Characterization of the mitochondrial genome of *Arge bella* sp. nov. (Hymenoptera: Argidae) (#26554)

First submission

Editor guidance

Please submit by **15 Oct 2018** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data. Download from the location described by the author.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

7 Figure file(s)
2 Latex file(s)
4 Table file(s)
5 Raw data file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

New species checks

- Have you checked our new species policies?
- Do you agree that it is a new species?
- Is it correctly described e.g. meets ICZN standard?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Characterization of the mitochondrial genome of *Arge bella* sp. nov. (Hymenoptera: Argidae)

Shiyu Du [1], **Gengyun Niu** [2], **Tommi Nyman** [3], **Meicai Wei** Corresp. [1]

[1] Central South University of Forestry and Technology, Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees (Central South University of Forestry and Technology), Ministry of Education, Changsha, Hunan, China
[2] Jiangxi Normal University, Life Science College, Nanchang, Jiangxi, China
[3] Norwegian Institute for Bioeconomy Research, Svanhovd Research Station, Svanvik, Norway

Corresponding Author: Meicai Wei
Email address: weimc@126.com

We describe *Arge bella* sp. nov., a large and beautiful species of Argidae from south China, and report its mitochondrial genome based on high-throughput sequencing data. We present the gene order, nucleotide composition of protein-coding genes (PCGs), and the secondary structures of tRNA genes. The nearly complete mitochondrial genome of *A. bella* has a length of 15,576 bp (GenBank accession NO. MF287761) and a typical set of 37 genes (22 tRNAs, 13 PCGs, and 2 rRNAs). Three tRNAs are rearranged in the *A. bella* mitochondrial genome as compared to the ancestral type in insects: *trnM* and *trnQ* are shuffled, while *trnW* is translocated from the *trnW-trnC-trnY* cluster to a location downstream of *trnI*. All PCGs are initiated by ATN codons, and terminated with TAA, TA or T as a stop codon. All tRNAs have a typical cloverleaf secondary structure, except for *trnS1*. H821 of *rrnS* and H976 of *rrnL* are redundant. A phylogenetic analysis based on mitogenome sequences of *A. bella*, 21 other symphytan species, 2 apocritan representatives, and 4 outgroup taxa support the placement of Argidae as sister to the Pergidae within the symphytan superfamily Tenthredinoidea.

Characterization of the mitochondrial genome of *Arge bella* sp. nov. (Hymenoptera: Argidae)

Shiyu Du[1] *, Gengyun Niu[2] *, Tommi Nyman[3], and Meicai Wei[1]

[1]**Central South University of Forestry and Technology, Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees (Central South University of Forestry and Technology), Ministry of Education, Changsha, Hunan, China**
[2]**Jiangxi Normal University, Life Science College, Nanchang, Jiangxi, China**
[3]**Norwegian Institute for Bioeconomy Research, Svanhovd Research Station, Svanvik, Norway**
[*]**These authors contributed equally to this work.**

Corresponding author:
Meicai Wei[1]

Email address: weimc@126.com

ABSTRACT

We describe *Arge bella* sp. nov., a large and beautiful species of Argidae from south China, and report its mitochondrial genome based on high-throughput sequencing data. We present the gene order, nucleotide composition of protein-coding genes (PCGs), and the secondary structures of tRNA genes. The nearly complete mitochondrial genome of *A. bella* has a length of 15,576 bp (GenBank accession NO. MF287761) and a typical set of 37 genes (22 tRNAs, 13 PCGs, and 2 rRNAs). Three tRNAs are rearranged in the *A. bella* mitochondrial genome as compared to the ancestral type in insects: *trnM* and *trnQ* are shuffled, while *trnW* is translocated from the *trnW-trnC-trnY* cluster to a location downstream of *trnI*. All PCGs are initiated by ATN codons, and terminated with TAA, TA or T as a stop codon. All tRNAs have a typical cloverleaf secondary structure, except for *trnS1*. H821 of *rrnS* and H976 of *rrnL* are redundant. A phylogenetic analysis based on mitogenome sequences of *A. bella*, 21 other symphytan species, 2 apocritan representatives, and 4 outgroup taxa support the placement of Argidae as sister to the Pergidae within the symphytan superfamily Tenthredinoidea.

INTRODUCTION

With about 950 valid species in the world, Argidae is the second-largest family of the paraphyletic suborder Symphyta of the order Hymenoptera (Choi et al., 2016). Eastern Asia is one of the three main centers of diversity of the family (Wei and Nie, 1997). Within China, about 170 species and 16 genera have been recorded (Choi et al., 2016). However, there are probably many more species to be described or newly recorded from this vast country.

Symphyta is the predominantly herbivorous and relatively less diverse suborder of the Hymenoptera, and contains more than 8500 described species (Taeger et al., 2010). The systematic arrangement of Symphyta, including the numbers of families and superfamilies are quite uncertain. Most researchers of symphytan systematics have divided the extant Symphyta into 14 families (Benson, 1938; Königsmann, 1977; Abe and Smith, 1991; Taeger et al., 2010) under four (Benson, 1938), six (Königsmann, 1977; Abe and Smith, 1991) or seven superfamilies (Taeger et al., 2010), and the six superfamilies in the system of Königsmann and of Abe and Smith are also different. Rasnitsyn (1988) divided extant taxa of Symphyta into two suborders, five infraorders, and 13 families. Wei and Nie (1997, 1998) divided the original Symphyta into five suborders, eleven superfamilies, and 20 families. It seems that a consensus on the systematics of Symphyta among sawfly researchers is difficult to get just based on morphological analysis, so using molecular-genetic data would be critical to test the different proposed systems and to approach a natural system of Symphyta.

The monophyly of Tenthredinoidea is supported by both morphological (Wei and Nie, 1997) and molecular data (Malm and Nyman, 2015) as well as combined analyses (Ronquist et al., 2012a; Sharkey et al., 2012; Klopfstein et al., 2013), but relationships among core tenthredinoids are less clear. Argidae was inferred as the sister to the remaining tenthredinoids by Wei and Nie (1997), but the disaccord of this result with several recent studies may follow from the limited dataset. Malm and Nyman (2015) analyzed nine protein-coding genes of 164 taxa to reconstruct the phylogenetic backbone of the Hymenoptera. In their analysis, 13 taxa were included to represent five out of seven subfamilies within Argidae; in the tree, Argidae and Pergidae form a monophylum as sister to the other non-blasticotomid tenthredinoids, which supports more comprehensive morphological (Vilhelmsen, 1997; Schmidt et al., 2006) and combined studies (Ronquist et al., 2012a; Sharkey et al., 2012; Klopfstein et al., 2013). A recent analysis of whole-body transcriptomes also inferred the monophylum of Argidae and Pergidae (Peters et al., 2017).

The mitochondrial genomes of 21 symphytan species have been reported (Table 1; data were collected at NCBI, available at https://www.ncbi.nlm.nih.gov/; accessed 3 Nov. 2017). Five phylogenetic analyses have been conducted based on nucleotide sequences of symphytan mitogenomes (Castro and Dowton, 2005; Dowton et al., 2009b; Song et al., 2015b, 2016; Doğan and Korkmaz, 2017), but none of them have provided clear insights into symphytan relationships because of the taxonomically restricted representation of sawfly families in the datasets: the mitochondrial genomes of Argidae, Xyelidae, Diprionidae, ~~Heptameliidae~~, Blasticotomidae, Megalodontesidae, Pamphiliidae, Xiphydriidae, Siricidae, and Anaxyelidae ~~(Wei and Nie, 1998)~~ have not been previously reported.

The small number of available sawfly mitochondrial genomes also limits our understanding of their genomic architecture. The conservation of rRNA secondary structures exceeds that of its nucleotides and, therefore, it is recommended that secondary structures guide decisions about the alignment of homologous positions for phylogenetic studies (Kjer, 1995). However, inferred secondary structures can only be considered as working hypotheses, and would be almost impossible to estimate without using a comparative approach (Misof and Fleck, 2003). Our understanding of the secondary structures of symphytan rRNAs has been developed only from seven Cephidae species (Dowton et al., 2009a; Korkmaz et al., 2015, 2016, 2017). More representatives and comparative analyses are therefore required within the Symphyta.

Here, we describe a large and beautiful new species of *Arge* Schrank (1802) and report its near-complete mitochondrial genome sequence, as the first representative of the family Argidae. We also characterize the nucleotide composition, codon usage, gene rearrangements and secondary structure of RNAs. The structural differences between *A. bella* and *Cephus* species are described to establish structural features as potentially useful characters for symphytan systematics. Finally, we report the results of phylogenetic analyses that we used to verify the phylogenetic placement of *A. bella* based on sequences of 13 protein-coding genes and 2 rRNA genes of ~~21~~ species of Symphyta, 2 representatives of Apocrita, and four outgroup taxa.

MATERIALS AND METHODS

Description of new species

Specimens were examined with a Leica S8APO dissection microscope. Adult images were taken with a Nikon D700 digital camera, and sequentially focused images were montaged using Helicon Focus (HeliconSoft), while detailed images were taken with Leica Z16 APO/DFC550. All images were further processed with Adobe Photoshop CS 6.0.

The terminology of sawfly genitalia follows Ross (1945), and that of general morphology follows Viitasaari (2002). Abbreviations used are: OOL = distance between the eye and outer edge of lateral ocellus; POL = distance between the mesal edges of the lateral ocelli; OCL = distance between a lateral ocellus and the occipital carina or hind margin of the head.

The holotype and all paratypes of the new species are deposited in the Insect Collection of Central South University of Forestry and Technology, Changsha, Hunan, China (CSCS).

All nomenclatural acts, authors and literature are registered in Zoobank as per the recent proposed amendment to the International Code of Zoological nomenclature for a universal register for animal names (Polaszek et al., 2005a,b; Pyle et al., 2008; ICZN, 2008). Rules for spelling Chinese personal and place names follow GB/T 16159-1996 and ISO 7098: 1991: “Chinese people’s names are to be written separately with the surname first, followed by the personal name written as one word, with the initial letters of both capitalized.”. “Chinese place names should be alphabetized according to the “Spelling

Rules for Chinese Geographical Place Names," document no. 17 (1984) of the State Committee on Chinese Geographical Place Names. Separate the geographical proper name from the geographical feature name and capitalize the first letter of both" (Niu et al., 2012).

The electronic version of this article in Portable Document Format (PDF) will represent a published work according to the International Commission on Zoological Nomenclature (ICZN), and hence the new names contained in the electronic version are effectively published under that Code from the electronic edition alone. This published work and the nomenclatural acts it contains have been registered in ZooBank, the online registration system for the ICZN. The ZooBank LSIDs (Life Science Identifiers) can be resolved and the associated information viewed through any standard web browser by appending the LSID to the prefix http://zoobank.org/. The LSID for this publication is: urn:lsid:zoobank.org:pub:A94BD62A-D4BE-40F9-8718-84F425875C7C. The online version of this work is archived and available from the following digital repositories: PeerJ, PubMed Central and CLOCKSS.

Library construction and sequencing

Total genomic DNA of a single specimen was used for library preparation with insert size of 250 bp following the manufacturer's instructions, and then 150 bp PE sequenced on an Illumina HiSeq 4000 platform for around 2.5 Gb of data at BGI-Shenzhen, China.

Mitochondrial genome assembly

Pre-analysis data filtering included: (i) Clean data was generated following published protocols (Zhou et al., 2013; Tang et al., 2014, 2015), by removing reads with adaptor contamination, >10% low-quality bases, or >5 bp Ns; (ii) Clean data was then compared with reference mitochondrial genomes downloaded from GenBank (716 RefSeq genomes, including 699 arthropods, seven starfish and 10 cyprinid fish; accessed on 10 March 2014) to screen out candidate mitochondrial reads using relaxed criteria: BLAST identity >30% and *E*-value <0.00001; (iii) A 51-mer set was then generated from these candidate mito-reads and used as reference for a second round of data filtering for the discarded reads from step 2; (iv) *De novo* assembly was performed by *SOAPdenovo-Trans* (Xie et al., 2014) (-K 71, -L 100, -t 1), *SOAPdenovo* 2.0 (Luo et al., 2012; Li et al., 2010) (-K 61, -k 45), *IDBA-UD* (Peng et al., 2012) (kMaxShortSequence = 256, –num threads 12), and mitochondrial protein-encoding assemblies and mitochondrial genome-sequence candidates were annotated by a custom *Perl* script (Zhou et al., 2013) using RefSeq mitochondrial genomes of target animal taxa (604 arthropod species, two asteriid starfish and the zebrafish; downloaded from GenBank on 13 June 2013) downloaded from NCBI as reference. The mitochondrial genome was constructed, corrected and manually checked as previously described (Tang et al., 2015).

Mitochondrial genome annotation and secondary structure prediction

All of the tRNAs were identified using MITOS (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt et al., 2013) using the default settings. The initiation and termination codons of PCGs were determined in Geneious v8.0.5 (Kearse et al., 2012) (available from http://www.geneious.com) using reference sequences from other symphytan species, and then checked manually. Secondary structures of rRNAs were inferred using alignment to the models predicted for *Trichiosoma anthracinum* and *Labriocimbex sinicus* (Yan et al., in press). The predicted secondary structures of tRNAs and rRNAs were drawn using VARNA v3-93 (Darty et al., 2009) and RNAviz 2.0.3 (De Rijk et al., 2003). Helix numbering follows the convention established at the CRW site (Cannone et al., 2002) and *Apis mellifera* rRNA secondary structure (Gillespie et al., 2006) with minor modifications.

The A + T content of nucleotide sequences and relative synonymous codon usage (RSCU) were calculated using MEGA v7.0 (Kumar et al., 2016) (Table 2). Strand asymmetry was calculated using the formulae by Perna and Kocher (1995): GC–skew = (G – C) / (G + C) and AT–skew = (A – T) / (A + T).

Phylogenetic analyses

Phylogenetic analyses was performed based on aligned sequences of the 13 PCGs and 2 rRNAs of the near-complete mitochondrial genome of *A. bella* and 21 other symphytan mitochondrial genomes downloaded from GenBank (Table 1). These additional taxa represented 5 families: Tenthredinidae (Wei et al., 2013, 2014; Song et al., 2015a, 2016), Cimbicidae (Song et al., 2016; Doğan and Korkmaz, 2017), Pergidae (Castro and Dowton, 2005), Orussidae (Dowton et al., 2009a), and Cephidae (Dowton et al., 2009a; Korkmaz et al., 2015, 2016, 2017). As the Symphyta is paraphyletic with respect to the

suborder Apocrita, we also included the mitochondrial genomes of the apocritan species *Parapolybia crocea* (GenBank: KY679828) and *Taeniogonalos taihorina* (GenBank: NC027830) in the analysis. As outgroups, we included *Neopanorpa pulchra* (GenBank: FJ169955) from Mecoptera, *Anopheles gambiae* (GenBank: L20934) from Diptera, *Neochauliodes parasparsus* (GenBank: KX821680) from Megaloptera and *Paroster microsturtensis* (GenBank: MG912997) from Coleoptera.

Nucleotide sequences of the 13 PCGs from the mitochondrial genomes of *A. bella* and the 27 other included species were translated into amino acid sequences and then aligned by MUSCLE (Codons) in MEGA v7.0. Nucleotide sequences of 2 rRNAs from the included taxa were aligned by MAFFT (https://www.ebi.ac.uk/Tools/msa/mafft/). The nucleotide sequence alignments of the 13 PCGs and 2 rRNAs were concatenated using SequenceMatrix v1.7.8 (Vaidya et al., 2011) and used in phylogenetic analyses under the Maximum-likelihood (ML) criteron and Bayesian inference (BI). Partition schemes and substitution models were calculated simultaneously in PartitionFinder 1.1.1 (Lanfear et al., 2012). The branch lengths and search strategy of schemes were set as linked and greedy, and models were selected based on AICc and BIC. The GTR+I+G model was chosen as the best-fitting model for all partitions for both ML and BI analyses.

The ML analysis was performed using the IQ-TREE web server (http://iqtree.cibiv.univie.ac.at/) (Trifinopoulos et al., 2016), using default parameters except for 0.1 as the perturbation strength and 1000 as the IQ-TREE stopping rule. The BI analysis was performed using MrBayes v3.2.6 (Ronquist et al., 2012b) on the CIPRES Science Gateway (Miller et al., 2010). Rate and substitution-model parameters were unlinked across partitions. Two independent runs with four simultaneous Markov chains (one cold, three incrementally heated at T = 0.2) were run for five million generations, with sampling of parameters and trees occurring every 1000 generations. The Maximum clade credibility tree showing all compatible groupings was calculated with a burn-in fraction of 10%, after confirming in Tracer v1.6.0 (Rambaut and Drummond, 2013) (Available at:http://beast.bio.ed.ac.uk/Tracer) that both runs had converged and that appropriate effective sample sizes were achieved for sampled parameters. Trees were edited in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS AND DISCUSSION

Arge bella Wei & Du sp. nov.

urn:lsid:zoobank.org:act:9AF00C3F-D5EE-474B-BD9E-4794BECACA4F

Etymology. This species is named after its beautiful body colour.

Holotype. Female. China: Hunan Province, Guidong Conty, Mt. Qiyun, Hydropower Station Valley, alt. 752m, 25°45.361′ N, 113°55.598′ E, April 4, 2015, Yuchen Yan, Ting Liu leg. (CSCS).

Paratypes. 2 Females. Locality and collecting time as the holotype, Hang Zhao, Mengmeng Liu leg. (CSCS).

Distribution. China (Hunan).

Remarks. This new species is somewhat similar to *Arge nigricrux* Malaise (1943) and *A. vitalisi* Turner (1919), but differs from these two species by the followings: the antennal flagellum entirely black; dorsum of mesonotum blue black except for posterior of mesoscutellum; abdomen yellow brown, terga 1-2 and 4-7 largely blue black; wings distinctly yellowish, without transverse macula; hind tarsus entirely blue black; mesepisternum blue black with upper fourth yellow brown. In *A. nigricrux* and *A. vitalisi*, the antennal flagellum and mesonotum in female entirely yellow; abdomen yellow brown, terga 4 largely black, terga 5-8 with short middle black maculae; wings weakly yellowish, with distinct transverse smoky macula just below pterostigma; hind tarsus entirely yellow brown; mesepisternum entirely black or entirely yellow brown.

Description

Female. Body length 13 mm (Figure 1). Body and leg yellow brown; apex of mandible, apical half of pedicel and flagellum entirely black, without bluish tinge; ocellar area, entire mesoscutal middle lobe, dorsum of mesoscutal lateral lobe, anterior triangular lobe of mesoscutellum, mesepisternum except for dorsal fourth, katepimeron except for margins, small central macula on metapleuron, large transversal band on tergum 1, dorsum of terga 2 and 4-7, small middle macula on tergum 3, apical 0.8 of middle and hind femora, apex of middle tibia, apex of middle basitarsus and tarsomeres 2-5, apical 0.4 of hind tibia and entire hind tarsus black with distinct bluish tinge, upper margin of black macula on mesepisternum convex (Figure 2c); most of body hairs yellow brown, hairs on flagellum, middle and hind femora, spines

on inner sides of sheath mostly black, hairs on mesonotum dark brown. Wings hyaline, with distinct yellowish tinge, veins C, r1 and A largely pale brown, pterostigma and other veins black brown.

Head smooth, shiny; basal half of mandible with large and dense punctures; clypeus, face, lower half of inner orbit and frontal wall distinctly and rather densely punctured, temple densely and minutely punctured, postocellar area hardly punctured, postorbit minutely and sparsely punctured, frontal basin smooth; pronotum finely and faintly punctured; mesonotum minutely and sparsely punctured, most of mesoscutellum smooth with some minute punctures, parapsis smooth, impunctate and without microsculptures; metanotum and mesepisternum smooth, hardly punctured; anepimeron minutely punctured, katepimeron finely microsculptured; center of mesosternum distinctly punctured; metapleuron impunctate, without microsculptures; dorsum of abdominal terga smooth, impunctate, terga 1-2 faintly microsculptured, sterna and sheath smooth.

Labrum about 2 times as broad as long, apex obtusely truncate; clypeus flat, anterior incision shallow and roundish; supraclypeal area roundly elevated, without middle ridge; lateral carinae between antennal toruli low and obtuse, almost parallel downwards, not merged together, largest breadth between lateral carinae about 1.5 times diameter of median ocellus (Figure 2a); middle fovea round, distinct, shallowly open to frontal basin; frons small, center evenly concave, frontal wall distinct; malar space 0.6 times diameter of median ocellus; inner margin of eyes parallel, distance between eyes 1.1 times longest axis of eye; POL: OOL: OCL = 20: 27: 22; postocellar area flat, breadth 1.7 times length; lateral furrows shallow and fine, slightly convergent backwards; postocellar furrow fine and shallow, interocellar furrow broad and shallow; in dorsal view (Figure 2d), head about 0.65 times as long as eye, hardly enlarged behind eyes; head in lateral view as in Figure 2i. Antenna slightly enlarged toward apex, hardly bent, longitudinal carina low and faint, pedicel as long as broad, flagellum 0.8 times as long as thorax and about 1.35 times head breadth (Figure 2i). Middle furrow on mesoscutal middle lobe faint, notauli distinct; mesoscutellum 1.2 times as broad as long, anterior 0.2 with a short middle furrow; distance between cenchri 0.2 times breadth of a cenchrus. Forewing: vein R short, about 0.7 times as long as vein and 0.6 times as long as free abscissa of vein Sc, vein R+M about half length of vein 1r-m, second abscissa of vein Rs clearly longer than third abscissa of Rs, third abscissa of Rs 1.4 times as long as fourth abscissa of Rs, cell 1Rs clearly longer than cell 2Rs, upper and lower margins of cell 2Rs equal in length; vein 2Rs roundly convex outwards, vein cu-a meeting cell 1M at basal 0.4, basal anal cell closed. Hind wing: cell Rs 1.25 times as long as cell M, cell M about 2.1 times as long as broad; anal cell closed; relative length of cells A, petiole of anal cell and cu-a about 80: 57: 27; outer margin of fore and hind wings naked (Figure 1). Middle and hind tibiae each with 1 preapical spur (Figure 2j); hind basitarsus slightly longer than following 3 tarsomeres together. Ovipositor sheath as long as hind femur, basal third distinctly concave in lateral view (Figure 2f); apex of sheath round in dorsal view (Figure 2m); subapical part of lance weakly narrowed (Figure 2l); lancet broad, annular spines very short, serrulae strongly protruding (Figure 2g), basal and middle serrulae as in Figure 2g.

Male. Unknown.

Architecture and nucleotide composition of *Arge bella* mitochondrial genome

We sequenced the nearly complete mitochondrial genome of *A. bella*, ~~and the mitochondrial genome sequence of *A. bella* have been~~ deposited in GenBank of NCBI under accession number MF287761. The sequenced region is 15,576 bp in length, with 13 protein-coding genes, two rRNA genes and 22 tRNA genes. Of these, 23 genes (9 PCGs and 14 tRNAs) were encoded by the J strand, while the remaining ones were encoded by the N strand (Table 3). We did not succeed in sequencing a fragment spanning the A + T-rich region and genes flanking the A + T-rich region, due to the extremely high A + T content and presence of highly repetitive sites.

Compared with the ancestral gene arrangement of insects (Figure 3), the mitochondrial genome of *A. bella* exhibited only few rearrangements: *trnM* and *trnQ* have swapped positions, and *trnW* has been translocated from the *trnW-trnC-trnY* cluster to downstream of *trnI* (Figure 3). No inversions were found.

There were totally 23 overlapping nucleotides between neighboring genes in 6 locations, and the length of the overlapping sequences ranged from 1 to 7 bp (Table 3). Some overlapping nucleotides were conserved in the *A. bella* mitochondrial genome: ATGATAA for *ATP8* and *ATP6*, and ATGTTAA for *ND4* and *ND4L*, which is also a common feature of many other insect mitochondrial genomes (Chai et al., 2012). There were totally 192 non-coding positions between neighboring genes in 15 locations, and the length of non-coding sequences ranged from 1 to 44 bp (Table 3). There were 5 locations where the length

was over 15 bp: 33 bp between *trnW* and *trnI*, 30 bp between *trnI* and *trnM*, 25 bp between *trnR* and *trnN*, and 44 bp between *COX2* and *trnK*.

Protein-coding genes and codon usage

All PCGs were initiated by ATN codons: four genes (*ND2*, *ATP8*, *ND3* and *ND6*) used ATA as start codon, four genes (*COX1*, *ND5*, *ND4L* and *ND1*) started with ATT, and five genes (*COX2*, *ATP8*, *COX3*, *ND4* and *CYTB*) were initiated with ATG (Table 3).

The stop codons of *A. bella* were generally TAA, except for *ND6*, which ended with TA, and *COX1*, *COX3* and *ND4*, which ended with T (Table 3). Incomplete stop codons have been reported for all symphytan mitochondrial genomes sequenced to date.

The nucleotide composition of the mitochondrial genome of *A. bella* was A and T rich, with an 80.7% A + T content (Table 4). In PCGs, the highest A + T content was observed in the third codon position, the highest T content in the second codon position and the lowest G content in the third codon position. The highest A + T content was observed in the *ATP8* gene (89.7%).

It has been reported that the parental N strand remains as a single strand for a longer time during replication of mitochondrial genomes, resulting in deamination of A and C (Reyes et al., 1998). This leads to an A- and C-skew on the J strand and a T- and G-skew on the N strand. In the case of *A. bella*, we observed that the AT skew was slightly positive (0.0706). On the contrary, GC skew was negative (-0.2708) when considering the whole mitogenome (Table 4), which shows that the occurrence of A was higher than that of T, and the occurrence of C was higher than that of G, which is a general phenomenon in symphytan mitochondrial genomes (Castro and Dowton, 2005; Dowton et al., 2009b; Wei et al., 2015). PCGs on the J strand were slightly T-skewed (-0.0218) and slightly C-skewed (-0.2273), whereas PCGs encoded by the N strand were all slightly T-skewed (-0.2285) and moderately G-skewed (0.3226).

Codon usage in the *A. bella* mitochondrial genome is presented in Table 2. As in other insect mitogenomes (Foster et al., 1997), a significant correlation between codon usage and nucleotide composition was found. Leu, Ser, Gly and Ala were most frequently used amino acids (Table 2). UUA-Leu had the highest relative synonymous codon usage (4.83) (Table 2). A relationship between the nucleotide compositions of codon usage and amino acid occurrence was noticed. The relationship can be calculated by the ratio of G + C rich codons (Pro, Ala, Arg, and Gly) and A + T rich codons (Phe, Ile, Met, Tyr, Asn, and Lys). The ratio found in *A. bella* (0.27) is similar to or lower than that of other symphytan species (0.28–0.31) (Korkmaz et al., 2015).

Transfer RNA genes

The position and orientation of the predicted tRNAs and anticodon sequences were identical to most of the hitherto reported symphytan mitochondrial genomes (Table 3). 14 tRNAs were encoded by the J strand, and the others by the N strand. All tRNAs folded into a usual clover-leaf structure except for *trnS1* (AGN). Compared with other symphytan species, *trnS1* (AGN) lacked a dihydrouridine (DHU) arm. The size of tRNAs ranged from 64 bp (*trnS1*) to 74 bp (*trnH* and *trnN*) (Figure 4), placed well within the observed ranges in insects. The observed size differences resulted from changes in the length of the variable loop, dihydrouridine (DHU) arm and TΨC arm (Clary and Wolstenholme, 1985). Anticodon sequences of the tRNA genes were identical with previously-reported symphytan mitogenomes (Table 3). Pairing mismatches occurred mainly in the DHU arm, AA arm and AC arm, and sometimes in TΨC arm. All of the 18 mismatches were G-U pairs.

Ribosomal RNA genes

In *A. bella*, *rrnS* was located downstream of *trnV*, and *rrnL* was located between *trnL1* and *trnV* (Table 3). Both rRNAs were encoded by the N strand, and their lengths were 839 bp and 1,390 bp, and A + T contents 83.0% and 84.0%, respectively.

The *rrnS* secondary structure of *A.bella* contained four domains and 27 helices (Figure 5). Previous studies have shown that the second half of the domain III sequence can be difficult to align precisely, even when information on secondary structure is considered. The *rrnS* domain III model of *A.bella* was very similar to the one described in Hickson et al. (1996). H821 was redundant compared with other symphytan species like *Cephus* (Korkmaz et al., 2015, 2016, 2017), which includes 26 helices. The predicted structure of H921 was well conserved in symphytan species, but loop size in H47 is variable. The predicted structures of H500, H769, H944 and H1047 were conserved in symphytan species. H1399 and H1506 helices were well conserved in *A. bella*, as well as in other insect species (Cameron and

Whiting, 2008; Cannone et al., 2002; Gillespie et al., 2006; Misof and Fleck, 2003; Wei et al., 2009, 2010b).

The length of the *rrnL* gene was 1,390 bp (Table 1), with an 84.0% A + T content. The secondary structure of the *rrnL* gene in *A. bella* conformed to models proposed for other insects, with the 45 helices belonging to six domains (Figure 6) (Cameron and Whiting, 2008; Cannone et al., 2002; Gillespie et al., 2006; Misof and Fleck, 2003; Wei et al., 2009, 2010a,b). H563, H671, H1925 and H2043 were conserved, and H1775 almost with 3 pairs in symphytan species. H991 was different from *Perga condei*, *Orussus occidentalis*, and *Monocellicampa pruni* with regards to helical length and loop size.

Phylogenetic relationships

We investigated the phylogenetic position of *A. bella* by combining our new mitogenome sequences with previously-reported data from 21 species of Symphyta representing five other families (Table 1), as well as with sequences from 2 apocritan species and 4 non-hymenopteran outgroup taxa.

Both ML and BI analyses placed *A. bella* as sister to *Perga condei* with high support (Figure 7). This Pergidae + Argidae monophylum, as well as its placement as sister to the remaining non-blasticotomid tenthredinoids, corresponds to results from comprehensive morphological (Vilhelmsen, 1997; Schmidt et al., 2006) and molecular (Malm and Nyman, 2015) studies. The phylogenetic location of *A. bella* can be considered to constitute the basal branch of the normal Tenthredinoidea (Ross, 1937; Benson, 1938; Taeger et al., 2010), or the suborder Tenthredinomorpha (Wei and Nie, 1998). On a wider phylogenetic scale, our results supported a grouping of ((Cephidae, (Orussidae, Apocrita)), ((Argidae, Pergidae), (Cimbicidae, Tenthredinidae))) within the Hymenoptera.

CONCLUSIONS

Arge bella is a new species belonging to *A. vitalisi* group. It is similar to *A. nigricrux* Malaise (1943) and *A. vitalisi* Turner (1919) from south Asia, but differs from them by the antennal flagellum entirely black, the dorsum of mesonotum mainly blue black, and the abdominal tergites 1, 3-6 each with a large and broad bluish black macula.

The nearly complete mitochondrial genome of *A. bella* (15,576 bp) displays a highly conserved structure and composition as compared to the mitogenomes of other symphytans as well as insects in general. The main differences include minor rearrangements or translocations of three tRNAs, a non-clover-leaf-like structure of trnS1 (AGN), and redundancy of H821 of rrnS and H976 of rrnL. ML and BI phylogenetic analyses based on aligned mitogenome sequences of *A. bella*, 21 other symphytan taxa, 2 apocritan representatives and 4 outgroups resulted in a hymenopteran tree with the structure ((Cephidae, (Orussidae, Apocrita)), ((Argidae, Pergidae), (Cimbicidae, Tenthredinidae))) with high nodal supports. Hence, mitogenome sequencing of additional symphytan taxa in the future can clearly produce useful data for resolving hymenopteran relationships.

ACKNOWLEDGMENTS

Thanks are due to Ms Wei Songyun of Western Washington University for proofreading of the English. Thanks are due to Dr. Marko Prous of Senckenberg Deutsches Entomologisches Institut for making valuable comments on a previous version of the manuscript. Thanks are also due to Ms Yan Yuchen, Ms Liu Ting, Ms Zhao Hang and Ms Liu Mengmeng of Central South University of Forestry and Technology for collecting the interesting specimens, and to Ms Wu Xiaotong and Ms Zhang Yaoyao for their kind help in this research. Thanks are also due to Ms Tang Min of China Agricultural University for her kind help.

REFERENCES

Abe, M. and Smith, D. R. (1991). The genus-group names of Symphyta (Hymenoptera) and their type species. *Esakia, Fukuoka*, 31:1–115.

Benson, R. B. (1938). On the Classification of Sawflies (Hymenoptera Symphyta). *The Transactions of the Royal Entomological Society of London*, 87(15):353–384.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, 69(2):313–319.

Cameron, S. L. and Whiting, M. F. (2008). The complete mitochondrial genome of the tobacco hornworm, Manduca sexta, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene*, 408(1-2):112–123.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC bioinformatics*, 3:1–31.

Castro, L. R. and Dowton, M. (2005). The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of Perga condei (Hymenoptera: Symphyta: Pergidae). *Molecular Phylogenetics and Evolution*, 34(3):469–479.

Chai, H. N., Du, Y. Z., and Zhai, B. P. (2012). Characterization of the Complete Mitochondrial Genomes of Cnaphalo- crocis medinalis and Chilo suppressalis (Lepidoptera: Pyralidae). *International Journal of Biological Sciences*, 8(4):561–579.

Choi, J. K., Lee, S. B., Wei, M. C., and Lee, J. W. (2016). Three new species of the genus Arge (Hymenoptera: Symphyta: Argidae) from South Korea with keys to genera and species of the family Argidae. *Journal of Asia-Pacific Biodiversity*, pages 183–193.

Clary, D. O. and Wolstenholme, D. R. (1985). The mitochondrial DNA molecule of Drosophila yakuba: Nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, 22:252–271.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA : Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, 25(15):1974–1975.

De Rijk, P., Wuyts, J., and Wachter, R. D. (2003). RnaViz 2 : An improved representation of RNA secondary structure. *Bioinformatics*, 19(2):299–300.

Dowton, M., Cameron, S. L., Austin, A. D., and Whiting, M. F. (2009a). Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the hymenoptera–a lineage with both rapidly and slowly evolving mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 52(2):512.

Dowton, M., Cameron, S. L., Dowavic, J. I., Austin, A. D., and Whiting, M. F. (2009b). Characterization of 67 mitochondrial tRNA gene rearrangements in the hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution*, 26(7):1607–1617.

Doğan, Ö. and Korkmaz, E. M. (2017). Nearly complete mitogenome of hairy sawfly, Corynis lateralis (Brullë, 1832) (Hymenoptera: Cimbicidae): rearrangements in the IQM and ARNS1EF gene clusters. *Genetica*, pages 1–10.

Foster, P. G., Jermiin, L. S., and Hickey, D. A. (1997). Nucleotide Composition Bias Affects Amino Acid Content in Proteins Coded by Animal Mitochondria. *Journal of Molecular Evolution*, 44:282–288.

Gillespie, J. J., Johnston, J. S., Cannone, J. J., and Gutell, R. R. (2006). Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of Apis mellifera (Insecta: Hymenoptera): Structure, organization, and retrotransposable elements. *Insect Molecular Biology*, 15(5):657–686.

Hickson, R. E., Simon, C., Cooper, A., Spicer, G. S., Sullivan, J., and Penny, D. (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Molecular Biology and Evolution*, 13(1):150–169.

ICZN (2008). Proposed amendment of the International Code of Zoological Nomenclature to expand and refine methods of publication. *Zootaxa*, 1908:57–67.

Kearse, M., Moir, R., Wilson, A., Stoneshavas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., and Duran, C. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649.

Kjer, K. M. (1995). Use of rRNA Secondary Structure in Phylogenetic Studies to Identify Homologous Positions: An Example of Alignment and Data Presentation from the Frogs. *Molecular Phylogenetics and Evolution*, 4:314–330.

Klopfstein, S., Vilhelmsen, L., Heraty, J. M., Sharkey, M., and Ronquist, F. (2013). The Hymenopteran Tree of Life: Evidence from Protein-Coding Genes and Objectively Aligned Ribosomal Data). *PLoS ONE*, 8(8):e69344.

Königsmann, E. (1977). Das phylogenetische System der Hymenoptera. Teil 2: "Symphyta". *Deutsche entomologische Zeitschrift, Neue Folge*, 24(1-3):1–40.

Korkmaz, E. M., Aydemir, H. B., Temel, B., Budak, M., and Başıbüyük, H. H. (2017). Mitogenome

evolution in Cephini (Hymenoptera: Cephidae): Evidence for parallel adaptive evolution. *Biochemical Systematics and Ecology*, 71(4):137–146.

Korkmaz, E. M., Budak, M., Ördek, M. N., and Başibüyük, H. H. (2016). The complete mitogenomes of Calameuta filiformis (Eversmann, 1847) and Calameuta idolon (Rossi, 1794) (Hymenoptera: Cephidae): The remarkable features of the elongated A+T rich region in Cephini. *Gene*, 576(1):404–411.

Korkmaz, E. M., Doğan, Ö., Budak, M., and Başibüyük, H. H. (2015). Two nearly complete mitogenomes of wheat stem borers, cephus pygmeus (l.) and cephus sareptanus dovnar-zapolskij (hymenoptera: Cephidae): An unusual elongation of rrns gene. *Gene*, 558(2):254–264.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7):1870–1874.

Lanfear, R., Calcott, B., Ho, S. S., and Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analysis. *Molecular Biology and Evolution*, 29:1695–1701.

Li, R. Q., Zhu, H. M., Ruan, J., Qian, W. B., Fang, X. D., Shi, Z. B., Li, Y. R., Li, S. T., Shan, G., and Kristiansen, K. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, 20(2):265–272.

Luo, R. B., Liu, B. H., Xie, Y. L., Li, Z. Y., Huang, W. H., Yuan, J. Y., He, G. Z., Chen, Y. X., Pan, Q., Liu, J. B., Wu, G. X., Zhang, H., Shi, Y. J., Liu, Y., Yu, C., Wang, B., Yao, L., Han, C. L., Cheung, D. W., Yiu, S. M., Peng, S. L., Zhu, X. Q., Liu, G. M., Liao, X. K., Li, Y. R., Yang, H. M., Wang, J., Lam, T. W., and Wang, J. (2012). SOAPdenovo2: an empirically improvedmemory-efficient short-read de novo assembler. *GigaScience*, 1:18.

Malaise, R. (1943). Neue exotische Blattwespen. - Folium Entomolocicum. *Festschrift zum*, 60:8–13.

Malm, T. and Nyman, T. (2015). Phylogeny of the symphytan grade of Hymenoptera: new pieces into the old jigsaw (fly) puzzle. *Cladistics*, 31:1–17.

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). Creating the cipres science gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop*, 14:1–8.

Misof, B. and Fleck, G. (2003). Comparative analysis of mt LSU rRNA secondary structures of Odonates: Structural variability and phylogenetic signal. *Insect Molecular Biology*, 12(6):535–547.

Niu, G. Y., Wei, M. C., and Taeger, A. (2012). Revision of the Siobla metallica group (Hymenoptera: Tenthredinidae). *Zootaxa*, 3196:1–49.

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). Idba-ud: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11):1420–1428.

Perna, N. T. and Kocher, T. D. (1995). Unequal Base Frequencies and the Estimation of Substitution Rates. *Molecular Biology and Evolution*, 12(2):359–361.

Peters, R. S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P. A., Heraty, J., Kjer, K. M., Klopfstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S. l., Zhou, X., Wappler, T., Rust, J., Misof, B., and Niehuis, O. (2017). Evolutionary History of the Hymenoptera. *Current Biology*, 27:1–6.

Polaszek, A., Agosti, D., Alonso-Zarazaga, M., Beccaloni, G., de Place Bjørn, P., Bouchet, P., Brothers, D. J., Earl of Cranbrook, Evenhuis, N. L., Godfray, H. C. J., Johnson, N. F., Krell, F. T., Lipscomb, D., Lyal, C. H. C., Mace, G. M., Mawatari, S. F., Miller, S. E., Minelli, A., Morris, S., Ng, P. K. L., Patterson, D. J., Pyle, R. L., Robinson, N., Rogo, L., Taverne, J., Thompson, F. C., van Tol, J., Wheeler, Q. D., and Wilson, E. O. (2005a). Commentary: A universal register for animal names. *nature*, 437:447.

Polaszek, A., Alonso-Zarazaga, M., Bouchet, P., Brothers, D. J., Evenhuis, N. L., Krell, F. T., Lyal, C. H. C., Minelli, A., Pyle, R. L., Robinson, N., Thompson, F. C., and van Tol, J. (2005b). Five new species of the damselfish genus Chromis (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. *Bulletin of Zoological Nomenclature*, 62:210–220.

Pyle, R. L., Earle, J. L., and Greene, B. D. (2008). Five new species of the damselfish genus Chromis (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. *Zootaxa*, 1671:3–31.

Rambaut, A. and Drummond, A. (2013). Tracer v1.6.0. available at:http://beast.bio.ed.ac.uk/tracer.

Rasnitsyn, A. P. (1988). Sepulki i proischozhdenie steblevjch pilishtshikov (Hymenoptera: Cephidae, Sepulcidae). *Trudy vsesojuznogo entomologitscheskogo obschtschestva*, 70:68–73.

Reyes, A., Gissi, C., Pesole, G., and Saccone, C. (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Molecular Biology and Evolution*, pages 957–966.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmester, S., Murray, D. L., and Rasnitsyn, A. P. (2012a). A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera. *Systematic Biology*, 61(6):973–999.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012b). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61:539–542.

Ross, H. H. (1937). A Generic Classification of the Nearctic Sawflies (Hymenoptera, Symphyta). *Illinois biological monographs*, 15(2):1–173.

Ross, H. H. (1945). Sawfly genitalia: terminology and study techniques. *Entomological News*, 61(10):261–268.

Schmidt, S., Walter, G. H., Grigg, J., and Moore, C. J. (2006). Sexual Communication and Host Plant Associations of Australian Pergid Sawflies (Hymenoptera: Symphyta: Pergidae). *Recent Sawfly Research: Synthesis and Prospects*, pages 173–194.

Schrank, F. v. P. (1802). Fauna boica: Durchgedachte geschichte der in baiern einheimischen und zahmen thiere. zweiter band. zweite abtheilung. *Ingolstadt, bey Johann Wilhelm Krüll*, 2(2):412pp.

Sharkey, M. J., Carpenter, J. M., Vilhelmsen, L., Heraty, J., Dowling, A. P. G., Schulmeister, S., Murray, D., Andrew, R., Ronquist, F., Krogmann, L., and Wheeler, W. C. (2012). Phylogenetic relationships among superfamilies of Hymenoptera Michael. *Cladistics*, 28:80–112.

Song, S. N., Tang, P., Wei, S. J., and Chen, X. X. (2016). Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. *Scientific Reports*, 6:1–11.

Song, S. N., Wang, Z. H., Li, Y., Wei, S. J., and Chen, X. X. (2015a). The mitochondrial genome of tenthredo tienmushana (takeuchi) and a related phylogenetic analysis of the sawflies (insecta: Hymenoptera). *Mitochondrial Dna Part A Dna Mapping Sequencing and Analysis*, 27(4):2860.

Song, S. N., Wang, Z. H., Li, Y., Wei, S. J., and Chen, X. X. (2015b). The mitochondrial genome of Tenthredo tienmushana (Takeuchi) and a related phylogenetic analysis of the sawflies (Insecta: Hymenoptera). *Mitochondrial DNA*, 27(4):2860.

Taeger, A., Blank, S. M., and Liston, A. D. (2010). World Catalog of Symphyta (Hymenoptera). *Zootaxa*, 2580:1–1064.

Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E. D., Wang, J., Yang, C., Bruce, C., Nevard, T., Potts, S. G., Zhou, X., and Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution*, 6:1034–1043.

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., Liu, S., Song, W., Li, Y., Wu, Q., Zhang, A., and Zhou, X. (2014). Ultra-deep sequencing enables high-fielity recovery of biodiversity for bulk arthropod samples without PCR amplifiation. *Nucleic Acids Research*, 42:e166.

Trifinopoulos, J., Nguyen, L. T., Haeseler, A. V., and Minh, B. Q. (2016). W-iq-tree: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(Web Server issue):W232–W235.

Turner, R. E. (1919). On Indo-Chinese Hymenoptera collected by R. Vitalis de Salvaza II. *The Annals and Magazine of Natural History, including Zoology, Botany, and Geology; Ninth Series, London*, 3:483–487.

Vaidya, G., Lohman, D. J., and Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, 27:171–180.

Viitasaari, M. (2002). *Sawflies (Hymenoptera, Symphyta) I. A review of the suborder, the Western Palaearctic taxa of Xyeloidea and Pamphilioidea*. Helsinki Tremex Press Ltd.

Vilhelmsen, L. (1997). The phylogeny of lower Hymenoptera (Insecta), with a summary of the early evolutionary history of the order. *Journal of Zoological Systematics and Evolutionary Research*, 35:49–70.

Wei, M. C. and Nie, H. Y. (1997). Studies on the biogeography of Tenthredinoidea (Hymenoptera) I. Analysis of the geographical distribution of Tenthredinoidea at family level. *Entomotaxonomia*, pages 127–132.

Wei, M. C. and Nie, H. Y. (1998). Generic list of Tenthredinoidea s. str. in new systematic arrangement with synonyms and distribution data. *Journal of Central South Forestry University*, 18(3):28–31.

Wei, S. J., Niu, F. F., and Du, B. Z. (2014). Rearrangement of trnQ-trnM in the mitochondrial genome of Allantus luctifer (Smith) (Hymenoptera : Tenthredinidae) Rearrangement of trnQ-trnM in the mitochondrial genome of Allantus. *Mitochondrial DNA*, pages 1–3.

Wei, S. J., Shi, M., Chen, X. X., Sharkey, M. J., Achterberg, C. V., Ye, G. Y., and He, J. H. (2010a). New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE*, 5(9):1–10.

Wei, S. J., Shi, M., He, J. H., Sharkey, M., and Chen, X. X. (2009). The complete mitochondrial genome of Diadegma semiclausum (hymenoptera: ichneumonidae) indicates extensive independent evolutionary events. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada*, 52(4):308–319.

Wei, S. J., Shi, M., Sharkey, M. J., Achterberg, C. V., and Chen, X. X. (2010b). Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. *BMC genomics*, 11:371.

Wei, S. J., Wu, Q. L., Kees, V. A., and Chen, X. X. (2015). Rearrangement of the nad1 gene in Pristaulacus compressus (Spinola) (Hymenoptera: Evanioidea: Aulacidae) mitochondrial genome. *Mitochondrial DNA*, 26(4):629–630.

Wei, S. J., Wu, Q. L., and Liu, W. (2013). Sequencing and characterization of the monocellicampa pruni (hymenoptera: Tenthredinidae) mitochondrial genome. *Mitochondrial Dna*, 26(1):157–8.

Xie, Y. L., Wu, G. X., Tang, J. B., Luo, R. B., Patterson, J., Liu, S. L., Huang, W. H., He, G. Z., Gu, S. C., Li, S. K., Zhou, X., Lam, T. W., Li, Y. R., Xu, X., Wong, G. K.-S., and Wang, J. (2014). Soapdenovo-trans: de novo transcriptome assembly with short rna-seq reads. *Bioinformatics*, 30(12):1660–1666.

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J., and Huang, Q. (2013). Ultra-deep sequencing enables high-fielity recovery of biodiversity for bulk arthropod samples without PCR amplifiation. *Gigascience*, 2:4.

Notes

Figure 1. *Arge bella* sp. nov.

Arge bella Wei & Du sp. nov. Adult female, dorsal view. Scale bar = 1 mm.

Figure 2. *Arge bella*.

Figures a–m. *Arge bella* Wei & Du sp. nov. a. Head of female, front view; b. Head of female, dorsal view; c. Mesopleuron of female, lateral view; d. Head of female, dorsal view; e. Sheath of female, lateral view; f. Sheath of female, ventral view; g. Serrula of female; h. Antenna of female; i. Head of female, lateral view; j. Tibia of hind leg; k. Lancet; l. Lance; m. Sheath of female, dorsal view.

Figure 3. Mitochondrial genome organisation of *Arge bella*.

Mitochondrial genome organisation of insects in general (above) and *Arge bella* (below). Genes are transcribed from left to right, except for those underlined. PCGs are marked by white, the A + T-rich region by grey, rRNA genes by red, and tRNA genes by yellow and single-letter amino acid codes. Gene rearrangements are shown with connecting lines that indicate the translocation of *trnW* from the *trnW-trnC-trnY* cluster to a position downstream of *trnI*, and the swapped position of *trnM* and *trnQ*.

Figure 4. *Arge bella* tRNAs.

Predicted secondary structures of the 22 tRNA genes of *Arge bella*. Dashes indicate Watson–Crick base pairing and dots indicate G–U base pairing.

Figure 5. *Arge bella rrnS*.

The predicted secondary structure of *rrnS* in the *Arge bella* mitochondrial genome. The numbering of helices follows Gillespie et al. (2006), roman numbers refer to domain names.

Figure 6. *Arge bella rrnL*.

Predicted secondary structure of *rrnL* in the *Arge bella* mitochondrial genome. The numbering of helices follows Gillespie et al. (2006), roman numbers refer to domain names.

Figure 7. Phylogenetic tree of Symphyta and selected apocritan and outgroup taxa, based on a Maximum-likelihood analysis of sequence data from 13 PCGs and 2 rRNA genes.

The numbers above and below branches are Maximum-likelihood bootstrap values and Bayesian posterior probabilities, respectively. The scale bar indicates the number of substitutions per site.

Table 1. General information of the mitochondrial genomes of Symphyta.

Table 2. Codon usage of 13 PCGs in the mitochondrial genome of *Arge bella*.

No., frequency of each codon; RSCU, relative synonymous codon usage.

Table 3. Characteristics of the mitochondrial genome of *Arge bella*.

J, major; N, minor; IGN, intergenic nucleotides. Minus indicates overlapping sequences between adjacent genes.

Table 4. Nucleotide composition of *Arge bella* mitogenome.

Figure 1

Arge bella sp. nov.

Arge bella Wei & Du sp. nov. Adult female, dorsal view. Scale bars = 1 mm.

Figure 2

Arge bella

Figures a-m. *Arge bella* Wei & Du sp. nov. a. Head of female, front view; b. Head of female, dorsal view; c. Mesopleuron of female, lateral view; d. Head of female, dorsal view; e. Sheath of female, lateral view; f. Sheath of female, ventral view; g. Serrula of female; h. Antenna of female; i. Head of female, lateral view; j. Tabia of hind leg; k. Lancet; l. Lance; m. Sheath of female, dorsal view.

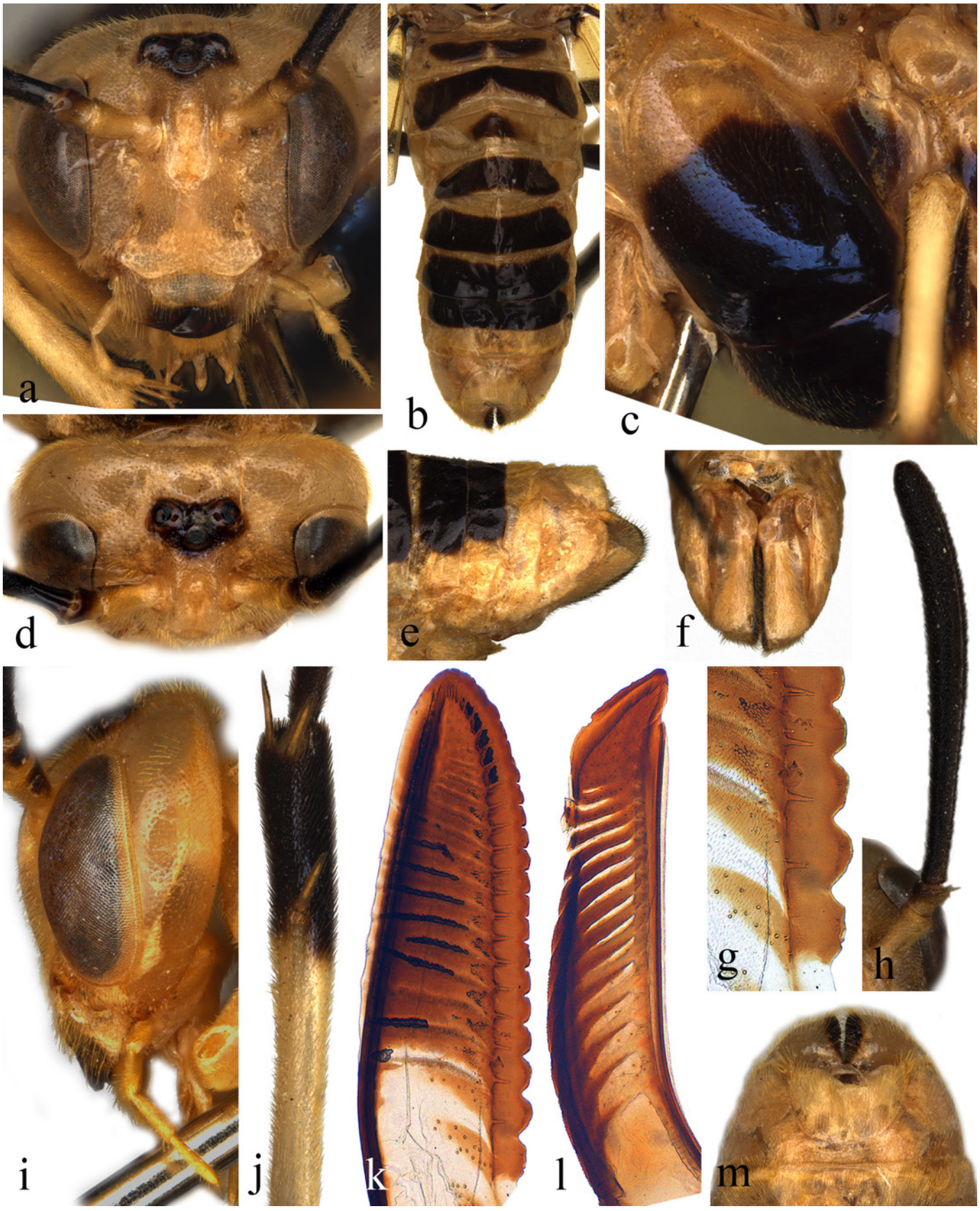
a
b
c
d
e
f
g
h
i
j
k
l
m

Figure 3

Mitochondrial genome organisation of *Arge bella*

Mitochondrial genome organisation of insects in general (above) and *Arge bella* (below). Genes are transcribed from left to right, except for those underlined. PCGs are marked by white, the A + T rich region by grey, rRNA genes by red, and tRNA genes by yellow and single-letter amino acid codes. Gene rearrangements are shown with connecting lines that indicate the translocation of *trnW* from the *trnW-trnC-trnY* cluster to a position downstream of trnI and the swapped position of *trnM* and *trnQ*.

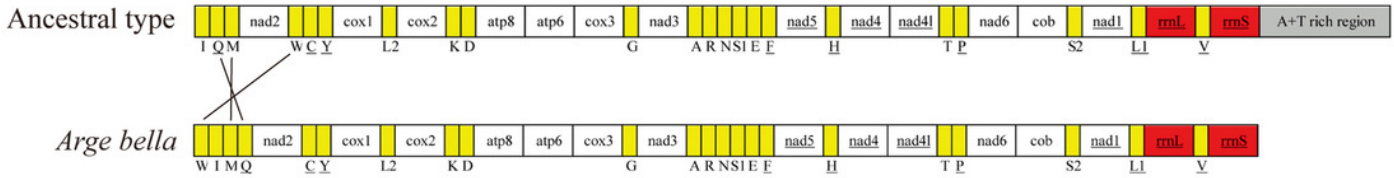

Figure 4

Arge bella tRNAs

Predicted secondary structures of the 22 tRNA genes of *A. bella*. Dashes indicate Watson–Crick base pairing and dots indicate G–U base pairing.

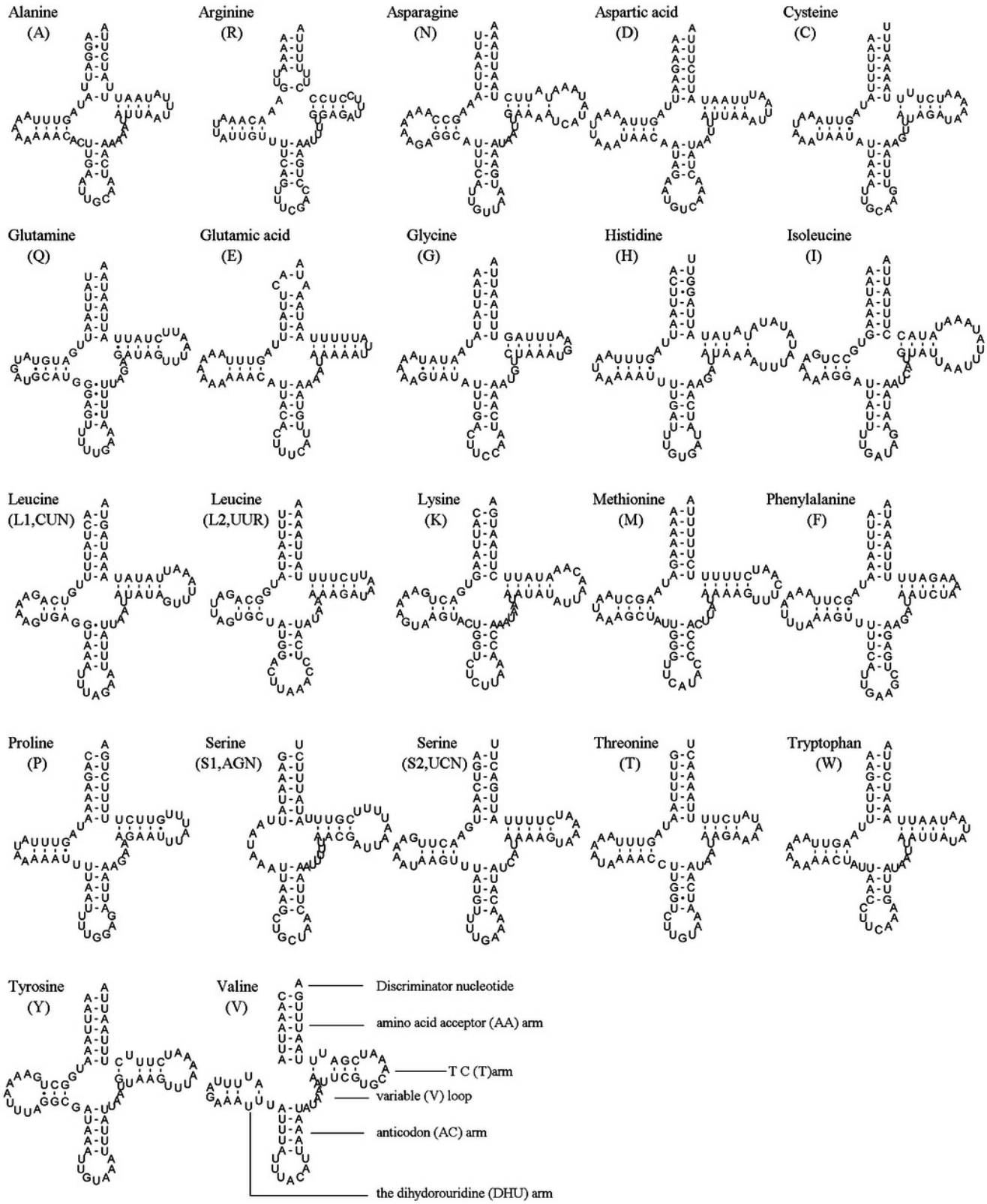
Alanine
(A)
Arginine
(R)
Asparagine
(N)
Aspartic acid
(D)
Cysteine
(C)
Glutamine
(Q)
Glutamic acid
(E)
Glycine
(G)
Histidine
(H)
Isoleucine
(I)
Leucine
(L1,CUN)
Leucine
(L2,UUR)
Lysine
(K)
Methionine
(M)
Phenylalanine
(F)
Proline
(P)
Serine
(S1,AGN)
Serine
(S2,UCN)
Threonine
(T)
Tryptophan
(W)
Tyrosine
(Y)
Valine
(V)
Discriminator nucleotide
amino acid acceptor (AA) arm
T C (T)arm
variable (V) loop
anticodon (AC) arm
the dihydorouridine (DHU) arm

Figure 5

Arge bella rrnS

The predicted secondary structure of *rrnS* in the *Arge bella* mitochondrial genome. The numbering of helices follows Gillespie et al. (2006), roman numbers refer to domain names.

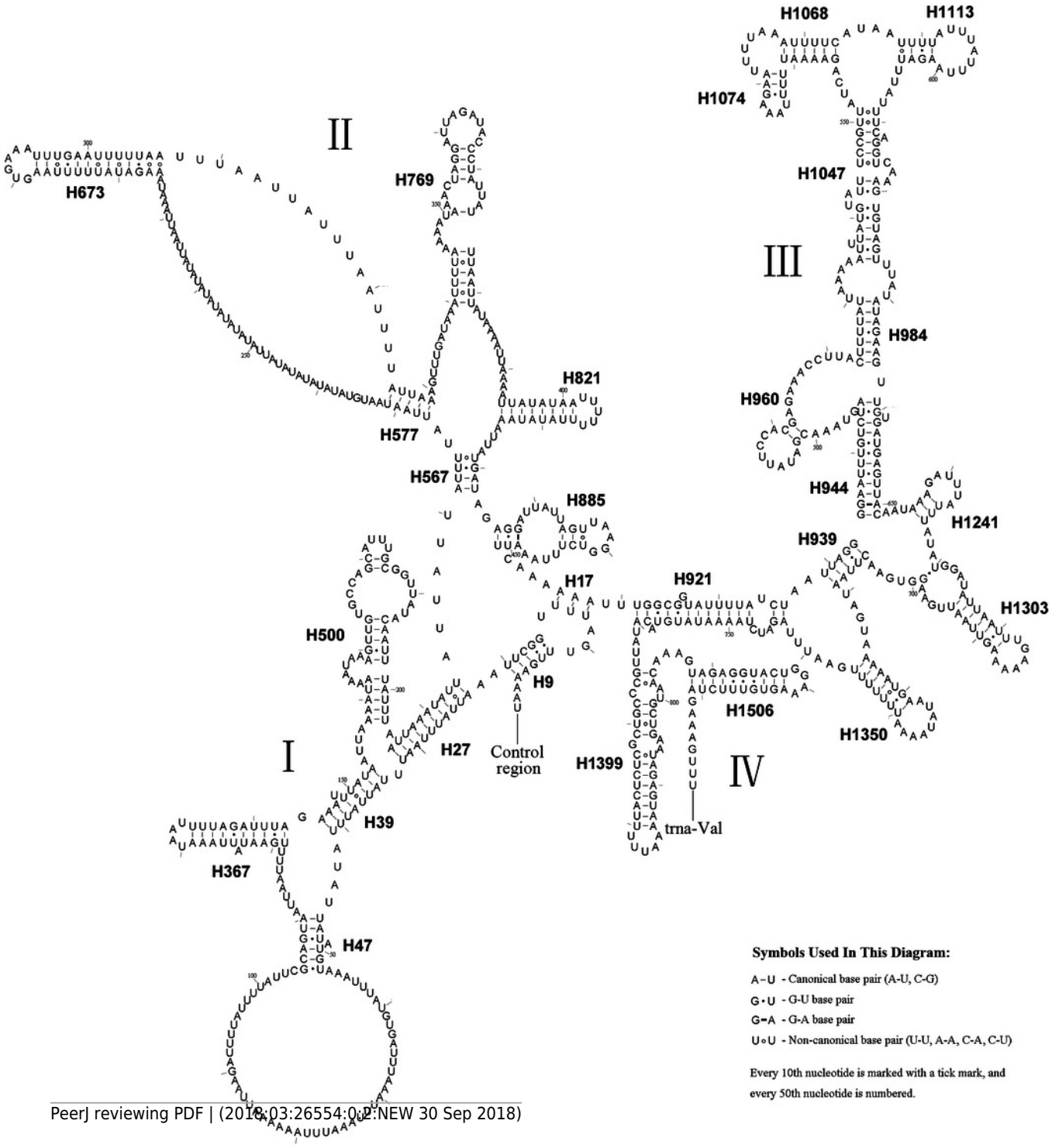

Figure 6

Arge bella rrnL

Predicted secondary structure of *rrnL* in the *Arge bella* mitochondrial genome. The numbering of helices follows Gillespie et al. (2006), roman numbers refer to domain names.

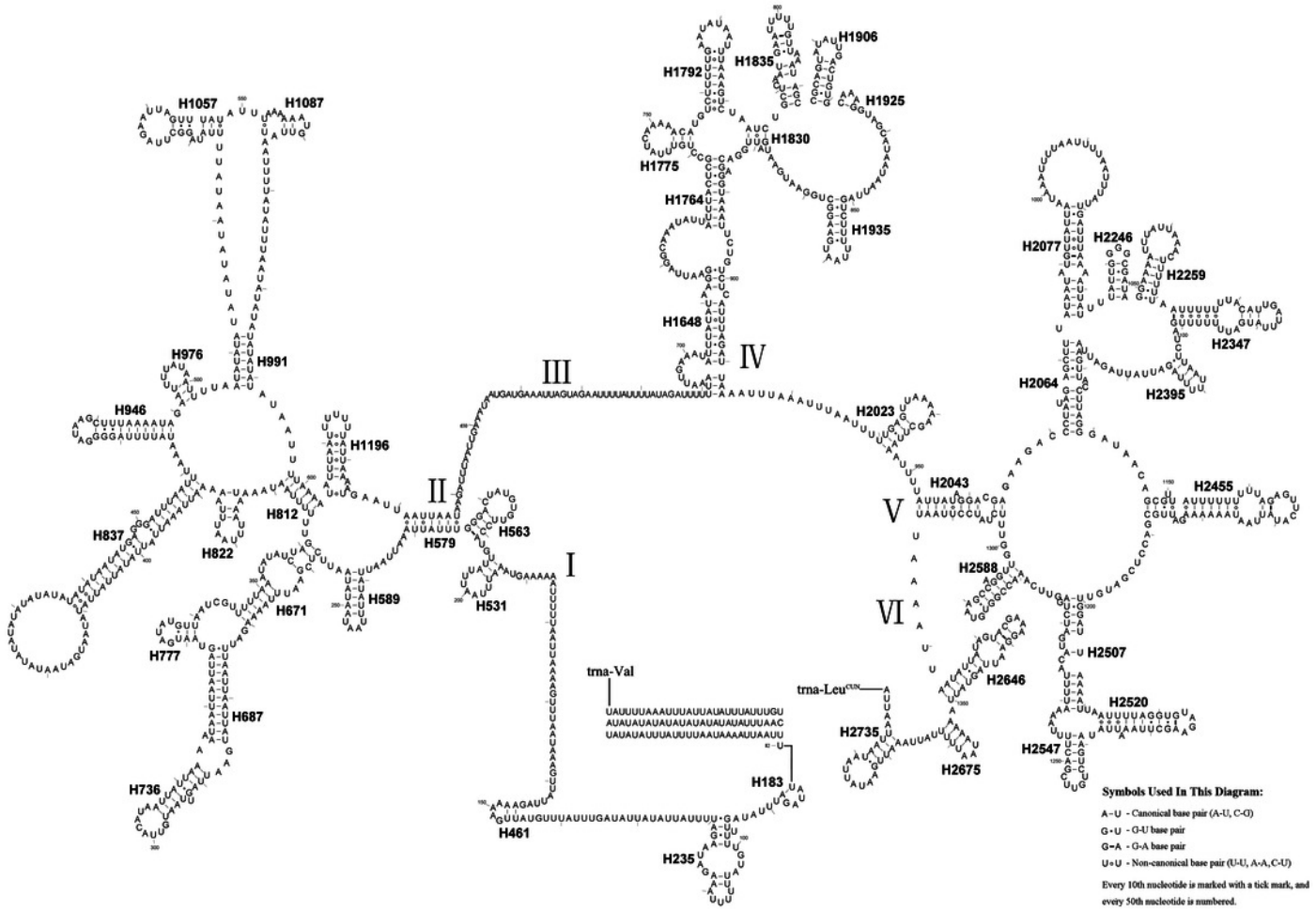

Figure 7

Phylogenetic tree of Symphyta and selected apocritan and outgroup taxa, based on a Maximum-likelihood analysis of sequence data from 13 PCGs and 2 rRNA genes.

The numbers above and below branches are Maximum-likelihood bootstrap values and Bayesian posterior probabilities, respectively. The scale bar indicates the number of substitutions per site.

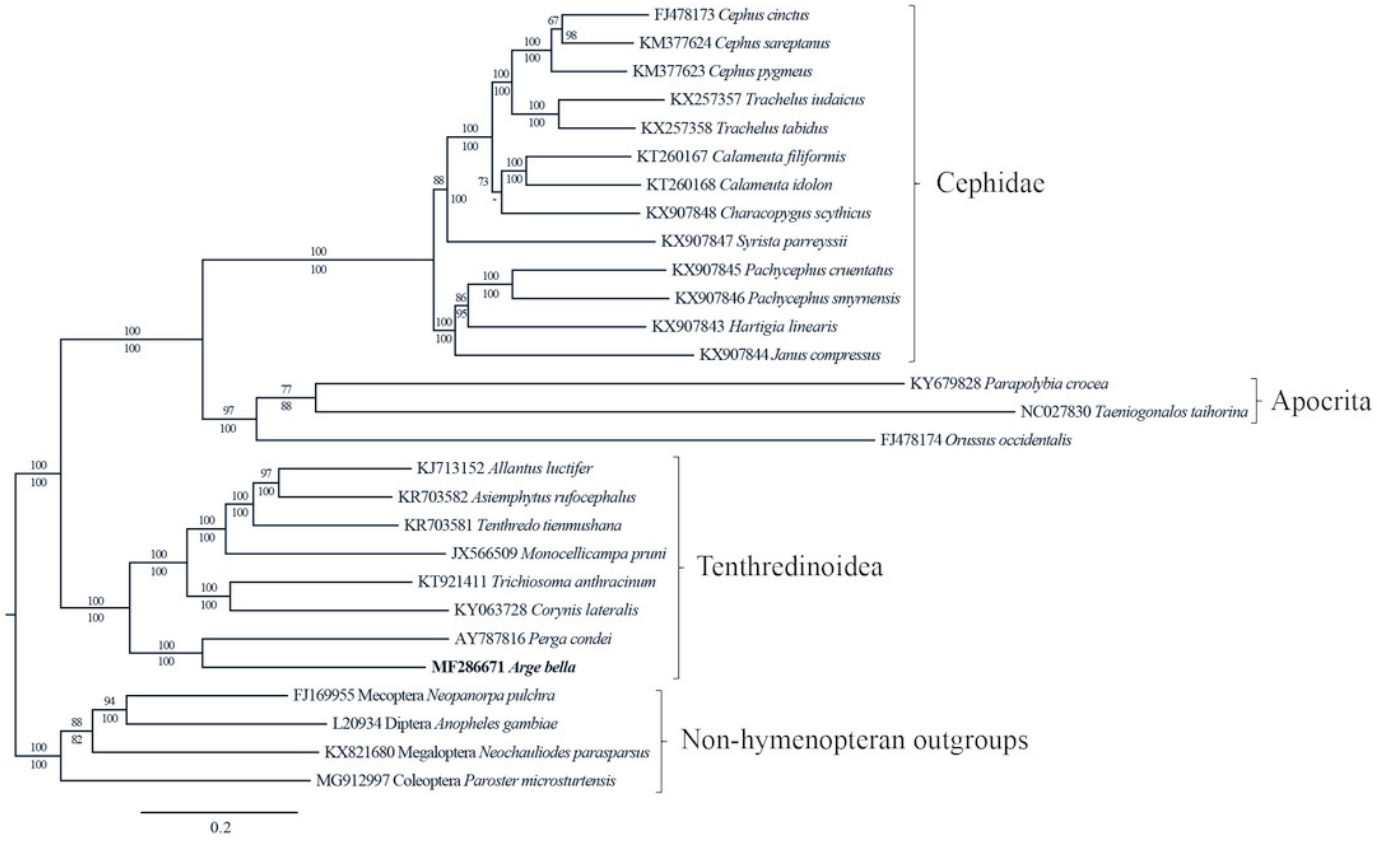

Table 1(on next page)

General information of the mitochondrial genomes of Symphyta

1 General information of the mitochondrial genomes of Symphyta

Species	Length (bp)	Completeness	Family	Subfamily	Accession number	Resources
Perga condei	13,416bp	partial	Pergidae	Perginae	AY787816	Castro and Dowton, 2005
Orussus occidentalis	15,947bp	complete	Orussidae	Orussinae	FJ478174	Dowton et al, 2009
Trichiosoma anthracinum	15,392bp	partial	Cimbicidae	Cimbicinae	KT921411	Song et al, 2016
Corynis lateralis	14,999bp	partial	Cimbicidae	Coryninae	KY063728	Doğan and Korkmaz, 2017
Monocellicampa pruni	15,169bp	partial	Tenthredinidae	Hoplocampinae	JX566509	Wei et al, 2013
Allantus luctifer	15,418bp	complete	Tenthredinidae	Allantinae	KJ713152	Wei et al, 2014
Asiemphytus rufocephalus	14,864bp	partial	Tenthredinidae	Allantinae	KR703582	Song et al, 2016
Tenthredo tienmushana	14,942bp	partial	Tenthredinidae	Tenthrediniae	KR703581	Song et al, 2015
Cephus cinctus	19,339bp	complete	Cephidae	Cephinae	FJ478173	Dowton et al, 2009
Cephus pygmeus	16,145bp	partial	Cephidae	Cephinae	KM377623	Korkmaz et al, 2015
Cephus sareptanus	15,212bp	partial	Cephidae	Cephinae	KM377624	Korkmaz et al, 2015
Calameuta filiformis	20,055bp	complete	Cephidae	Cephinae	KT260167	Korkmaz et al, 2016
Calameuta idolon	19,746bp	complete	Cephidae	Cephinae	KT260168	Korkmaz et al, 2016
Trachelus iudaicus	20,370bp	complete	Cephidae	Cephinae	KX257357	Korkmaz et al, 2017
Trachelus tabidus	18,539bp	complete	Cephidae	Cephinae	KX257358	Korkmaz et al, 2017
Hartigia linearis	20,116bp	partial	Cephidae	Hartigiinae	KX907843	unpublished
Janus compressus	16,700bp	partial	Cephidae	Hartigiinae	KX907844	unpublished
Pachycephus cruentatus	14,568bp	partial	Cephidae	Hartigiinae	KX907845	unpublished
Pachycephus smyrnensis	15,203bp	partial	Cephidae	Hartigiinae	KX907846	unpublished
Syrista parreyssii	15,924bp	partial	Cephidae	Hartigiinae	KX907847	unpublished
Characopygus scythicus	10,558bp	partial	Cephidae	Hartigiinae	KX907848	unpublished

2

Table 2(on next page)

Codon usage of 13 PCGs in mitochondrial genome of *Arge bella*

No., frequency of each codon; RSCU, relative synonymous codon usage.

1 Codon usage of 13 PCGs in mitogenome of *Arge bella*

Amino acid	Codon	NO.	RSCU	Amino acid	Codon	NO.	RSCU
Phe	UUU	355	1.83	Tyr	UAU	147	1.76
	UUC	32	0.17		UAC	20	0.24
Leu	UUA	459	4.85	End	UAA	0	0
	UUG	29	0.31		UAG	0	0
Leu	CUU	31	0.33	His	CAU	53	1.66
	CUC	6	0.06		CAC	11	0.34
	CUA	45	0.47	Gln	CAA	55	1.75
	CUG	0	0		CAG	8	0.25
Ile	AUU	396	1.81	Asn	AAU	214	1.75
	AUC	42	0.19		AAC	30	0.25
Met	AUA	308	1.9	Lys	AAA	128	1.83
	AUG	17	0.1		AAG	12	0.17
Val	GUU	62	2.12	Asp	GAU	53	1.71
	GUC	0	0		GAC	9	0.29
	GUA	52	1.78	Glu	GAA	66	1.74
	GUG	3	0.1		GAG	10	0.26
Ser	UCU	87	2.21	Cys	UGU	39	1.9
	UCC	7	0.18		UGC	2	0.1
	UCA	112	2.85	Trp	UGA	83	1.93
	UCG	3	0.08		UGG	3	0.07
Pro	CCU	63	1.92	Arg	CGU	7	0.6
	CCC	16	0.49		CGC	1	0.09

	CCA	52	1.59		CGA	36	3.06
	CCG	0	0		CGG	3	0.26
Thr	ACU	62	1.46	Ser	AGU	27	0.69
	ACC	13	0.31		AGC	2	0.05
	ACA	93	2.19		AGA	75	1.91
	ACG	2	0.05		AGG	1	0.03
Ala	GCU	55	2.08	Gly	GGU	32	0.7
	GCC	8	0.3		GGC	2	0.04
	GCA	42	1.58		GGA	130	2.86
	GCG	1	0.04		GGG	18	0.4

2

Table 3(on next page)

Mitochondrial genome characteristics of *Arge bella*

J, major; N, minor; IGN, intergenic nucleotides. Minus indicates overlapping sequences between adjacent genes.

1 Mitochondrial genome characteristics of *Arge bella*

Gene	Strand	Start	Stop	Length(bp)	Start codon	Stop codon	Anticodon	IGN
trnW	J	152	219	68			TCA	
trnI	J	253	323	71			GAT	33
trnM	J	354	422	69			CAT	30
trnQ	N	420	488	69			TTG	-3
ND2	J	486	1,562	1,077	ATA	TAA		-3
trnC	N	1,575	1,643	69			GCA	12
trnY	N	1,651	1,722	72			GTA	7
COX1	J	1,723	3,271	1,549	ATT	T		0
trnL2	J	3,272	3,336	65			TAA	0
COX2	J	3,337	4,017	681	ATG	TAA		0
trnK	J	4,062	4,133	72			CTT	44
trnD	J	4,135	4,205	71			GTC	1
ATP8	J	4,207	4,380	120	ATA	TAA		1
ATP6	J	4,374	5,048	675	ATG	TAA		-7
COX3	J	5,048	5,828	781	ATG	T		-1
trnG	J	5,829	5,894	66			TCC	0
ND3	J	5,895	6,248	354	ATA	TAA		0
trnA	J	6,264	6,328	65			TGC	15
trnR	J	6,329	6,393	65			TCG	0
trnN	J	6,419	6,491	73			GTT	25
trnS1	J	6,492	6,554	63			GCT	0
trnE	J	6,556	6,622	67			TTC	1

trnF	N	6,621	6,688	68			GAA	-2
ND5	N	6,694	8,409	1,716	ATT	TAA		5
trnH	N	8,410	8,482	73			GTG	0
ND4	N	8,483	9,821	1,339	ATG	T		0
ND4L	N	9,815	10,111	297	ATT	TAA		-7
trnT	J	10,114	10,179	66			TGG	2
trnP	N	10,180	10,246	67			TGT	0
ND6	J	10,248	10,777	530	ATA	TA		1
CYTB	J	10,778	11,911	1,134	ATG	TAA		0
trnS2	J	11,917	11,985	69			TGA	5
ND1	N	11,996	12,949	954	ATT	TAA		10
trnL1	N	12,950	13,018	69			TAG	0
rrnL	N	13,019	14,408	1,390				0
trnV	N	14,409	14,475	67			TAC	0
rrnS	N	14,476	15,314	839				0

2

Table 4(on next page)

Nucleotide composition of *Arge bella* mitogenome

1 Nucleotide composition of *Arge bella* mitogenome

Feature	Length(bp)	A%	C%	G%	T%	A+T%	AT-skew	GC-skew
Whole genome	15,576	43.2	12.2	7.0	37.5	80.7	0.0706	-0.2708
Protein coding genes	11,222	35.6	10.7	9.9	43.8	79.4	-0.1033	-0.0388
First codon position	3,741	33.2	11.7	13.0	42.1	75.3	-0.1182	0.0526
Second codon position	3,741	33.4	11.9	8.4	46.4	79.8	-0.1629	-0.1724
Third codon position	3,741	40.2	8.6	8.4	42.9	83.1	-0.0325	-0.0118
Protein coding genes-J	6,928	38.2	13.5	8.5	39.9	78.1	-0.0218	-0.2273
First codon position	2,310	39.8	12.5	7.4	40.3	80.1	-0.0062	-0.2563
Second codon position	2,309	39.1	14.1	9.4	37.4	76.5	0.0222	-0.2000
Third codon position	2,309	35.7	13.8	8.6	41.9	77.6	-0.0799	-0.2321
Protein coding genes-N	4,294	31.4	6.3	12.3	50.0	81.4	-0.2285	0.3226
First codon position	1,432	34.4	3.8	12.2	49.6	84.0	-0.1810	0.5250
Second codon position	1,431	25.6	9.4	15.2	49.8	75.4	-0.3210	0.2358
Third codon position	1,431	34.0	5.7	9.5	50.8	84.8	-0.1981	0.2500
ATP6	675	39.1	14.4	7.0	39.6	78.7	-0.0064	-0.3458
ATP8	174	40.8	9.2	1.1	48.9	89.7	-0.0903	-0.7864
COX1	1,549	34.9	14.2	12.0	38.9	73.8	-0.0542	-0.0840
COX2	681	40.7	13.7	9.0	36.7	77.4	0.0517	-0.2070
COX3	781	36.4	14.3	10.9	38.4	74.8	-0.0267	-0.1349
CYTB	1,134	36.1	14.5	9.3	40.1	76.2	-0.0525	-0.2185
ND1	954	49.9	14.0	6.5	29.6	79.5	0.2553	-0.3659
ND2	1,077	42.6	11.6	4.4	41.4	84.0	0.0143	-0.4500
ND3	354	37.6	11.9	8.2	42.4	80.0	-0.0600	-0.1841
ND4	1,339	50.0	12.7	6.0	31.3	81.3	0.2300	-0.3583

ND4L	297	52.5	10.8	5.4	31.3	83.8	0.2530	-0.3333
ND5	1,716	49.8	11.2	6.5	32.5	82.3	0.2102	-0.2655
ND6	530	41.9	12.1	5.3	40.8	82.7	0.0133	-0.3908
rrnL	1,390	45.2	11.1	5.0	38.8	84.0	0.0762	-0.3789
rrnS	839	43.7	11.6	5.4	39.3	83.0	0.0530	-0.3647

2