 conserved PCGs (*atp6*, *cox1-3*, and *nad2*) showed that *A. agkistrodontis* and *Armillifer armillatus* constituted a clade Pentastomida which was a sister group of the Branchiura. The complete mt genome sequence of *A. agkistrodontis* provides important genetic markers for both phylogenetic and epidemiological studies of pentastomids.

Key words: *Armillifer agkistrodontis*, mitochondrial genome, pentastomid, pentastome, phylogenetics

Human pentastomiasis is a zoonotic parasitic disease, which is caused by infection with the nymphal stage of the pentastomid. Although more than 10 pathogenic species associated with human disease have so far been identified, *Linguatula ser-rata* and *Armillifer armillatus* are implicated in the majority of cases [1,2]. Two of these pathogenic species, *Armillifer monili-formis* and *Armillifer agkistrodontis*, are found exclusively in China, where the pit viper snake (*Deinagkistrodon acutus*) is the most common final host for *A. agkistrodontis* [3]. Infections are acquired by the ingestion of snake meat, blood, and bile contaminated with mature eggs containing the primary larva. Recently, there has been an increase in infection cases of *A. monili-formis* and *A. agkistrodontis* raising the profile of this previously neglected disease [2,4,5].

The nymphs that cause human visceral pentastomiasis are mainly located in or on the liver, the intestinal wall, and mesentery organs, as a calcified cyst. Since the characteristics of the nymph are not completely developed, it is necessary to combine both morphological and molecular approaches to facilitate the disease diagnosis and species identification. The sequence of the mitochondrial (mt) genome has been widely

used as a specific molecular marker for species identification as it contains more information than single genes or fragments [6].

In this work, we have released the complete sequence for the mt genome of *A. agkistrodontis*, annotated for gene order and codon usage for protein-coding genes (PCGs). Furthermore, we have attempted to use of the mt PCGs data for phylogenetic analysis to define the relationship with other arthropods, especially Mandibulata and Pancrustacea. We have provided evidence, based on the analysis of this mt genome, supporting the placement of Pentastomida as a sister group to Branchiura within the Pancrustacea clade. This result suggested that the best strategy for utilizing the mt genome for phylogenetic analysis should be the 6 conserved PCGs (*atp6*, *cox1-3*, and *nad2*).

Adult individuals of *A. agkistrodontis* were collected from the visceral organ and air sac of captive *D. acutus* purchased from the market in Taizhou city, Zhejiang province, P. R. China. These pentastomids were washed with saline buffer (0.7% NaCl) before being fixed in 70% (v/v) ethanol at 4°C. Species and sexual identification of pentastomids was defined by morphological characteristics [7], after which the tegument was isolated from individual females by excision under a stereomicroscope. Total genomic DNA was extracted from this tegument tissue using a DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, and stored at -20°C until use. The identity of *A. agkistrodontis* specimens

•Received 10 October 2016, revised 22 November 2016, accepted 29 November 2016.

*Corresponding author (huw@fudan.edu.cn)

© 2016, Korean Society for Parasitology and Tropical Medicine

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

was also confirmed by sequencing a gene fragment, coding for 18S rRNA, amplified by PCR from total DNA [3].

The complete mt genomic sequence of *A. agkistrodontis* was amplified, as described previously, in 6 overlapping fragments, using 6 pairs of oligonucleotide primers designed based on the conserved regions from published complete mtDNA sequences of *A. armillatus* (no. NC_005934.1) (Table 1) [8,9].

Table 1. Primer sequences used to amplify PCR fragments of *Armillifer agkistrodontis*

Primer name	Sequences (5'-3')	Size (kb)
AA-1	F: ACTTATTAGTCGAACAGA R: AATCCTGTTCATACGCTA	~6
AA-2	F: ACCTCAAAGCTGGAAT R: AAGAGCTGGAATGGAAGT	~2.2
AA-3	F: CTTGCCAACTCCTCCATTGATACTGCC R: GTTGGCTATCAGGCGTAG	~1.8
AA-4	F: TATAACCATAATTGAACTCCTCAGC R: TCTACGTTGAGGCAGCTA	~2.6
AA-5	F: CATTTCATCGCACATCCTG R: GTCCACCTCATAGTCAGT	~3.4
AA-6	F: TTCTCCTCATAATACTTC R: GTACATATTGCCCGTCGC	~3

These amplicons were sequenced in both directions by primer walking using an ABI 3730 DNA analyzer (Genewiz, Inc., Suzhou, China).

The complete mt genome sequence of *A. agkistrodontis* was assembled manually from individual sequence runs and then aligned against the mt genome sequences of *A. armillatus* by the program MAFFT 7.122. Thirteen PCGs were identified and translated into amino acid sequences using MEGA 6.0 with the invertebrate mitochondrial code option. The tRNA genes were identified using online services ARWEN (<http://130.235.46.10/ARWEN/>) and tRNAscan-SE (<http://selab.janelia.org/tRNAscan-SE/>). tRNA genes that could not be determined by either program, were identified by comparison with the annotations of for the *A. armillatus* mt genome.

To investigate the sequence similarity, 12 concatenated amino acid sequences translated from each of 12 PCGs, except *atp8*, from the mt genomes were aligned with those of 21 published mt genomes from selected Arthropoda species: *Argulus americanus*, *Ammothoa carolinensis*, *A. armillatus*, *Anopheles quadrimaculatus*, *Abacion magnum*, *Blattella bisignata*, *Calanus*

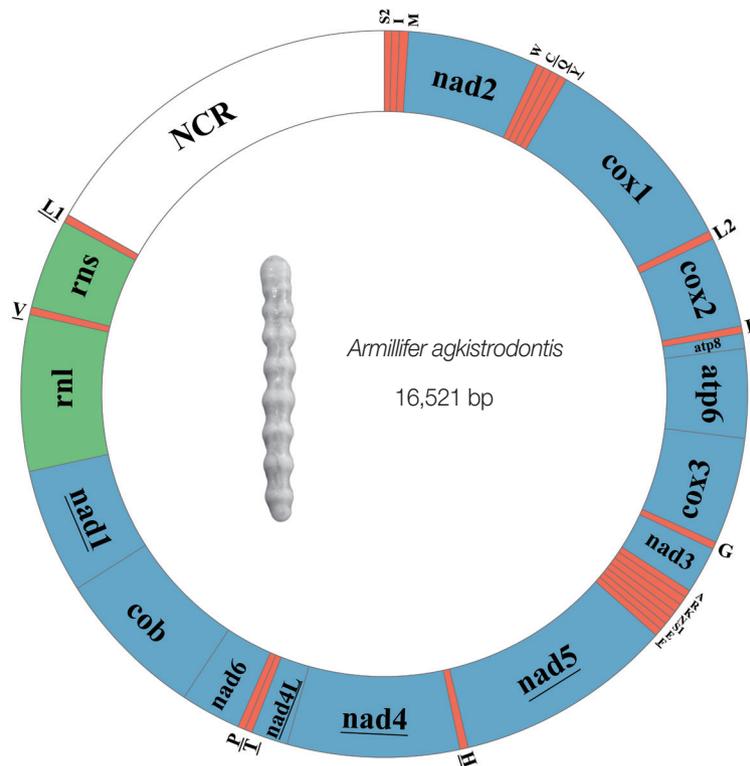


Fig. 1. The mitochondrial genome of *Armillifer agkistrodontis*. The underlined genes are transcribed anticlockwise, while those without underlining are transcribed clockwise. All the tRNA genes are indicated with the 1-letter code of their corresponding amino acids. There are 2 tRNA genes for leucine: L1 for codons CUN and L2 for UUR; and 2 tRNA genes for serine: S1 for codons AGN and S2 for UCN. “NCR” refers to the non-coding region.

hyperboreus, *Colossendeis megalonyx*, *Daphnia magna*, *Euphausia pacifica*, *Hutchinsoniella macracantha*, *Ixodes ricinus*, *Lepas australis*, *Nobia grandis*, *Rhipicephalus microplus*, *Scutigera causeyae*, *Strigamia maritima*, *Sericinus montela*, *Trichophthalma punctata*, *Opisthopatus cinctipes*, and *Metaperipatus inae*.

Alignments of both nucleotide and amino acid sequences were performed using Muscle 3.8.31 [10]. The amino acid sequences were extracted from 22 selected mitochondrial genomes (including our data), and 12 of the PCGs, except for *atp8*, were aligned. The nucleotide sequence from the genes most conserved (*atp6*, *cox1-3*, and *nad2*) in arthropod species [11] were aligned, using the amino acid alignment as a guide, with online software TranslatorX [12]. All the ambiguous gaps were excluded using GBlock [13], resulting in 3870 bases in the final nucleotide alignment.

Amino acid sequence sets were compared using a Bayesian phylogenetic analysis with a mtREV replacement model where the gamma distribution and a fraction of sites were constrained to be invariable. In order to minimize long-branch bias, a Neutral Transitions Excluded method (NTE) was used for this phylogenetic analysis [14]. All Bayesian analyses were generated by MrBayes 3.2 with 5 independent Markov chains run for

2,000,000 generations, a sampling frequency of 1,000 and a burn-in of 250 trees [15].

The complete mt genome of *A. agkistrodantis* was 16,521 bp in size (no. KX686568), slightly smaller than that of *A. armillatus* (16,747 bp), with a difference in length mainly due to the non-coding region (NCR). The gene content was typical, including 13 PCGs (*atp6*, *atp8*, *cox1-3*, *nad1-6*, *cytb*, and *nad4L*), 22 tRNA genes, 2 rRNA genes (*rns* and *rnl*), and 1 NCR (Fig. 1; Table S1). The gene order and transcriptional orientation were identical to that of *A. armillatus*, which has been hypothesized to be the ancestral arrangement for the mt genome of insects [8]. The overlapping nucleotides between mt genes in *A. agkistrodantis* ranged from 1 to 22 bp, the longest overlap located between *cytb* and *nad1* (Table S1). The nucleotide content was 34.20%, 27.33%, 31.01%, and 7.46%, for A, C, T, and G, respectively, with the overall A+T content being 65.21%, ranging from 61.44% (*cytb* and *nad1*) to 72.97% (*atp8*) across the 13 PCGs, 2 rRNA genes, and NCR.

The 13 PCGs, 10,652 bp in length (including the terminator codons) accounted for 64.46% of the entire mt genome. These PCGs had a consensus start codon of ATN (1 with ATT, 2 with ATC, 4 with ATA, and 5 with ATG) with only 1 being

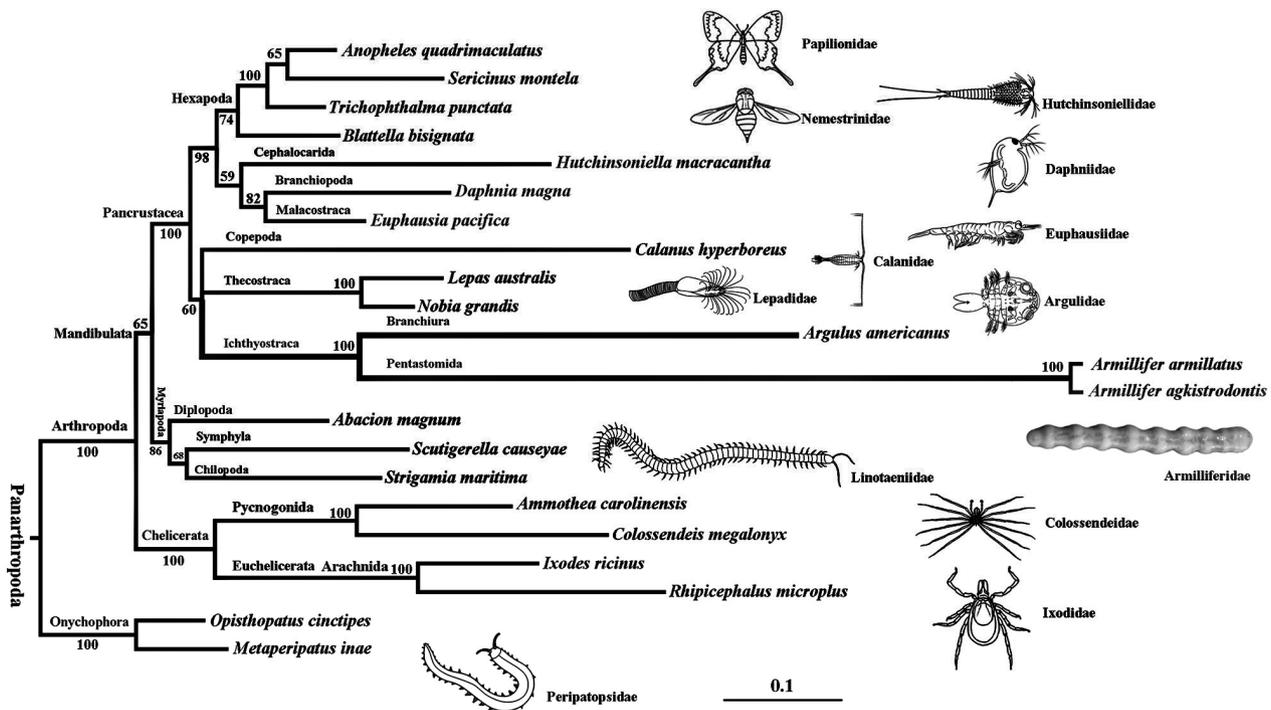


Fig. 2. A phylogenetic tree of *Armillifer agkistrodantis* with other arthropods based on 6 conserved mitochondrial protein coding genes. *Opisthopatus cinctipes* and *Metaperipatus inae* (Onychophora) were used as outgroups.

CTG (*cox1*). Conventional stop codons of TAA and TAG were found in 8 of 13 PCGs; however, 5 PCGs (*cox1*, *cox2*, *cox3*, *nad4*, and *nad5*) had incomplete termination codons (T) (Table S1). The codon usage pattern of the 13 PCGs is shown in Table S1, with A+T rich codons most frequently used; ATT (Ile: 5.24%), TTA (Leu: 4.90%), and ATA (Met: 4.68%). The least used codons were CGG (Arg: 0.11%), ACG (Thr: 0.11%), and GCG (Ala: 0.17%).

The 22 tRNA genes typically found in metazoan mt genomes were also identified in *A. agkistrodontis*, ranging in length from 53 bp to 65 bp (Table S1), with 16 of them determined by ARWEN and tRNAscan-SE. The remaining 6 tRNA genes (*tRNA-Val*, *tRNA-His*, *tRNA-Phe*, *tRNA-Ser^{AGN}*, *tRNA-Cys*, and *tRNA-Met*) not identified in this manner, were assigned by comparison with the tRNA genes of *A. armillatus* (no. NC_005934.1). Two rRNA genes (*rns* and *rml*) separated by a *tRNA-Val* were identified in the mt genome, with the boundaries defined by contiguous genes (*nad1* and *tRNA-Leu^{UUR}*). The *rns* and *rml* were found to be 659 bp and 1,144 bp, respectively.

The major NCR of 2,769 bp found between *tRNA-Leu^{UUR}* and *tRNA-Ser^{UCN}*, was smaller than that of *A. armillatus* (2,993 bp). This region could be divided into 3 parts, with the first being 837 bp in length and having an A+T content of 52.93%. The second region was the longest and contained 2 tandem repeats (1,496 bp in length) with 72.41% A+T nucleotides, while the third part was 62 bp in length and located before the *tRNA-Ser^{UCN}* gene.

Phylogenetic analysis based on 6 PCGs (*atp6*, *cox1-3*, and *nad2*), placed *A. agkistrodontis* and *A. armillatus* as a definitive cluster in the clade Pentastomida, grouped with *Argulus americanus* (Branchiura) in the clade Ichthyostraca [16]. The resulting phylogenetic tree also revealed that all the species in Arthropoda are distributed into 3 independent clades: Chelicerata; Mandibulata, and Pancrustacea (Fig. 2). This is consistent with the phylogenomic analysis of nuclear protein-coding sequences [17], in which this clade was named as Oligostraca to represent a subclass in the clade Maxillopoda (Branchiura, Pentastomida, and Mystacocarida) and clade Ostracoda. However, when we constructed a phylogenetic tree based on the analysis of 12 PCGs (except for *atp8*), the clade Ichthyostraca (containing Branchiura and Pentastomida) was located as a sister group with Chelicerata (Fig. 1S). We suspect this analysis is skewed by long-branch attraction bias caused by the noise introduced with the inclusion of these more rapidly evolving genes. For this reason, we suggest using the 6 most conserved

PCGs (*atp6*, *cox1-3*, and *nad2*) be used to analyze the phylogenetic relationships among the Pentastomida within other species. While evidence from palaeontological [18-20], comparative spermatology [21-23], and molecular biomarkers [8,16] has been used to classify and group pentastomids, the actual evolutionary relationship between these species is, as yet, unresolved and remains controversial. Therefore, complete genome information from more pentastomid species is required to clarify this issue.

In conclusion, this study presents the first complete mt genome sequence from *A. agkistrodontis*. Our phylogenetic analysis supports the previous observation that *A. agkistrodontis* has strong similarity to the Branchiura and should be grouped in the Crustacea. Finally, the data presented here have provided more molecular biomarkers for future studies of epidemiology, discrimination of cryptic species, and the analysis of genetic diversity, and systematic classification of pentastomids.

ACKNOWLEDGMENT

This work was supported, in part, by the National Science and Technology Major Project (no. 2009ZX10004-302) to the corresponding author (Hu Wei).

CONFLICT OF INTEREST

There is no conflict of interest related to this work.

REFERENCES

- Riley J. The biology of pentastomids. *Adv Parasitol* 1986; 25:45-128.
- Yao MH, Wu F, Tang LF. Human pentastomiasis in China: case report and literature review. *J Parasitol* 2008; 94: 1295-1298.
- Chen SH, Liu Q, Zhang YN, Chen JX, Li H, Chen Y, Steinmann P, Zhou XN. Multi-host model-based identification of *Armillifer agkistrodontis* (Pentastomida), a new zoonotic parasite from China. *PLoS Negl Trop Dis* 2010; 4: e647.
- Wang HY, Zhu GH, Luo SS, Jiang KW. Childhood pentastomiasis: a report of three cases with the following-up data. *Parasitol Int* 2013; 62: 289-292.
- Ye F, Sheng ZK, Li JJ, Sheng JF. Severe pentastomiasis in children: a report of 2 cases. *Southeast Asian J Trop Med Public Health* 2013; 44: 25-30.
- Shao R, Barker SC. Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. *Parasitology* 2007; 134: 153-167.
- Self JJ, Kuntz RE. New Pentastomida, *Sambonia parapodium* n. sp.

- from *Varanus salvator*, and *Armillifer agkistrodantis* n. sp. from *Agkistrodon acutus*. *Trans Am Microscop Soc* 1966; 85: 256-260.
8. Lavrov DV, Brown WM, Boore JL. Phylogenetic position of the Pentastomida and (pan)crustacean relationships. *Proc Biol Sci* 2004; 271: 537-544.
 9. Hu M, Jex AR, Campbell BE, Gasser RB. Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. *Nat Protoc* 2007; 2: 2339-2344.
 10. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32: 1792-1797.
 11. 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. *Syst Biol* 2005; 54: 277-298.
 12. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* 2010; 38: 7-13.
 13. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; 17: 540-552.
 14. Hassanin A. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. *Mol Phylogenet Evol* 2006; 38: 100-116.
 15. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001; 17: 754-755
 16. Zrzavý J, Hypša V, Vlášková M. Arthropod phylogeny: taxonomic congruence, total evidence and conditional combination approaches to morphological and molecular data sets. In Fortey RA, Thomas RH eds, *Arthropod relationships*. London, UK: Springer. 1998, pp 97-107.
 17. Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* 2010; 463: 1079-1083.
 18. Waloszek D, Repetski JE, Maas A. A new Late Cambrian pentastomid and a review of the relationships of this parasitic group. *Earth Environ Sci Trans R Soc Edinb* 2005; 96: 163-176.
 19. Castellani C, Maas A, Waloszek D, Haug JT. New pentastomids from the Late Cambrian of Sweden-deeper insight of the ontogeny of fossil tongue worms. *Palaeontogr Abt A* 2011; 293: 95-145.
 20. Siveter DJ, Briggs DE, Siveter DJ, Sutton MD. A 425-million-year-old silurian pentastomid parasitic on ostracods. *Curr Biol* 2015; 25: 1632-1637.
 21. Wingstrand KG. Comparative spermatology of a pentastomid, *Raillietiella hemidactyli*, and a branchiuran crustacean, *Argulus foliaceus*, with a discussion of pentastomid relationships. *Biol Skrifter* 1972; 19: 1-72
 22. Riley J, Banaja AA, James JL. The phylogenetic relationships of the pentastomida: The case for their inclusion within the crustacea. *Int J Parasitol* 1978; 8: 245-254.
 23. Storch V, Jamieson BGM. Further spermatological evidence for including the Pentastomida (tongue worms) in the Crustacea. *Int J Parasitol* 1992; 22: 95-108.

Table S1. Organization of the mitochondrial genome of *Armillifer agkistrodantis*

Gene/region	Strand	Positions	Size (bp)	No. of aa1	Ini/Ter codons	Anticodons	In
<i>tRNA-Ser^{UCN}</i> (S2)	+	1-54	54			TGA	0
<i>tRNA-Ile</i> (I)	+	61-120	60			GAT	6
<i>tRNA-Met</i> (M)	+	122-182	61			CAT	1
<i>nad2</i>	+	183-1,145	963	320	ATT/TAG		0
<i>tRNA-Trp</i> (W)	+	1,144-1,200	57			TCA	-2
<i>tRNA-Cys</i> (C)	-	1,193-1,252	60			GCA	-8
<i>tRNA-Gln</i> (Q)	-	1,251-1,315	65			TTG	-2
<i>tRNA-Tyr</i> (Y)	-	1,318-1,377	60			GTA	2
<i>cox1</i>	+	1,379-2,903	1,525	508	CTG/T		1
<i>tRNA-Leu^{JUR}</i> (L2)	+	2,904-2,966	63			TAA	0
<i>cox2</i>	+	2,967-3,630	664	221	ATA/T		0
<i>tRNA-Asp</i> (D)	+	3,631-3,685	55			GTC	0
<i>atp8</i>	+	3,686-3,796	111	36	ATA/TAA		0
<i>atp6</i>	+	3,793-4,446	654	218	ATA/TAA		-4
<i>cox3</i>	+	4,446-5,226	781	260	ATG/T		-1
<i>tRNA-Gly</i> (G)	+	5,227-5,282	56			TCC	0
<i>nad3</i>	+	5,283-5,627	345	114	ATC/TAA		0
<i>tRNA-Ala</i> (A)	+	5,626-5,682	57			TGC	-2
<i>tRNA-Arg</i> (R)	+	5,682-5,744	63			TCG	-1
<i>tRNA-Lys</i> (K)	+	5,744-5,800	57			TTT	-1
<i>tRNA-Asn</i> (N)	+	5,800-5,852	53			GTT	-1
<i>tRNA-Ser^{AGN}</i> (S1)	+	5,852-5,908	57			TCT	-1
<i>tRNA-Glu</i> (E)	+	5,909-5,963	55			TTC	0
<i>tRNA-Phe</i> (F)	-	5,962-6,018	57			GAA	-2
<i>nad5</i>	-	6,019-7,624	1,606	535	ATG/T		0
<i>tRNA-His</i> (H)	-	7,626-7,687	62			GTG	1
<i>nad4</i>	-	7,687-8,947	1,261	420	ATG/T		-1
<i>nad4L</i>	-	8,941-9,210	270	89	ATA/TAG		-7
<i>tRNA-Thr</i> (T)	-	9,227-9,282	56			TGT	16
<i>tRNA-Pro</i> (P)	-	9,283-9,338	56			TGG	0
<i>nad6</i>	+	9,340-9,783	444	148	ATC/TAA		1
<i>cytb</i>	+	9,783-10,892	1,110	370	ATG/TAA		-1
<i>nad1</i>	-	10,915-11,832	918	305	ATG/TAA		22
<i>rnl</i>		11,833-12,976	1,144				0
<i>tRNA-Val</i> (V)	-	12,977-13,033	57			TAC	0
<i>rns</i>		13,034-13,692	659				0
<i>tRNA-Leu^{CUN}</i> (L1)	-	13,693-13,752	60			TAG	0
<i>NCR</i>		13,753-16,521	2,769				0

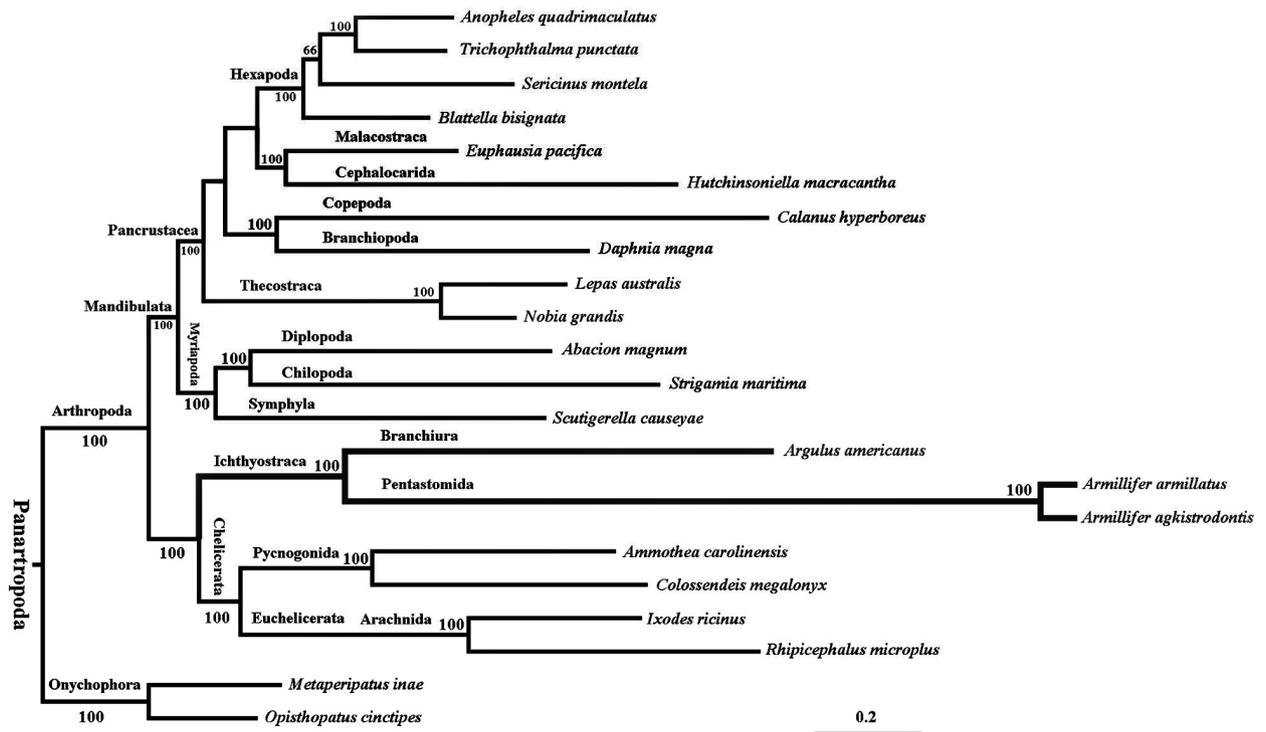


Fig. S1. A phylogenetic tree of *Armillifer agkistrodontis* with other arthropods based on concatenated amino acid sequence data representing 12 mitochondrial protein coding genes. *Opisthoptatus cinctipes* and *Metaperipatus inae* (Onychophora) were used as outgroups.