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 serrata and Armillifer armillatus are implicated in the majority of cases [1,2]. Two of these pathogenic species, Armillifer moniliformis and Armillifer agkistrodonis, are found exclusively in China, where the pit viper snake (Deinagkistrodon acutus) is the most common final host for A. agkistrodonis [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. moniliformis and A. agkistrodonis 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. agkistrodonis, 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) showed that A. agkistrodonis and Armillifer armillatus constituted a clade Pentastomida which was a sister group of the Branchiura. The complete mt genome sequence of A. agkistrodonis provides important genetic markers for both phylogenetic and epidemiological studies of pentastomids.
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].

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 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*, *Ammothea carolinensis*, *A. armillatus*, *Anopheles quadrimaculatus*, *Abacion magnum*, *Blattella bisignata*, *Calanus*

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

| Primer name | Sequences (5'-3') | Size (kb) |
|-------------|-------------------|-----------|
| AA-1        | F: ACTTATTAGTCGAACAGA  
R: AATCCCTGTTCTACGCTA  | ~6        |
| AA-2        | F: ACCCTCAAAGCTGGAAAT  
R: AAGAGCCTGGAATGGAAGT  | ~2.2      |
| AA-3        | F: CTGGCCAACCTCCTCAAGATCTG  
R: GTGGGCTCTCAAGGCGTAG  | ~1.8      |
| AA-4        | F: TATAACCATAATTGAACCTCAGC  
R: TCTAAGGTAGGGAGCAGCTA  | ~2.6      |
| AA-5        | F: CATTGATGCAGCACCCTCCTG  
R: GCTCAGCTAAATGTAAGCT  | ~3.4      |
| AA-6        | F: TCTCCTCTGATATAACTTC  
R: GACATATTGGGCGGTGTCG  | ~3        |

**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, Scutigerella causeyae, Strigamia maritima, Sericinus montela, Trichophthalamia punctate, 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. agkistrodontis 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 (rrnS and rrnL), 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. agkistrodontis 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 agkistrodontis 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 AR-WEN and tRNAscan-SE. The remaining 6 tRNA genes (rrnS and rrnL) separated by a tRNA-Val were identified in the mt genome, with the boundaries defined by contiguous genes (nad1 and tRNA-Leu^UUR). The rrnS and rrnL 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-Val 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-Val 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 Ichthyosura [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 Ichthyosura (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**

1. Riley J. The biology of pentastomids. Adv Parasitol 1986; 25:45-128.
2. Yao MH, Wu F, Tang LF. Human pentastomiasis in China: case report and literature review. J Parasitol 2008; 94: 1295-1298.
3. 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.
4. 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.
5. Ye F, Sheng ZK, Li J, Sheng JF. Severe pentastomiasis in children: a report of 2 cases. Southeast Asian J Trop Med Public Health 2013; 44: 25-30.
6. Shao R, Barker SC. Mitochondrial genomes of parasitic arthropods: implications for studies of population genetics and evolution. Parasitology 2007; 134: 153-167.
7. Self JT, Kuntz RE. New Pentastomida, Sambonia parapodum n. sp.
from Varanus salvarior, and Armillifer agkistrodontis 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 analysis. 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 pentastomids: 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.
| Gene/region            | Strand | Positions | Size (bp) | No. of aa1 | Ini/Ter codons | Anticodons | ln   |
|------------------------|--------|-----------|-----------|------------|----------------|------------|------|
| tRNA-Ser<sup>UCN</sup> (S2) | +      | 1-54      | 54        |            | TGA            | 0          |      |
| tRNA-Leu (L)           | +      | 61-120    | 60        |            | GAT            | 6          |      |
| tRNA-Met (M)           | +      | 122-182   | 61        |            | CAT            | 1          |      |
| nad2                   | +      | 183-1,145 | 963       | 320        | ATT/TAG         | 0          |      |
| tRNA-Trp (W)           | +      | 1,144-1,200| 57       |            | TCA            | -2         |      |
| tRNA-Cys (C)           | -      | 1,193-1,252| 60       |            | GCA            | -8         |      |
| tRNA-Gln (G)           | -      | 1,251-1,315| 65       |            | TG             | -2         |      |
| tRNA-Tyr (Y)           | -      | 1,318-1,377| 60       |            | GTA            | 2          |      |
| cox1                   | +      | 1,379-2,903| 1,525    | 508        | CTG/T          | 1          |      |
| tRNA-Leu<sup>L1</sup> (L2) | +    | 2,904-2,966| 63       |            | TAA            | 0          |      |
| cox2                   | +      | 2,967-3,630| 664      | 221        | ATA/T          | 0          |      |
| tRNA-Asp (D)           | +      | 3,631-3,685| 55       |            | GTC            | 0          |      |
| atp8                   | +      | 3,686-3,796| 111      | 36         | ATA/TAA        | 0          |      |
| atp6                   | +      | 3,793-4,446| 654      | 218        | ATA/TAA        | 0          |      |
| cox3                   | +      | 4,446-5,226| 781      | 260        | ATG/T          | -1         |      |
| tRNA-Glu (G)           | +      | 5,227-5,282| 56       |            | TCC            | 0          |      |
| nad3                   | +      | 5,283-5,627| 345      | 114        | ATC/TAA        | 0          |      |
| tRNA-Ala (A)           | +      | 5,626-5,682| 57       |            | TGC            | -2         |      |
| tRNA-Arg (P)           | +      | 5,682-5,744| 63       |            | TCG            | -1         |      |
| tRNA-Lys (K)           | +      | 5,744-5,800| 57       |            | TTT            | -1         |      |
| tRNA-Asn (N)           | +      | 5,800-5,852| 53       |            | GTT            | -1         |      |
| tRNA-Ser<sup>AGN</sup> (S1) | + | 5,852-5,906| 57       |            | TCT            | -1         |      |
| tRNA-Glu (E)           | +      | 5,909-5,963| 55       |            | TTC            | 0          |      |
| tRNA-Phe (F)           | -      | 5,962-6,018| 57       |            | GAA            | -2         |      |
| nad5                   | -      | 6,019-7,624| 1,606    | 535        | ATG/T          | 0          |      |
| tRNA-His (H)           | -      | 7,626-7,887| 26       |            | GTG            | 1          |      |
| nad4                   | -      | 7,867-8,947| 1,261    | 420        | ATG/T          | -1         |      |
| nad4L                  | -      | 8,941-9,210| 270      | 89         | ATA/TAG        | -7         |      |
| tRNA-Thr (T)           | -      | 9,227-9,282| 56       |            | TGT            | 16         |      |
| tRNA-Pro (P)           | -      | 9,283-9,338| 56       |            | TGG            | 0          |      |
| nad6                   | +      | 9,340-9,783| 444      | 148        | ATC/TAA        | 1          |      |
| cyt b                  | +      | 9,783-10,892| 1,110   | 370        | ATG/TAA        | -1         |      |
| nad1                   | -      | 10,915-11,832| 918   | 305        | ATG/TAA        | 22         |      |
| ml                     |        | 11,833-12,976| 1,144  |            |                | 0          |      |
| tRNA-Val (V)           | -      | 12,977-13,033| 57       |            | TAC            | 0          |      |
| ms                     |        | 13,034-13,692| 659   |            |                | 0          |      |
| tRNA-Leu<sup>UUR</sup> (L2) | - | 13,690-13,752| 60      |            | TAG            | 0          |      |
| NCR                    |        | 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. *Opisthopatus cinctipes* and *Metaperipatus inae* (Onychophora) were used as outgroups.