Neelipleona is an understudied group of Collembola. It has the fewest described species (about 60 species according to http://www.collembola.org, last accessed 2021 Jun 02) among four present recognized Collembola orders (Deharveng 2004). Neelipleona has a single family, Neelidae, which plays a significant role in revealing collembolan’s evolutionary history. The molecular sequences from Neelidae species are relatively scarce, which impedes solving its phylogenetic status. For example, the close affinity between Neelipleona and Symphypleona was traditionally supported by morphological characters like globular shape, neosminthuroid setae, mucro gutter-like etc. (Schneider et al. 2011), which needs to be further tested from the molecular phylogenetic perspective (Xiong et al. 2008). What’s more, Wolbachia infection was detected from Megalothorax incertus Börner, 1903 in our previous studies (Ma et al. 2017). As both Wolbachia and mitochondrial genome are maternally inherited, whether coevolution of them contributed to the parthenogenetic lifestyle remains to be investigated. Here we determined the mitochondrial genome of M. incertus using a combination of multiple experimental and bioinformatic procedures, including long-range PCR, whole genome amplification, high-throughput shotgun sequencing and assembly. The annotated mitochondrial genome sequence of M. incertus was deposited in NCBI database with the accession of MW916537.

Specimens of M. incertus were first collected from Shanghai Expo Houtan Park, Shanghai, China (31°18’N, 121°47’E) by Yan Gao and colleagues in 2012. It was then cultured in the laboratory and confirmed as parthenogenetic through direct observation of reproduction from single unfertilized females. A specimen was deposited at Shanghai Entomological Museum, Chinese Academy of Sciences, Shanghai, China (http://www.shem.com.cn/, contact person: Cheng-Wang Huang, cwhuang@cemps.ac.cn) under the voucher number YIN20210611. Total DNA was extracted for four individuals (labeled as D1–D4) separately with the DNeasy Blood and Tissue Kits (Qiagen, Germany). The barcoding region of cox1 gene was PCR-amplified using universal primer HCO/LCO and sequenced by sanger sequencing to verify the species identity (Simon et al. 1994). Among the four DNA samples, one was used for whole genome amplification (WGA) (sample D1) and another for long-PCR amplification (sample D3), which were followed by illumina shotgun sequencing. The WGA were performed using the REPLI-g Cell WGA & WTA Kit (Qiagen, Germany) following the manufacturer’s recommendations. The amplified genomic DNA was subjected to library preparation and sequencing on illumina HiSeq X Ten System, being carried out at Sangon Biotech (Shanghai) Co., Ltd as a commercial service. The circularized mitochondrial genome sequence was assembled from over 7 Gb of 150 bp pair-end reads using Novoplynas (version 3.6) (Dierckxsens et al. 2016), with cox7 sequence as seed, and further annotated using MITOS2 web portal (Donath et al. 2019). The annotations were also manually checked by
comparing with the reported collembolan mitochondrial genomes. The Long PCR products of the region from cox1 to cytb (∼9 Kb, covering 11 of 13 mitochondrial PCGs) in four overlapped fragments were also achieved using conserved mitochondrial universal primers (Simon et al. 1994). The raw sequencing reads of those mixed PCR products were assembled with MEGAHIT software (Li et al. 2015). The assembled single 9Kb contig verified the sequence accuracy of Novoplasty assembled mitochondrial genome of the same region.

The mitochondrial genome of M. incertus is 14,994 bp in length, like typical hexapod mitogenomes, displaying a general A+T-bias (A+T content: 64.0%). The AT-skew and GC-skew were calculated as AT-skew = (A−T)/(A+T) and GC-skew = (G−C)/(G+C) to assess the strand asymmetry. The AT-skew is −0.001 and GC-skew is −0.341, which suggests A and T has a similar base content while C is favored compared to G. The mitochondrial genome of M. incertus encodes 37 typical mitochondrial genes common to Metazoa. However, the trnR gene was not predicted by MITOS2 initially. After comparing the sequence in the region between trnA and trnN with other collembolan trnR sequences, a potential abbreviated trnR sequences, a potential abbreviated trnR were predicted in this locus, the trnR identity was also supported by harboring a presumed ‘UCG’ anticodon in the RNAfold predicted secondary structure (Lorenz et al. 2011). Four PCGs (cox1, cox2, atp6, and nad5) use incomplete stop codons (TA−/T−), most likely post-transcriptionally restored into complete stop codons. Eleven out of 13 PCGs were annotated to use canonical ATN start codons (five using ATT, four using ATG, and two using ATA). Unexpectedly, cox1 and nad5 might use TTG as the start codon. The gene order is the same as the ancestral gene order for Pancrustacea and in agreement with another nearly complete mitochondrial genome from Neelus murinus Folsom, 1896 (Neelipleona) (MH155200) (Leo et al. 2019).

All available complete mitochondrial genomes of Collembola were retrieved from the GenBank RefSeq database in March 2021. The nearly complete mitochondrial genome of N. murinus was also added to the phylogenetic analysis. The Japyx solifugus (Diplura, Japygidae) was used as an outgroup. Please refer to Figure 1 for accession numbers of each sequence. The 13 PCGs were extracted from GenBank format files using an in-house Python script. The amino acid sequences were aligned individually with MUSCLE (Edgar 2004). Nucleotide sequences were then retro-aligned using the PAL2NAL script (Suyama et al. 2006). Three datasets (nt12, nt123, aa) were used for phylogenetic analysis. All data sets were analyzed using both RAxML and MrBayes algorithms embedded in GENEIOUS R11 software using default settings. All analyses yield almost the same topology,
represented by the nt12 RAxML tree shown in Figure 1. The *M. incertus* clustered with *N. murinus* and in support of monophyly of Neelipleona. Neelipleona is also recovered in a basal position in Collembola. The sister relationship between Neelipleona and Symphyleona was not supported.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This research was supported by the National Natural Science Foundation of China [Grant No. 31501849]. YM was supported by China Postdoctoral Science Foundation [2017M622407].

Data availability statement
The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/) under the accession no. MW916537. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA738273, SRR14826175, and SAMN19717180 respectively.

References
Deharveng L. 2004. Recent advances in Collembola systematics. Pedobiologia. 48(5–6):415–433.