RESEARCH ARTICLE

The Complete Mitochondrial Genomes of Three Bristletails (Insecta: Archaeognatha): The Paraphyly of Machilidae and Insights into Archaeognathan Phylogeny

Yue Ma1,2, Kun He2, Panpan Yu1,2, Danna Yu1, Xuefang Cheng1,2, Jiayong Zhang1,2*

1 Institute of Ecology, Zhejiang Normal University, Jinhua, Zhejiang Province, China, 2 College of Chemistry and Life Science, Zhejiang Normal University, Jinhua, Zhejiang Province, China
* zhang3599533@163.com

Abstract

The order Archaeognatha was an ancient group of Hexapoda and was considered as the most primitive of living insects. Two extant families (Meinertellidae and Machilidae) consisted of approximately 500 species. This study determined 3 complete mitochondrial genomes and 2 nearly complete mitochondrial genome sequences of the bristletail. The size of the 5 mitochondrial genome sequences of bristletail were relatively modest, containing 13 protein-coding genes (PCGs), 2 ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes and one control region. The gene orders were identical to that of Drosophila yakuba and most bristletail species suggesting a conserved genome evolution within the Archaeognatha. In order to estimate archaeognathan evolutionary relationships, phylogenetic analyses were conducted using concatenated nucleotide sequences of 13 protein-coding genes, with four different computational algorithms (NJ, MP, ML and BI). Based on the results, the monophyly of the family Machilidae was challenged by both datasets (W12 and G12 datasets). The relationships among archaeognathan subfamilies seemed to be tangled and the subfamily Machilinae was also believed to be a paraphyletic group in our study.

Introduction

Bristletails (Insecta: Archaeognatha) were primarily wingless insects inhabiting seashore or inland habitats all around the world, and they typically lived under rocks, logs, and the loose bark of standing trees. The most conspicuous features of these insects were the large compound eyes that meet dorsally, the well developed ocelli and the ability of these insects to spring into the air by a sudden flexure of the body [1–2]. Except for a rare few, bristletails were nocturnal and typically hid in crevices during the day. Archaeognatha was an ancient insect order with modest species diversity (about 500 species in two extant families: Machilidae and Meinertellidae) and the family Machilidae was divided into three subfamilies: Petrobiellinae, Petrobiinae and Machilinae [3]. The Meinertellidae occured predominantly in the Southern Hemisphere [4].
The Machilidae, considered by many authors to possess more primitive features than the Meinertellidae [3], were mostly found in the Northern Hemisphere. Relatively few cladistics analyses had addressed relationships among the three subfamilies of Machilidae and most hypotheses on infraordinal relationships were based on morphological evidence [5–6]. Among the three subfamilies, Machilinae was believed to occupy a series of primitive morphological characteristics, while the species of Petrobiinae and Petrobiellinae were more derived compared with Machilinae species based on the apomorphies such as the absence of scales of antennae and differences of penis feature and abdominal styli. Consequently, a tentative phylogenetic relationship (Machilinae + (Petrobiinae + Petrobiellinae)) was proposed [3].

True insects contained about two dozen orders of winged insects (Pterygota) and two orders of primitively wingless insects, Archaeognatha and Zygentoma. Zygentoma was the putative sister group of the pterygote insects; these two together formed the clade Dicondylia, with Archaeognatha as its sister group. Non-insect hexapods (the so-called Entognatha) comprised the orders Protura, Collembola, and Diplura. The phylogenetic relationships between and within the different orders which composed the primitively wingless hexapods, the so-called apterygote insects, and also their relations with the pterygote hexapods, had been subject to controversial assumptions [7–10]. Even though basal hexapods had modest species diversity, interesting relict distributions, and they clearly were phylogenetically significant, the phylogenetic work on these orders had so far been surprisingly preliminary [11–14].

The insect mitochondrial genome was a compact circular molecule typically 15–18 kb in size and typically encoded for 13 proteins, 22 tRNAs and 2 rRNAs. In addition to the genes, there was a large non-coding region termed the control region as it regulated the transcription and replication of the mtDNA [15–16]. Mitochondrial genome sequences were becoming increasingly important for comprehensive evolutionary studies. MtDNA had been widely regarded as the molecular marker for the phylogenetic analysis in metazoans because of its abundance in animal tissues, the small genome size, faster rate of evolution, low or absence of sequence recombination, and evolutionary conserved gene products [17–18]. Mitochondrial genome sequences were not only more informative than shorter sequences of individual genes, but also provided a set of genomic characters, such as the relative position of different genes, RNA secondary structures and modes of control of replication and transcription [19–22]. To date, over 100 publications reported insect phylogenies built using mitochondrial genome data [23].

Here we described the complete mitochondrial genomes of 3 bristletails, Petrobiellus bannaensis, Allopsontus halawuensis, Allopsontus helanensis, and 2 nearly complete mitochondrial genome sequences, Pedetontinus luanchuanensis and Petrobiellus puerensis, all of which belonged to the family Machilidae. P. bannaensis and P. puerensis were members of the subfamily Petrobiellinae, and P. luanchuanensis belonged to the subfamily Petrobiinae, while the other two bristletails were in the Machilinae subfamily. The mitochondrial genome sequences were compared with those of other bristletails available in the GenBank database (Pedetontinus silvestrii, Petrobius brevistylis, Trigoniophthalmus alternatus, Songmachilis xinxiangensis and Neso-machilis australica). The protein-coding gene sequences were used to construct datasets for phylogenetic analysis of Archaeognatha. In addition, the genomic sequences were also employed for the inference of the phylogenetic relationships among basal lineages of Hexapoda.

Material and Methods

Ethics statement

No specific permits were required for the insect collected for this study. The field studies did not involve endangered or protected species. The species in the order Archaeognatha were
common small insects living under rocks or dead leaves and were not included in the “List of Protected Animals in China”.

Specimen collection
Specimens of *P. bannaensis* were collected from Xishuangbanna of Yunnan province (E101° 14', N21° 55’), *P. puerensis* from Puer of Yunnan province (E101° 2', N23° 5’), *A. helanensis* (E105° 51', N38° 58’) and *A. halawuensis* (E106° 8’, N39° 1’) from Ho-Ian Mountains of Gansu province and *P. luanchuanensis* (E111° 45’, N33° 42’) from Luanchuan of Henan province. Voucher specimens were deposited in College of Chemistry and Life Sciences, Zhejiang Normal University, Zhejiang, China. All specimens were preserved in 100% ethanol in field. After being transported to the laboratory, they were stored at −70°C until used for DNA extraction. All the five species were identified by morphological characteristics.

DNA extraction, PCR amplification and sequencing
Total genomic DNA from each individual was extracted using the QIAGEN DNeasy Blood& Tissue Kit following the manufacturer’s protocol. DNA samples were stored at −20°C until further use. Based on alignments and comparisons of complete mitochondrial sequences of other bristletails and our previous researches [24], ten pairs of universal primers were designed for amplification of the mitochondrial genomes (S1 Table). All PCRs were performed using a Bio-RAD MJ Mini Personal Thermal Cycler. Based on the sequence information got from earlier PCR by universal primers, several pairs of specific primers were designed to obtain the whole mitochondrial genome. PCR reactions were carried out in a 50 μL reaction volume consisting of 32.75 μL sterile deionized water, 5.0 μL 10×PCR Buffer (Mg2+Free), 5.0 μL MgCl2 (25 mM), 4.0 μL dNTP Mixture (2.5 mM each), 1.0 μL DNA template, 1.0 μL each primer (10 ppm), 0.25 μL Takara Taq DNA polymerase (5 U/μL) with the following conditions: after an initial denaturation at 94°C for 4 min, then 94°C for 40 s (denaturation), 46–62°C for 50 s (annealing), 72°C for 1–5 min (extension) for 35 cyclers, followed by 72°C for 10 min (final extension). Each amplicon (5μL) was examined by agarose gel electrophoresis to validate amplification efficiency. After being purified by Axygen agarose-out kit, all the amplicons were sent to Sangon Biotech Company (Shanghai, China) for sequencing from both directions by using the primers used in the PCR amplification.

Sequence analysis and annotation
All of the typical protein-coding genes were identified by BLAST searches of NCBI database. Sequences were checked and assembled using the programs Seqman (DNASTAR, Inc.) and then adjusted manually. PCGs and rRNA genes were identified based on sequences similarity with published bristletail mitochondrial genome sequences from public database (e.g., GenBank). The base composition, codon usage, and open-reading frames were analyzed using program Mega 5.0 [25]. Translation initiation and translation termination codons were identified based on comparison with the mitochondrial genomes of other known bristletails. The overlapping regions and intergenic spacers between genes were counted manually. The tRNA genes were identified by their cloverleaf secondary structure using tRNAscan-SE Search Server v.1.21 [26] with default settings. The tRNA genes which could not be determined by tRNAscan-SE were determined by sequence similarity to tRNAs of other bristletails. Sequence data had been deposited in the NCBI database, and the accession numbers of *Petrobiellus bannaensis*, *Allopsontus halawuensis*, *Allopsontus helanensis*, *Pedetontinus luanchuanensis* and *Petrobiellus puerensis* were KJ754503, KJ754500, KJ754501, KJ754502 and KJ754504, respectively.
Phylogenetic analyses of Archaeognatha and basal Hexapoda

To elucidate the phylogenetic relationships among basal Hexapoda, 35 sequences of the complete mitochondrial genomes were obtained from GenBank database (Table 1). Among the mitochondrial genomes we selected, there were 5 from the earlier published bristletails [2, 24, 27–29], 3 from another primitive wingless order Zygentoma [30–32], 8 from the Paleoptera order: Ephemeroptera and Odonata [33–36], 4 from the Polyneoptera order: Blattaria [37], Isoptera [38], Mantodea [39] and Orthoptera [40]. To optimize the outgroup selection, we selected 14 entognathian species and 2 Crustacean species as outgroups to construct phylogenetic trees, respectively [28, 30, 41–46]. The nucleotide and putative amino acid regions for each of the 13 mitochondrial protein-coding genes were aligned using Clustal W in the program Mega 5.0 [25]. All the alignments were analyzed with the program Gblock 0.91b using default settings, to select conserved regions of the putative amino acids [47]. A saturation analysis was performed on the first, second and third codon positions of PCGs using DAMBE 4.2.13 [48]. According to the results from saturation analysis, third codon positions were saturated, so they were excluded from the final nucleotide alignment. Phylogenetic analyses were conducted using four methods: neighbor-Joining (NJ), maximum parsimony (MP), maximum likelihood (ML) and Bayesian (BI). GTR+I+G model was chosen for the likelihood and Bayesian analyses. The Maximum likelihood analysis (ML) of the nucleotide alignment was performed using PAUP* 4.0b10 with 100 bootstrap replicates [49]. Maximum parsimony analysis (MP) and Neighbor-Joining (NJ) were performed using PAUP* 4.0b10. 1000 bootstrap replicates were generated for the MP analysis with 10 replicates with random taxon order. BI analysis was performed using MrBayes, ver.3.1.2 [50]. For Bayesian Inference, 4 independent Markov chains were simultaneously running for 1000000 generations. Trees were sampled every 100 generations and the first 20% of the generations were discarded as burn-in and the remaining trees were used to calculate Bayesian posterior probabilities.

We selected Onychiurus orientalis from Collembola and Campodea lubbocki from Diplura as outgroups to construct the phylogeny of the order Archaeognatha, and for the first time, tried to elucidate the relationships among subfamilies of Archaeognatha using mitochondrial genomes. W12 dataset was constructed based on the nucleotide sequences of all the 13 protein-coding genes, which included all the indels after being aligned; G12 dataset was constructed using the program Gblock 0.91b, this dataset possessed conserved regions of all the 13 protein-coding genes, so all the poorly aligned positions of nucleotide alignment were removed. Then, W12 and G12 datasets were utilized to constructed phylogenetic trees with NJ, MP, ML and BI algorithms, respectively.

Results and Discussion

Genome organization

The complete mitochondrial genomes of *P. bannaensis*, *A. halawuensis*, and *A. helanensis* were obtained with the size of 15843 bp, 15532 bp, and 15538 bp, respectively (Table 2). But unfortunately, we were not able to successfully amplify the control region of *P. luanchuanensis* and *P. puerensis* with the length of amplified sequences of 15196 bp and 14562 bp, respectively. The length of the five mitochondrial genomes were all within the range of typical size for metazoan mt genomes, which was considered to be 14–18 kb [16]. All the 5 mitochondrial genomes contained 13 protein-coding genes (CO1–3, ND1–6, ND4L, atp6, atp8 and cob), a small subunit ribosomal RNA gene (rrnS), a large subunit ribosomal RNA gene (rrnL), and 22 transfer RNA genes, and 23 of which were transcribed on the majority J-strand and the others mapped to the minority N-strand. The gene orders of the five mitochondrial genomes were congruent with *Drosophila yakuba*, which was regarded as the ancestral metazoan mitochondrial genome [51].
In the mitochondrial genome of *A. halawuensis*, there was a total of 43 bp gene overlap, which was the highest in the 5 mt genomes, and the lowest was *A. helanensis* with a total of 20 bp overlap. The 43 bp gene overlap in 9 positions of *A. halawuensis* mt genome was ranging in size from 1 to 16 bp and the longest is a 16 bp overlap between tRNAPro and ND6, which was
Table 2. Annotation of the mitochondrial genomes of *Petrobiellus bannaensis* (Pb), *Allopsontus halawuensis* (Aha) and *Allopsontus helanensis* (Ahe).

| Gene/strand | Position | Start/stop codon |
|-------------|----------|------------------|
|             | Pb       | Aha              | Ahe              | Pb       | Aha              | Ahe              |
| **trnI / J** | 1–69     | 1–68             | 1–67             |          |                  |                  |
| **trnQ / N** | 83–150   | 69–138           | 68–136           |          |                  |                  |
| **trnM / J** | 149–217  | 142–207          | 140–205          |          |                  |                  |
| **ND2 / J** | 227–1246 | 226–1239         | 209–1231         |          |                  |                  |
| **trnW / J** | 1247–1315| 1239–1306        | 1230–1298        |          |                  |                  |
| **trnC / N** | 1328–1396| 1306–1372        | 1298–1362        |          |                  |                  |
| **trnY / N** | 1407–1471| 1373–1440        | 1363–1430        |          |                  |                  |
| **Cox1 / J** | 1506–3005| 1433–2972        | 1423–2962        |          |                  |                  |
| **trnL2 / J** | 3032–3097| 2973–3037        | 2963–3026        |          |                  |                  |
| **Cox2 / J** | 3098–3784| 3043–3730        | 3034–3721        |          |                  |                  |
| **trnK / J** | 3783–3854| 3731–3802        | 3722–3792        |          |                  |                  |
| **trnD / J** | 3854–3918| 3804–3869        | 3795–3861        |          |                  |                  |
| **Atp8 / J** | 3919–4080| 3870–4031        | 3862–4023        |          |                  |                  |
| **Atp6 / J** | 4074–4751| 4028–4702        | 4017–4691        |          |                  |                  |
| **Cox3 / J** | 4753–5544| 4705–5485        | 4695–5477        |          |                  |                  |
| **trnG / J** | 5549–5612| 5503–5567        | 5482–5547        |          |                  |                  |
| **ND3 / J** | 5610–5966| 5565–5921        | 5548–5901        |          |                  |                  |
| **trnA / J** | 5965–6027| 5925–5988        | 5905–5968        |          |                  |                  |
| **trnR / J** | 6033–6099| 5996–6060        | 5971–6037        |          |                  |                  |
| **trnN / J** | 6098–6164| 6060–6125        | 6037–6106        |          |                  |                  |
| **trnS1 / J** | 6165–6231| 6127–6191        | 6107–6174        |          |                  |                  |
| **trnE / J** | 6233–6298| 6193–6258        | 6175–6243        |          |                  |                  |
| **trnF / N** | 6308–6373| 6261–6326        | 6244–6308        |          |                  |                  |
| **ND5 / N** | 6374–8104| 6328–8062        | 6311–8045        |          |                  |                  |
| **trnH / N** | 8102–8165| 8064–8127        | 8047–8111        |          |                  |                  |
| **ND4 / N** | 8171–9499| 8133–9482        | 8115–9455        |          |                  |                  |
| **ND4L / N** | 9511–9813| 9476–9775        | 9458–9757        |          |                  |                  |
| **trnT / J** | 9816–9878| 9778–9841        | 9760–9823        |          |                  |                  |
| **trnP / N** | 9880–9946| 9842–9906        | 9824–9888        |          |                  |                  |
| **ND6 / J** | 9948–10472| 9891–10418      | 9892–10395       |          |                  |                  |
| **CYTP / J** | 10472–11608| 10426–11562      | 10395–11531      |          |                  |                  |
| **trnS2 / J** | 11614–11680| 11561–11628      | 11532–11595      |          |                  |                  |
| **ND1 / N** | 11702–12634| 11648–12586      | 11616–12536      |          |                  |                  |
| **trnL1 / N** | 12652–12717| 12593–12659      | 12558–12624      |          |                  |                  |
| **rrnL / N** | 12718–14040| 12660–13991      | 12625–13946      |          |                  |                  |
| **trnV / N** | 14041–14113| 13992–14063      | 13947–14018      |          |                  |                  |
| **rrnS / N** | 14114–14961| 14064–14876      | 14019–14824      |          |                  |                  |
| **A+T-rich region** | 14962–15843| 14877–15532      | 14825–15538      |          |                  |                  |

Note: J and N refer to the majority and minority strand, respectively. Position numbers refer to positions on the majority strand.

doi:10.1371/journal.pone.0117669.t002
also the longest overlap between individual genes of the 5 mt genomes. Compared to previously published 5 bristletail mt genomes, the gene overlap occurred between atp8 and atp6 seemed to be common across bristletail mt genomes and this 7 nucleotides “ATGATAA” overlap also existed in other insects [52–53].

In addition to the large non-coding region, several small non-coding intergenic spacers were also found in the 5 mitochondrial genomes. The non-coding regions of mtDNA of P. bannaensis is 187 bp in total, which is the highest in the 5 mt genomes, and made up of 19 intergenic spacer sequences, ranging in size from 1 bp to 34 bp. There were two relatively large spacers flanked by the protein-coding genes CO1 and ND1 in the P. bannaensis mt genome. Interestingly, the intergenic spacers flanked by ND1 gene seemed to be a common feature across bristletail mitochondrial genomes. The total nucleotides of intergenic spacers between tRNA{Ser}^{2}, ND1 and tRNA{Leu}^{1}1 in P. bannaensis, A. halawuensis, A. helanensis and P. puerensis were 38 bp, 19 bp, 41 bp and 36 bp, respectively. But notably, such feature was not found in the P. luanchuanensis mt genome, a 8 bp overlap existed between tRNA{Ser}^{2} and ND1 gene.

The nucleotide composition of P. bannaensis, A. halawuensis, A. helanensis, P. luanchuanensis and P. puerensis were biased toward A/T nucleotides, which account for 69.7%, 68.1%, 68.9%, 73.2% and 70.7%, respectively, and these values were well within the range detected so far in mitochondrial genomes of bristletail (67.9%–74.4%).

Protein-coding genes

13 protein-coding genes were identified in the 5 mitochondrial genomes of P. bannaensis, A. halawuensis, A. helanensis, P. luanchuanensis and P. puerensis, with characteristics similar to previously published 5 bristletail mitochondrial genomes. Among the 13 protein-coding genes, 4 genes located on the minority N-strand, and the others were transcribed on the majority J-strand. The commonly used protein-coding gene start codons were ATN, GTG [54] and TTG [55], but there was also reported no typical ATN start codon for CO1 gene in mitochondrial genomes of some other species, such as CCG [56], CGA [57], ACC [36], ACG [58], TTA [59] and even a tetranucleotide codon, ATTA, had been proposed in the mitogenome of Locusta migratoria [60]. All the 13 protein-coding genes of the 10 bristletail mitochondrial genomes (5 from our work and 5 from NCBI database) started with typical ATN and GTG (Table 2). In the mitochondrial genome of P. bannaensis, 6 genes shared ATG initiation codon, 3 shared ATT and ATA respectively, and only 1 gene started with ATC. With the similar characteristics in P. bannaensis mitochondrial genome, most protein-coding genes in bristletail mt genomes started with ATG and ATT. While interestingly, 7 protein-coding genes in N. australica (from family Meineriellidae) mt genome had a ATA start codon, which was quite different with the other 9 species.

In the P. bannaensis and P. puerensis mitochondrial genomes, all the PCGs terminated with the conventional stop codons TAG or TAA. While in terms of the A. halawuensis mt genome, 4 genes (CO1, CO2, CO3 and ND5) had incomplete stop codon T adjacent to downstream tRNA genes. And the incomplete stop codon T was also found in A. helanensis (CO1, CO2 and ND5) and P. luanchuanensis (CO2, CO3 and ND5) mitochondrial genomes. The presence of incomplete stop codon was common in metazoan mitochondrial genomes, and these truncated stop codons was presumed to be complete by post-transcriptional polyadenylation [61].

The A+T content of protein-coding gens, excluding stop codons, was 67.3%, 66.6%, 66.9%, 72.2% and 69.6% in P. bannaensis, A. halawuensis, A. helanensis, P. luanchuanensis and P. puerensis, respectively. In the 5 mitochondrial genomes, the first, second and third codon positions was calculated separately and the third codon positions seemed to have the highest A+T content (75.7%, 75.5%, 75.8%, 86.9% and 81.3%, respectively), the strongest bias toward T was in the second codon positions of the 5 species (Table 3).
Transfer RNA, Ribosomal RNA genes and A+T rich region

The complete set of 22 tRNAs (one specific for each amino acid and 2 each for leucine and serine) found in typical metazoan mitochondrial genomes were identified in *P. bannaensis*, *A. halawuensis*, *A. helanensis*, *P. luanchuanensis* and *P. puerensis*. Among these tRNA genes, 14 tRNAs were encoded by the J-strand with the rest by the N-strand. All the tRNA genes could be folded into the typical cloverleaf structure using tRNAscan-SE 1.21 except for tRNAser(AGN), in which its dihydrouridine (DHU) stem simply formed a loop, as was often found in other insect mitochondrial genomes. The anticodons of the 5 species were identical to those of the respective tRNAs found in previously published bristletail mitochondrial genomes. Unmatched base pairs had been detected in stems of tRNA secondary structure. Among the 5 mitochondrial genomes, *P. bannaensis* possessed 42 unmatched base pairs, consisting of 35 G-U pairs, 2 U-U and 3 U-C mismatches, which was the highest compared to other 5 species. Except for G-U, U-U, U-C pairs and mismatches, the other types of mismatches like A-C and A-A had also been identified. tRNA gene cluster rearrangement occurred between protein-coding genes ND3 and ND5 had been found in other bristletails. In the mt genomes of *T. alternatus* and *P. brevistylis*, the tRNA gene orders between ND3 and ND5 were tRNAArg-tRNAAsn-tRNASer1-tRNAGlu-tRNAAla-tRNAPhe and tRNAAla-tRNAArg-tRNAAsn-tRNASer1-tRNAGlu-tRNATyr-tRNAPhe, respectively. But in our study, there was no tRNA gene rearrangement that had been detected, and the 5 bristletail species owned the typical insects tRNA gene order within ND3 and ND5, which was tRNAAla-tRNAArg-tRNAAsn-tRNASer1-tRNAGlu-tRNAPhe.

The ends of ribosomal RNA genes were impossible to be precisely determined by DNA sequencing alone, so the ends of rRNA genes were assumed to be at the boundaries of flanking genes. As in the 5 mitochondrial genomes, the 1rRNA gene was between tRNALeu1 and tRNAVal, while the srRNA was flanked by tRNAVal and A+T rich region and both genes were encoded on the minor strand. The length of srRNA in *P. bannaensis* mitochondrial genome was 848 bp, which was relatively higher than that of *A. halawuensis* (813 bp) and *A. helanensis* (806 bp) mt genomes. We could not determine the length of *P. luanchuanensis* and *P. puerensis* because of our inability to successfully amplify the srRNA gene and A+T rich region. The length of 1rRNA gene was similar to other bristletail mt genomes.

The largest non-coding region of bristletail mitochondrial genome was the control region, which was located between srRNA and the tRNAlue-tRNAGlu-tRNAMet gene cluster. Compared to previously published bristletail mitochondrial genomes, the length of the control region

---

### Table 3. Nucleotide compositions of *Petrobiellus bannaensis* (Pb), *Allopsontus halawuensis* (Aha) and *Allopsontus helanensis* (Ahe).

| feature          | A (%) | C (%) | G (%) | T (%) | A+T (%) |
|------------------|-------|-------|-------|-------|---------|
| Whole genome     | Pb    | Aha   | Ahe   | Pb    | Aha    | Ahe    |
|                  | 39.0  | 35.4  | 36.4  | 19.5  | 19.8   | 19.4   |
| Protein-coding   | Pb    | Aha   | Ahe   | Pb    | Aha    | Ahe    |
| genes            | 29.5  | 28.0  | 27.9  | 16.7  | 17.3   | 17.0   |
| 1st codon positions | 31.7  | 28.9  | 29.3  | 15.1  | 16.1   | 16.2   |
| 2nd codon positions | 18.1  | 18.5  | 18.8  | 20.3  | 20.6   | 20.2   |
| tRNA genes       | 35.5  | 35.2  | 37.3  | 14.5  | 15.2   | 14.5   |
| 3rd codon positions | 38.7  | 36.5  | 35.8  | 14.5  | 15.2   | 14.5   |
| rRNA genes       | 30.8  | 33.1  | 35.1  | 7.6   | 9.5    | 9.3    |
| A+T-rich region  | 42.7  | 40.2  | 43.1  | 11.9  | 9.3    | 11.5   |

Note: Stop codons were excluded.
ranged from 538 bp (N. australica) to 1149 bp (T. alternatus), and the size of control region in this study were well within this range. The control region length variation was probably due to tandem repeat regions and differences in copy numbers.

Phylogeny of Archaeognatha and basal Hexapoda

The Archaeognatha consisted of only 2 extant families and approximately 500 species: Meinertellidae and Machilidae [3]. The Machilidae was considered to possess more primitive features (e.g. relating to ovipositor, penis, parameres, coxal vesicles and tarsomeres) than the Meinertellidae. Kaplin subdivided the family Machilidae into 3 subfamilies: Machilinae, Petrobiinae and Petrobiellinae [5]. And Sturm proposed the monophyly of the 3 subfamilies had to be proved by additional arguments [4]. We performed phylogenetic analysis with nucleotide sequences of 13 mitochondrial protein-coding genes from 10 archaeognathan species and 2 outgroup species (Onychiurus orientalis [30] and Campodea lubbocki [46]). Four phylogenetic algorithms (NJ, MP, ML and BI) yielded 2 datasets (Fig. 1 and Fig. 2): W12 (13 PCG nucleotide sequences with poorly aligned positions of nucleotide alignment) and G12 (13 PCG nucleotide sequences analyzed by Gblock 0.91b). The 2 datasets showed the following features: firstly, the family Machilidae seemed to be a paraphyletic group, which was not supporting the taxonomic classification by analysis of morphological data. In the W12 dataset, all the analyses support this view, while in the G12 dataset, both NJ and BI analyses yielded the same results. Secondly, the monophyly of the subfamily Machilinae was also challenged by all analyses in this paper. The group consisting of (S. xinxiangensis + (A. halawensis + A. helanensis)) was strongly supported as being monophyletic across all analyses, but another Machilinae species T. alternatus was not in this clade. The main morphological characters for the diagnosis of this subfamily by Sturm were plesiomorphies, so he proposed a subdivision or an additional diagnosis was required, and our results suggested the phylogenetic relationships among Machilinae was not congruent with morphological taxonomy. Thirdly, the Machilidae subfamily Petrobiellinae seemed to have a closer relationship with the other family Meinertellidae. P. puerensis and P. bannaensis were clustered with the Meinertellidae species N. australica in most analyses with relatively high nodal supports. This subfamily comprised only one genus Petrobiellus, and besides some plesiomorphies, this genus shared an apomorphic feature with the Petrobius group: the weak and indistinct separation of the distal mandibular teeth. Sturm suggested that this feature could indicate the Petrobiellus group and Petrobius were monophyletic or the parallel evolution had taken place [3]. But in our analyses, the genus Petrobiellus and Petrobius were not in the same clade, interestingly, the 2 Petrobiellus species formed a monophyletic clade with the Meinertellidae species N. australica in most analyses (all analyses in W12 dataset and NJ, BI analyses in G12 dataset). Our result may provide evidence for resolving phylogenetic relationships of Archaeognatha, although the currently limited taxon sampling showed the preliminary nature of this analysis. The unsampled groups should be included in further studies to provide a more comprehensive phylogenetic estimate.

To elucidate the phylogenetic relationships among basal Hexapoda, 35 sequences of the complete mitochondrial genomes were obtained to construct phylogenetic trees using 14 Entognathan species and 2 Crustacean species as outgroups, respectively. Based on the results, the monophyly of Archaeognatha was an undoubted hypothesis strongly supported by both morphological features and molecular data that was also demonstrated in our study based on mitochondrial genomes (Figs. 3, 4, 5 and 6). The topologies showed the similar characteristics with datasets W12 and G12, corroborating the paraphyly of Machilinae and other features demonstrated by those two datasets. The traditional taxon Thysanura was the combination of bristletails and Zygentoma species. However, it had been recognized as an unnatural group in
recent years. The topologies showed that silverfish (Zygentoma) were more closely related to
the winged insects (pterygota) than to bristletails. Most of our analyses yielded a topology of
(Ephemeroptera + (Odonata + Neoptera)), but both NJ analyses yielded a (Odonata + (Ephe-
meroptera + Neoptera)) clade (S1 Fig. and S3 Fig.), hence, the phylogenetic relationships be-
tween the two Paleoptera orders needed a further investigation. The monophyly of Pterygota
and Neoptera was also obtained from our analyses. The monophyly of Collembola were veri-
fied, but the phylogenetic position of Diplura was challenged by our datasets. All the analyses
except for both ML analyses (with relatively low support values) showed that Diplura was a
paraphyletic group (S2 Fig. and S4 Fig.). Three dipluran species: Japyx solifugus, Campodea

![Phylogenetic tree](image1.png)

**Fig 1. Phylogenetic tree yielded by W12 dataset.** All indels after being aligned were reserved in this dataset. Phylogenetic analyses were based on first and second codon positions of 13 protein-coding genes. The tree was rooted with the sequences of two outgroup species: Onychiurus orientalis and Campodea lubbocki. From W12 dataset, the paraphyly of Machilidae and its subfamily Machilinae was obtained in all algorithms.

doi:10.1371/journal.pone.0117669.g001
lubbocki and Campodea fragilis were utilized for the phylogenetic analyses. *Japyx solifugus* seemed to have a closer relationship with Archaeognatha and formed a monophyletic group with true insects. The monophyly of Diplura had been questioned because the subgroups Cam­podeoidea and Japygoidea differed in cerci, sperm morphology and ovarian structure [62], but some analyses based on molecular data had led to incongruent results [63]. In addition, the phylogenetic position of Diplura was controversial [64]. Our phylogenetic estimate differed from that obtained by Luan et al. in their analysis of ribosomal RNA gene sequences. They inferred the relationships (Collembola + (Protura + Diplura)), with Diplura a monophyletic group, which was in accordance with our ML tree outgrouped by two Crustacean species (nodal support was relatively low in our analysis). Most other researches based on molecular data supported the monophyly of Diplura. Gao et al. [65] also retrieved a (Collembola +

**Fig 2.** Phylogenetic tree yielded by G12 dataset. All conserved regions after being aligned were retained. Phylogenetic analyses were based on first and second codon positions of 13 protein-coding genes. The tree was rooted with the sequences of two outgroup species: *Onychiurus orientalis* and *Campodea lubbocki*. Almost the same topologies were obtained in G12 dataset compared with W12 dataset.

doi:10.1371/journal.pone.0117669.g002
Fig 3. Phylogenetic relationships among basal Hexapods inferred from Maximum parsimony of 13 protein-coding sequences. 1 Protura species, 10 Collembola species and 3 Diplura species were used as the outgroup. Numbers denoted bootstrap values in percentages.

doi:10.1371/journal.pone.0117669.g003
Fig 4. Phylogenetic relationships among basal Hexapods inferred from Bayesian analysis of 13 protein-coding sequences. 1 Protura species, 10 Collembola species and 3 Diplura species were used as the outgroup. Numbers denoted posterior probabilities of nodes.

doi:10.1371/journal.pone.0117669.g004
Fig 5. Phylogenetic relationships among basal Hexapods inferred from Maximum parsimony analysis of protein-coding gene sequences. 2
Crustacean species Capitulum mitella and Pollicipes mitella were used as the outgroup. Numbers denoted bootstrap values in percentages.

doi:10.1371/journal.pone.0117669.g005
Fig 6. Phylogenetic relationships among basal Hexapods inferred from Bayesian analysis of protein-coding gene sequences. 2 Crustacean species *Capitulum mitella* and *Pollicipes mitella* were used as the outgroup. Numbers denoted posterior probabilities of nodes.

doi:10.1371/journal.pone.0117669.g006
(Protura + Diplura)) clade using 18S and 28S rRNA gene sequences. But he stated that the non-stationarity of nucleotide frequencies among dipluran and proturan sequences because of the high content of GC, so their result might not be solid. Given the situation above, we suggested a more detailed research, including more dipluran taxa and more characters, would be necessary for a better understanding of the phylogenetic relationships of Diplura and of the subgroups within basal Hexapoda.

Supporting Information

S1 Fig. Phylogenetic relationships among basal Hexapods inferred from Neighbor-Joining analysis of 13 protein-coding gene sequences. 1 Protura species, 10 Collembola species and 3 Diplura species were used as outgroup.

S2 Fig. Phylogenetic relationships among basal Hexapods inferred from Maximum likelihood analysis of 13 protein-coding gene sequences. 1 Protura species, 10 Collembola species and 3 Diplura species were used as outgroup.

S3 Fig. Phylogenetic relationships among basal Hexapods inferred from Neighbor-Joining analysis of protein-coding gene sequences. 2 Crustacean species Capitulum mitella and Pollicipes mitella were used as outgroup.

S4 Fig. Phylogenetic relationships among basal Hexapods inferred from Maximum likelihood analysis of protein-coding gene sequences. 2 Crustacean species Capitulum mitella and Pollicipes mitella were used as outgroup.

S1 Table. List of universal primers used for PCR amplification.

Acknowledgments

We are grateful to Weiwei Wang (Zhejiang Normal University) for his species collection, to Yubo Lin (Zhejiang Normal University) for assistance with the experiments. We also thank the anonymous reviewers for their important comments on the manuscript.

Author Contributions

Conceived and designed the experiments: JZ DY KH. Performed the experiments: YM KH PY XC. Analyzed the data: YM KH DY. Contributed reagents/materials/analysis tools: YM KH JZ. Wrote the paper: YM KH JZ. Collected samples: JZ.

References

1. Grimaldi DA (2010) 400 million years on six legs: On the origin and early evolution of Hexapoda. Arthropod Struct Dev 39: 191–203. doi: 10.1016/j.asd.2009.10.008 PMID: 19883792
2. Podsiadlowski L (2006) The mitochondrial genome of the bristletail Petrobius brevistyliis (Archaeognatha: Machilidae). Insect Mol Biol 5: 253–258.
3. Sturm H, Machida R (2001) Handbook of Zoology. Archaeognatha. Berlin: Walter de Gruyter. pp. 1–213.
4. Sturm H (1984) Zur systematik, biogeographic and evolution der sudamerikanischen Meinertellidae (Machiloidea, Archaeognatha, Insecta). J Zool Syst Evol Res 22: 27–44.
5. Kaplin VG (1985) On the classification and phylogeny of the Machilidae (Thysanura, Microcoryphia). Entomol Rev 64: 117–131.

6. Zhang JY, Zhou KY (2011) Descriptions of one new genus and six new species of Machilidae (Insecta: Archaeognatha) from China: morphological and molecular data. J NAT HIST 45: 19–20.

7. Hennig W (1953) Kritische Bemerkungen zum phylogenetischen System der Insekten. Beiträge zur Entomologie 3: 1–85. doi:10.1007/s00701-009-0297-2 PMID: 19387542

8. Hovmöller R, Pape T, Källersjö M (2002) The Palaeoptera problem: basal Pterygote phylogeny inferred from 18S and 28S rDNA sequences. Cladistics 18(3): 313–323.

9. Sharov AG (1966) Basic Arthropodan stock: with special reference to insects. Oxford: Pergamon Press. 271 p.

10. Kristensen NP (1997) The ground plan and basal diversification of the hexapods. In Fortey RA, Thomas RH, editors. Arthropod Relationships. London: Chapman & Hall. pp. 281–293.

11. Koch M (2003a) Character evolution in the Archaeognatha: consensus and conflict. Entomologische Abhandlungen 61: 120–122.

12. Koch M (2003b) Towards a phylogenetic system of the Zygentoma. Entomologische Abhandlungen 61: 122–125.

13. Mendes LF (2002) Taxonomy of Zygentoma and Microcoryphia: historical overview, present status and goals for the new millennium. Pedobiologia 46: 225–233.

14. Sturm H, Bach de Roca C (1993) On the systematics of the Archaeognatha (Insecta). Entomol Gen 18: 55–90.

15. Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. P Natl Acad Sci-Biol 76: 1967–1971.

16. Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27: 1767–1780. PMID: 10101183

17. Lin CP, Danforth BN (2004) How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol Phylogenet Evol 30: 686–702. PMID: 15012948

18. Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity 101: 301–320. doi: 10.1038/hdy.2008.62 PMID: 18612321

19. Dowton M, Castro LR, Austin AD (2002) Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates the examination of genome ‘morphology’. Invertebr Syst 16: 345–356.

20. Boore JL, Macey JR, Medina M (2005) Sequencing and comparing whole mitochondrial genomes of animals. Method Enzymol 395: 311–348. PMID: 15865975

21. Boore JL (2006) The use of genome-level characters for phylogenetic reconstruction. Trends Ecol Evol 21: 439–446. PMID: 16762445

22. Masta SE, Boore JL (2008) Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. Mol Biol Evol 25: 949–959. doi: 10.1093/molbev/msn051 PMID: 1829699

23. Cameron SL (2014) Insect mitochondrial genomics: Implications for evolution and phylogeny. Annu Rev Entomol 59: 95–117. doi: 10.1146/annurev-ento-011613-162007 PMID: 24160435

24. He K, Zhang JY, Deng KZ, Chen Z (2013) The complete mitochondrial genome of the bristletail Song-machilis xinxiangensis (Archaeognatha: Machilidae). Mitochondr DNA 24(2): 99–101.

25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739. doi: 10.1093/molbev/msr121 PMID: 21546353

26. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964. PMID: 9023104

27. Zhang JY, Song DX, Zhou KY (2008) The complete mitochondrial genome of the bristletail Pedetontus silvestrii (Archaeognatha: Machilidae) and an examination of mitochondrial gene variability within four bristletails. Ann Entomol Soc Am 101: 1131–1136.

28. Carapelli A, Liò P, Nardi F, Wath E, Frati F (2007) Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol Biol 7(Suppl 2): S8. PMID: 17767736

29. Cameron SL, Miller KB, D’Haese CA, Whiting MF, Barker SC (2004) Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). Cladistics 20: 534–557.

30. Cook CE, Yue Q, Akam M (2005) Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. P Roy Soc B-Biol Sci 272(1569): 1295–1304. PMID: 16024395
31. Ninomiya T, Yoshizawa K (2009) A revised interpretation of the wing base structure in Odonata. Syst Entomol 34: 334–345.
32. Comandi S, Carapelli A, Podsiadlowski L, Nardi F, Frati F (2009) The complete mitochondrial genome of Atelura formicaria (Hexapoda: Zygentoma) and the phylogenetic relationships of basal insects. Gene 439: 25–34. doi: 10.1016/j.gene.2009.02.020 PMID: 19285124
33. Zhang JY, Zhou CF, Gai YH, Song DX, Zhou KY (2008) The complete mitochondrial genome of Pararunurus voui (Insecta: Ephemeroptera) and the phylogenetic status of Ephemeroptera. Gene 424: 18–24. doi: 10.1016/j.gene.2008.07.037 PMID: 18725275
34. Lee EM, Hong MY, Kim MI, Kim MJ, Park HC, et al. (2009) The complete mitogenome sequences of the palaeopteran insets Ephemera orientalis (Ephemeroptera: Ephemeridae) and Davidius lunatus (Odonata: Gomphidae). Genome 52(9): 810–817. doi: 10.1139/g09-055 PMID: 19935929
35. Lin CP, Chen MY, Huang JP (2010) The complete mitochondrial genome and phylogenomics of a damselfly, Euphæa formosa suoopt a basal Odonata within the Pterygota. Gene 468: 20–29. doi: 10.1016/j.gene.2010.08.001 PMID: 20699111
36. Yamauchi MM, Miya MU, Nishida M (2004) Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. Insect Mol Biol 13: 435–442. PMID: 15271216
37. Chen AH (2013) Complete mitochondrial genome of the double-striped cockroach Blattella bisignata (Insecta: Blattaria). Mitochondr DNA 24: 14–16.
38. Wei SJ, Ni JF, Yu ML, Shi BC (2012) The complete mitochondrial genome of Macrotermes barneyi Light (Isoptera: Termitidae). Mitochondr DNA 23: 426–428.
39. Cameron SL, Barker SC, Whiting MF (2006) Mitochondrial genomics and the new insect order Mantophasmatodea. Mol Phylogenet Evol 38: 274–279. PMID: 16321547
40. Yang MR, Zhou ZJ, Chang YL, Zhao LH (2012) The mitochondrial genome of the quiet-calling katydids, Xiczus fascipes (Orthoptera: Tettigoniidae: Meconematinae). J Genet 91: 141–153. PMID: 22942084
41. Chen WJ, Bu Y, Carapelli A, Dallas R, Li S, et al. (2011) The mitochondrial genome of Sinentomon erythranum (Arthropoda: Hexapoda: Protura): an example of highly divergent evolution. BMC Evol Biol 11: 246. doi: 10.1186/1471-2148-11-246 PMID: 21871115
42. Carapelli A, Comandi S, Convey P, Nardi F, Frati F (2008) The complete mitochondrial genome of the Antarctic springtail Cryptopygus antarcticus (Hexapoda: Collembola), BMC Genomics 9: 315. doi: 10.1186/1471-2164-9-315 PMID: 18593463
43. Nardi F, Carapelli A, Fanciulli PP, Dallas R, Frati F (2001) The complete mitochondrial DNA sequence of the basal hexapod Tetrodontophora bielanensis: evidence for heteroplasmy and tRNA translocations. Mol Biol and Evol 18: 1293–1304. PMID: 11420368
44. Torricelli G, Carapelli A, Convey P, Nardi F, Boore JL, et al. (2010) High divergence across the whole mitochondrial genome in the “pan-Antarctic” springtail Friesea grisea: Evidence for cryptic species? Gene 449: 30–40. doi: 10.1016/j.gene.2009.09.006 PMID: 19782734
45. Carapelli A, Nardi F, Dallas R, Boore JL, Liò P (2005) Relationships between hexapods and crustaceans based on four mitochondrial genes, Crustac 16: 295–306.
46. Podsiadlowski L, Carapelli A, Nardi F, Dallas R, Koch M, et al. (2006) The mitochondrial genomes of Campodea fragilis and Campodea lubbocki (Hexapoda: Diplura): High genetic divergence in a morphologically uniform taxon. Gene 381: 49–61. PMID: 16919404
47. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552. PMID: 10742046
48. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92:371–373. PMID: 11535656
49. Swofford DL (2002) Paup*: Phylogenetic analysis using parsimony, version 4.0b10. Sunderland, MA: Sinauer Associates.
50. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755. PMID: 11524383
51. Clary DO, Wolstenholme DR (1985) The ribosomal RNA genes of Drosophila mitochondrial DNA. Nucleic Acids Res 13: 4029–4045. PMID: 2989784
52. Arnoldi FGC, Ogho K, Ohmiya Y, Viviani VR (2007) Mitochondrial genome sequences of the Brazilian luminescent click beetle Pyrophorus divergens (Coleoptera: Elateridae): Mitochondrial genes utility to investigate the evolutionary history of Coleoptera and its bioluminescence. Gene 405: 1–9. PMID: 17942246
53. Stewart JB, Beckenbach AT (2003) Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle Crioceris duodecimtincta. Mol Phylogenet Evol 26: 513–526. PMID:12644408
54. Li H, Liu H, Shi A, Styv P, Zhou X, et al. (2012) The Complete Mitochondrial Genome and Novel Gene Arrangement of the Unique-Headed Bug Stenopirates sp. (Hemiptera: Enicocephalidae). Plos One 7: 1–13.
55. Li H, Liu H, Cao L, Shi A, Yang H, et al. (2012) The Complete Mitochondrial Genome of the Damsel Bug Allooeorrhynchus bairi (Hemiptera: Nabidae). Int J Biol Sci 8: 93–107. PMID:22211108
56. Fenn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (Anabrus simplex: Tetingtonidae: Orthoptera) and an analysis of control region variability. Insect Mol Biol 16: 239–252. PMID:17316330
57. Wu YP, Li J, Zhao JL, Su Tj, Luo AR, et al. (2012) The complete mitochondrial genome of the rice moth, Corcyra cephalonica. J Insect Sci 12: 1–14. doi:10.1673/031.012.12601 PMID: 23465075
58. Miller AD, Nguyen TT, Burridge CP, Austin CM (2004) Complete mitochondrial DNA sequence of the Australian freshwater crayfish, Cherax destructor (Crustacea: Decapoda: Parastacidae): a novel gene order revealed. Gene 331: 65–72. PMID:15094192
59. Yamauchi M, Miya M, Nishida M (2002) Complete mitochondrial DNA sequence of the Japanese spiny lobster, Panulirus japonicus (Crustacea: Decapoda). Gene 295: 89–96. PMID:12242015
60. Flock PK, Rowell CH, Gellissen G (1995) The sequence, organization, and evolution of the Locusta migratoria mitochondrial genome. J Mol Evol 41: 928–941. PMID:8587138
61. Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470–474. PMID:7219536
62. Štys P, Biliński S (1990) Ovariole types and the phylogeny of hexapods. Biol Rev 65: 401–429.
63. Luan YX, Mallatt JM, Xie RD, Yin YM, Yin WY (2005) The phylogenetic positions of three basal-Hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. Mol Biol Evol 22: 1579–1592. PMID:15845456
64. Kukalová-Peck J (1987) New Carboniferous Diplura, Monura, and Thysanura, the hexapod ground plan, and the role of thoracic side lobes in the origin of wings (Insecta). Can J Zool 65: 2327–2345.
65. Gao Y, Bu Y, Luan YX (2008) Phylogenetic relationships of basal hexapods reconstructed from nearly complete 18S and 28S rRNA gene sequences. Zool Sci 25: 1139–1145. doi:10.2108/zsj.25.1139 PMID:19267625