Comparative mitogenome analyses uncover mitogenome features and phylogenetic implications of the subfamily Cobitinae

Peng Yu1,2, Li Zhou1,2, Wen-Tao Yang1,2, Li-jun Miao1,2, Zhi Li1, Xiao-Juan Zhang1, Yang Wang1,2* and Jian-Fang Gui1,2*

Abstract

Background: Loaches of Cobitinae, widely distributed in Eurasian continent, have high economic, ornamental and scientific value. However, the phylogeny of Cobitinae fishes within genera or family level remains complex and controversial. Up to now, about 60 Cobitinae mitogenomes had been deposited in GenBank, but their integrated characteristics were not elaborated.

Results: In this study, we sequenced and analyzed the complete mitogenomes of a female Cobitis macrostigma. Then we conducted a comparative mitogenome analysis and revealed the conserved and unique characteristics of 58 Cobitinae mitogenomes, including C. macrostigma. Cobitinae mitogenomes display highly conserved tRNA secondary structure, overlaps and non-coding intergenic spacers. In addition, distinct base compositions were observed among different genus and significantly negative linear correlation between AT% and AT-skew were found among Cobitinae, genus Cobitis and Pangio mitogenomes, respectively. A specific 3 bp insertion (GCA) in the atp8-atp6 overlap was identified as a unique feature of loaches, compared to other Cypriniformes fish. Additionally, all protein coding genes underwent a strong purifying selection. Phylogenetic analysis strongly supported the paraphyly of Cobitis and polyphyly of Misgurnus. The strict molecular clock predicted that Cobitinae might have split into northern and southern lineages in the late Eocene (42.11 Ma), furthermore, mtDNA introgression might occur (14.40 Ma) between ancestral species of Cobitis and ancestral species of Misgurnus.

Conclusions: The current study represents the first comparative mitogenomic and phylogenetic analyses within Cobitinae and provides new insights into the mitogenome features and evolution of fishes belonging to the cobitinae family.

Keywords: Cobitinae, Loach, Mitochondrial genome, mtDNA introgression, Phylogeny, Divergence time
Background

Vertebrate mitogenome is a small (16–17 kb) and circular double-stranded molecule [1]. It contains 37 genes including 22 tRNA genes, 13 PCGs and two rRNA genes [1]. It also has two noncoding regions, O1 and CR, and the latter contains regulatory elements for controlling the transcription and replication of mtDNA molecule [2, 3]. Due to its unique features, such as high copy numbers in tissues, simple genomic organization, maternal inheritance, almost unambiguous orthology, haploid inheritance and high nucleotide substitution rate [4–6], mitogenome has been widely applied in species identification, i.e., DNA barcoding, as well as population genetics, conservation biology, molecular phylogenetics and evolutionary processes [7–13]. Gene arrangements of fish mitogenomes are generally conserved, only with a few exceptions [1]. However, the genome sequence length, the bias of base composition and start/stop codon, the overlap and IGSs are diverse among different species [14].

Cobitinae is a subfamily of Cobitidae that was first identified by Hora (1932). To date, it contains 214 species recorded in FishBase, covering 21 genera, such as Cobitis, Misgurnus and Paramisgurnus [15]. Loaches of subfamily Cobitinae are bottom-dwelling fishes and widely distributed in Eurasian continent. They usually possess high economic, ornamental and scientific research value. Loach commercial farming, including cobitid loach (M. anguillicaudatus) and large-scale loach (P. dabyranus), occupies a significant position in freshwater aquaculture of Asia, due to their enjoyable taste, high nutritional value, rapid growth and strong adaptation [16–18]. In China, loach is used as a diet therapy or folk remedy for patient’s recovery or treatment of many diseases, such as hepatitis, osteomyelitis, carbuncles, and cancers. Many Cobitis populations are mixed diploid-polyploid, even bisexual and unisexual forms co-existing in the same niche [19–21]. They are suitable as models to reveal the relationship among hybridization, polyploidization, reproduction, speciation and evolution [21–23]. Due to their great diversity, they are also used to trace the biogeographic history of freshwater systems and to reflect geologic events [24].

Cobitinae fishes usually inhabit various benthic habitats in rivers, lakes, streams and ponds [25]. However, diapomorphism of the ecological environment has led to a decrease of benthic organisms [26, 27]. Cobitinae fishes are seriously threatened and their wild populations are gradually decreasing [28]. On this account, the diversity of these benthic fishes have been used as a bioindicator to assess the quality of the ecological environment [29, 30]. In addition, many Cobitinae species, such as the “kuhi loaches”, are well-known in Southeast Asia and Europe as ornamental fish for their varied morphological patterns and the ability to ingest bottom organic residues.

Cobitinae fishes are difficult to be classified because of their morphological similarity and high plasticity in morphology [31]. Although the secondary sexual dimorphism is used to define genera, it is not always congruent with the current genera definitions. The molecular phylogeny of Cobitinae fishes has been studied at the genera or family level via one or two mitochondrial and/or nuclear genes [24, 31–36], and remains complex and controversial. For example, based on mitochondrial gene cytb and nuclear gene rag-1, Perdices et al. (2016) [37] reconstructed the phylogenetic relationship of Northern Clade of family Cobitidae that inhabit in Europe, and North and Northwest parts of Asia. The subfamily Cobitinae was divided into Cobitis sensu lato group (Cobitis, Iksookimia, Niwaella and Kichulchoia), Misgurnus sensu lato group (Misgurnus, Paramisgurnus and Koreocobitis), Microcobitis, and Sabanejewia. Although the monophyly of the groups were resolved, the relationships within the groups are discordant with current taxonomic status.

Up to now, about 60 mitogenomes, covering more than 40 species of Cobitinae, have been deposited into GenBank [38–55]. Although a few mitogenomes characteristics were described, the integrated characteristics of Cobitinae mitogenomes are still not well known. In this study, we sequenced the mitogenome of C. macrostigma, the type species of the genus Cobitis [25], and compared it with other 41 species (57 individuals) to amplify detailed features of the Cobitinae mitogenomes. Additionally, we assembled a large sequence matrix (11,442 bp) of 58 Cobitinae mitogenomes and two outgroups to investigate the phylogenetic status and the origin time of Cobitinae fishes.

Results

General features of C. macrostigma mitogenome

The mitogenome of C. macrostigma was sequenced, annotated and compared with 57 Cobitinae mitogenomes (Table 1). It contains 13 PCGs (nd1–6, nd4l, cox1–3, cytb, atp6 and atp8), 22 tRNA genes, two rRNA genes (12S rRNA and 16S rRNA) and two non-coding regions (O1, and CR) (GenBank: MT259034). Gene order and orientation are same to most teleost mitogenomes (Fig. 1, Table 2). PCGs range from 168 bp (atp8) to 1551 bp (cox1) in size, with a total length of 11,427 bp. tRNAs vary from 66 bp (tRNAVal(Cys)(C)) to 76 bp (tRNAVal(K)) in size, with a total length of 1557 bp. The length of small encoding subunit 12S rRNA and large subunit 16S rRNA are 952 bp and 1675 bp, respectively. They are flanked by tRNAPro and tRNALeu(UUR) and interposed by tRNA-

Val. Among 58 mitogenomes analyzed, the entire mitogenome of C. macrostigma has the highest (99.6%)
Table 1: Species, GenBank accession number and length of mitogenomes used in this study

| Genus     | Species                      | Accession ID | Sequence length (bp) | Reference |
|-----------|------------------------------|--------------|----------------------|-----------|
| Cobitis   | Cobitis macrostigma          | MT259034     | 16,636               | this study|
|           | Acanthopsis choirorhynchos   | AB242161.1   | 16,600               | [38]      |
|           | Acanthoposoides              | NC_029438.1  | 16,603               | Unpublished|
|           | Canthophrys                  | NC_031576.1  | 16,561               | Unpublished|
|           | Cobitis biwae                | NC_027663.1  | 16,642               | [39]      |
|           | Cobitis choii                | NC_010649.2  | 16,566               | [40]      |
|           | Cobitis elongatoides         | NC_023947.1  | 16,541               | [41]      |
|           | Cobitis granoei              | NC_023473.1  | 16,636               | [42]      |
|           | Cobitis lutheri              | NC_022717.1  | 16,639               | Unpublished|
|           | Cobitis minamori minamori   | AP013309.1   | 16,645               | Unpublished|
|           | Cobitis matsubarai           | NC_029441.1  | 16,636               | Unpublished|
|           | Cobitis naltanti             | MH49461.1    | 16,631               | [43]      |
|           | Cobitis sp. (1)              | AP013307.1   | 16,671               | Unpublished|
|           | Cobitis sp. (2)              | AP013306.1   | 16,570               | Unpublished|
|           | Cobitis sp. (3)              | AP013296.1   | 16,576               | Unpublished|
|           | Cobitis striata (1)          | AP010782.1   | 16,646               | [44]      |
|           | Cobitis striata (2)          | AB054125.1   | 16,572               | [45]      |
|           | Cobitis striata striata      | AP013311.1   | 16,631               | Unpublished|
|           | Cobitis sinensis             | NC_007229.1  | 16,553               | Unpublished|
|           | Cobitis takatsuensis (1)     | AP009306.1   | 16,647               | [44]      |
|           | Cobitis takatsuensis (2)     | AP011290.1   | 16,578               | [39]      |
|           | Iksookimia                   | NC_027850.1  | 16,624               | Unpublished|
|           | Kichulchoia                  | AP011337.1   | 16,643               | Unpublished|
|           | Koreocobitis                 | HM535625.1   | 16,567               | Unpublished|
|           | Kottelatlimia                | NC_031597.1  | 16,588               | Unpublished|
|           | Lepidocephalichthys          | AP013313.1   | 16,337               | Unpublished|
|           | Lepidocephalichthys annandaei| AP013315.1   | 16,368               | Unpublished|
|           | Lepidocephalichthys guntea   | NC_031593.1  | 16,567               | Unpublished|
|           | Lepidocephalichthys hasselti | AP013334.1   | 15,897               | Unpublished|
|           | Lepidocephalichthys micropagon| NC_031595.1  | 16,608               | Unpublished|
|           | Lepidocephalichthys sp.      | AP013314.1   | 15,917               | Unpublished|
|           | Lepidocephalus               | NC_031596.1  | 16,556               | Unpublished|
| Misgurnus | Misgurnus anguillicaudatus (1)| KC823274.1   | 16,646               | [46]      |
|           | Misgurnus anguillicaudatus (2)| KM186181.1   | 16,645               | Unpublished|
|           | Misgurnus anguillicaudatus (3)| KC881110.1   | 16,643               | [47]      |
|           | Misgurnus anguillicaudatus (4)| KC734881.1   | 16,643               | [48]      |
|           | Misgurnus anguillicaudatus (5)| KC884745.1   | 16,644               | [47]      |
|           | Misgurnus anguillicaudatus (6)| MG938590.1   | 16,646               | Unpublished|
|           | Misgurnus anguillicaudatus (7)| KC509900.1   | 16,646               | [49]      |
|           | Misgurnus anguillicaudatus (8)| MF579257.1   | 16,647               | Unpublished|
|           | Misgurnus anguillicaudatus (9)| KC509901.1   | 16,646               | [49]      |
|           | Misgurnus anguillicaudatus (10)| KC762740.1    | 16,645               | [46]      |
|           | Misgurnus anguillicaudatus (11)| HM856629.1   | 16,634               | [50]      |
|           | Misgurnus anguillicaudatus (12)| AP011291.1   | 16,641               | [39]      |
|           | Misgurnus anguillicaudatus (13)| DQ026434.1   | 16,565               | [51]      |
similarity with C. granoei and lowest (88.2%) with C. sinensis.

**Highly conserved tRNAs secondary structure, overlaps and non-coding intergenic spacers among Cobitinae mitogenomes**

Cobitinae mitogenomes range from 16,337 bp (L. annandalei) to 16,647 bp (M. anguillicaudatus and C. takatsuensis) in length (Table 1). Their gene composition, gene arrangement and strand bias are highly conserved (Fig.1 and Table 2). Among the 22 tRNAs, due to the absence of DHU arm, tRNAser(AGN) (S1) is the only one that is not folded into the typical clover-leaf secondary structure (Fig. 2a). In the Cobitinae mitogenomes, unmatched base pairs are widespread among tRNAs. Taking C. macrostigma as an example, there are 446 base pairs among the 22 tRNAs, and only one gene (tRNALeu(-CUN)) possesses a fully paired stem. In the 425 base pairs of other 21 tRNAs, there are 43 (10.1%) unmatched base pairs that contain 28 noncanonical matches of G-U and 15 other mismatches, including A-C (7), A-A (1), C-C (2), C-U (2), and U-U (3) (Fig. 2a). Most of them are located in the acceptor, DHU and anticodon stems.

We also compared the gene overlaps and IGSs among 58 Cobitinae mitogenomes. Two long overlaps (atp8-atp6 and nd4l-nd4) and two long IGSs (O1 and tRNAArg-cod2) were found in Cobitinae mitogenomes. Highly conserved motifs “ATGCTAA” and “ATGGCAA-TAA” were found in the overlapped junctions between nd4l and nd4, and between atp8 and atp6, respectively (Fig. 3a). There are also several small overlaps between adjacent tRNA genes, such as tRNAile - tRNAArg and tRNAThr - tRNAPro. O1 is located within the five gene cluster (WANCY) (Table 2, Fig.1) and its secondary structure shows a stable stem-loop hairpin, which is strengthened by six C-G base pairs (Fig. 2b). Among the 31 bp of O1, the C-G base pairs on stems are highly conserved while the loops in the middle are variable (Fig. 3b). Another long IGS, between tRNAAsp and cox2, is also conserved in the 5′ and 3′ end, and highly variable in the middle.

CR, located between tRNAPro and tRNAAsp, is the most variable region in Cobitinae mitogenomes and ranges from 872 bp (Lepidocephalus macrochir) to 990 bp (C. takatsuensis) (Supplementary Table 2)[44]. Three domains are conserved and can be recognized in Cobitinae mitogenomes (Fig. 3c). They are terminal associated sequences (TAS), the central conserved-blocks (CSB-D, CSB-E and CSB-F) and conserved sequence blocks (CSB-1, CSB – 2 and CSB-3).

**Usage bias of start and stop codon, codon distributions and relative synonymous codons in Cobitinae mitogenomes**

The typical start codon ATG is conservative and is used in 12 PCGs, while GTG is only used in cox1 in 98% (57/58) analyzed Cobitinae mitogenomes except one individual of M. anguillicaudatus (No. 11) (Fig. 4, Supplementary Table 3). Five types of stop codons were found, containing three canonical (TAA, TAG and AGA) and two truncated stop codons (TA- and T--) (Fig. 4). The two truncated termination codons are used in nd2, cox2,
atp6, cox3, nd3, nd4 and cytb, the 3’-ends of which are followed by a tRNA gene encoded with the same strand.

The codon distribution and relative synonymous codon usage (RSCU) of 58 Cobitinae mitogenomes were analyzed. Our results show that codon distribution is largely coincident among these Cobitinae mitogenomes (Supplementary Figure S1). As shown by six representative species of Cobitinae, the codons encoding Leu(CUN), Ala and Thr are the three most frequently present, while those encoding Cys are rare (Fig. 5a). Compared to the other five Cobitinae species, P. anguillaris uses more codons of Leu(CUN) and less codons of Leu(UUR). The patterns of RSCU are also consistent among the analyzed species (Fig. 5b). Degenerated codons are biased to use more A/T than G/C in the 3rd position of PCGs, which results in the content of A + T is higher than G + C in the 3rd position of Cobitinae PCGs. For example, the codons for Arginine CCA and the codes for Tryptophan UGU are prevalent, while their other synonymous codons are relatively less used.

A + T %, AT-skew and their linear correlations of Cobitinae mitogenomes

The A + T content and AT-skew of whole mitogenomes, PCGs, tRNAs, rRNAs and CR were calculated (Fig. 6a-b). The 58 Cobitinae mitogenomes all exhibit AT bias, and the A + T content is the lowest (54.8 ± 0.6%) in tRNAs and the highest (66.3 ± 0.9%) in CR (Fig. 6a, Supplementary Table 2). The AT-skew values are the largest and positive in rRNAs, while they are the smallest in
| Feature         | Position (bp) | Nucleotide size (bp) | Start codon | Stop codon | Amino acid | Anti-codon | Intergenic nucleotidea | Strandb |
|-----------------|---------------|----------------------|-------------|------------|------------|------------|-------------------------|----------|
| tRNAPhe (S)     | 1–69          | 69                   |             |            |            |            |                         | H        |
| 12S rRNA        | 70–1021       | 952                  |             |            |            |            |                         | H        |
| tRNAArg (L1)    | 1022–1093     | 72                   |             |            |            |            |                         | H        |
| 16S rRNA        | 1094–2768     | 1675                 |             |            |            |            |                         | H        |
| tRNAGlu(L1)     | 2769–2843     | 75                   |             |            |            |            |                         | H        |
| nd1             | 2845–3819     | 975                  | ATG         | TAA        | 324        |            |                         | H        |
| tRNAAsp (S)     | 3826–3897     | 72                   |             |            |            |            |                         | H        |
| 12S rRNA        | 3896–3966     | 71                   |             |            |            |            |                         | L        |
| tRNAVal (V)     | 3968–4036     | 69                   |             |            |            |            |                         | H        |
| 16S rRNA        | 4037–5081     | 1045                 | ATG         | T          | 348        |            |                         | H        |
| tRNAThr (M)     | 5082–5151     | 70                   |             |            |            |            |                         | H        |
| tRNAGln (Q)     | 5153–5221     | 69                   |             |            |            |            |                         | L        |
| tRNALeu(UUR)    | 5223–5295     | 73                   |             |            |            |            |                         | L        |
| L-strand replication origin (OL) | 5296–5325 | 30                   |             |            |            |            |                         | L        |
| tRNAAsn (N)     | 5326–5391     | 66                   |             |            |            |            |                         | L        |
| tRNAArg (R)     | 5392–5460     | 69                   |             |            |            |            |                         | L        |
| cox1            | 5462–7012     | 1551                 | GTG         | TAA        | 516        |            |                         | H        |
| tRNAGlu(L2)     | 7014–7084     | 71                   |             |            |            |            |                         | L        |
| tRNAAsp (O)     | 7087–7158     | 72                   |             |            |            |            |                         | H        |
| cox2            | 7172–7906     | 735                  | ATG         | TAA        | 244        |            |                         | H        |
| tRNAGlu (S)     | 7933–8008     | 76                   |             |            |            |            |                         | L        |
| atp6            | 8010–8177     | 168                  | ATG         | TAA        | 55         | −10        |                         | H        |
| cox3            | 8168–8851     | 684                  | ATG         | TAA        | 227        | −1         |                         | H        |
| tRNAGlu (S)     | 8851–9634     | 784                  | ATG         | T          | 261        |            |                         | H        |
| tRNAArg (R)     | 9635–9706     | 72                   |             |            |            |            |                         | H        |
| nd3             | 9707–10,055   | 349                  | ATG         | T          | 116        |            |                         | H        |
| tRNAAsp (O)     | 10,056–10,125 | 70                   |             |            |            |            |                         | H        |
| nd4             | 10,126–10,422 | 297                  | ATG         | TAA        | 98         | −7         |                         | H        |
| tRNAAsp (H)     | 10,416–11,797 | 1382                 | ATG         | TA         | 460        |            |                         | H        |
| tRNAVal (N)     | 11,790–11,866 | 96                   |             |            |            |            |                         | H        |
| tRNAGlu (S1)    | 11,867–11,934 | 68                   |             |            |            |            |                         | H        |
| tRNAGlu (L2)    | 11,936–12,008 | 73                   |             |            |            |            |                         | H        |
| nd5             | 12,009–13,847 | 1839                 | ATG         | TAG        | 612        | −4         |                         | H        |
| nd6             | 13,844–14,365 | 522                  | ATG         | TAA        | 173        |            |                         | L        |
| tRNAGlu (E)     | 14,366–14,434 | 69                   |             |            |            |            |                         | L        |
| cyt b           | 14,441–15,581 | 1141                 | ATG         | T          | 380        |            |                         | H        |
PCGs and most are negative except *Canthophrys gongota*, *Acanтопis choirohynchos*, *P. cuneovirgata*, *P. kuhlii*, *P. oblonga*, and *Kottelatimia pristes* (Fig. 6, Supplementary Table 2). These results indicate that PCGs are biased towards using T not A in most Cobitinae mitogenomes. To examine whether the A + T content and AT-skew are different in three codon position of PCGs, we also selected the six Cobitinae species for a more detailed analysis. The A + T content shows 1st < 2nd < 3rd in the three position of PCGs in all analyzed fishes. Meanwhile, the AT-skew of 1st and 3rd are positive while 2nd is negative (Table 3). This is due to the bias usage of relative synonymous codons (Fig. 5b). In all analyzed Cobitinae mitogenomes, CRs possess more A and C with all AT-skew values positive (0.002–0.112) and GC-skew negative (−0.341–−0.101) (Supplementary Table 2).

The correlations of Cobitinae mitogenomes (yA1 = −0.0166x − 0.9047, R² = 0.5991) genus *Cobitis* (yA2 = −0.012x + 0.5786, R² = 0.5197) and *Pangio* (yA3 = −0.0466x + 2.5813, R² = 0.5486) were calculated between A + T % versus AT-skew. All of them showed negative linear correlations, implying that AT-skew becomes more positive with the increasing of A + T content (Fig. 6c). The similar negative linear correlations were also found in G + C % versus GC-skew (Fig. 6d).

### Non-synonymous and synonymous substitutions

To better understand the role of selective pressure and evolutionary relations of Cobitinae fishes, the φ or dN/dS value of each PCG was calculated (Fig. 7). All the PCGs evolved under a purifying selection (φ < 0.5). The *atp8* gene showed the highest φ value (φ = 0.12) and the *cox* family genes were lowest (φ = 0.02 ± 0.01). This phenomenon is also found in most Metazoa [56], but the fold change (> 10 fold) is particularly high in Cobitinae. The lower φ value represents less variations in amino acids. Thus, *cox1*, *cox3* and *cytb* are potential bar-coding markers for Cobitinae species identification.

### Phylogenetic analysis of Cobitinae fishes

Molecular phylogenetic analyses were performed using 13 PCGs from 58 Cobitinae mitogenomes, belonging to 41 species from 14 genera. The ML and BI analyses generated similar topology with high bootstrap support / posterior probability values. Each tree was similarly divided into two main clades: *Cobitis*-Misgurnus-other genera (clade I) and *Pangio*-Lepidocephalichthys-other genera (clade II) (Fig. 8 and Supplementary Figure S2). Clade I included all analyzed species of *Cobitis*, *Paramisgurnus* and *Misgurnus*, and five species from other genus (*I. longicorpa*, *K. multifasciata*, *N. delicate*, *K. naktonensis*, and *Microcobitis sp.*). Four *Pangio* species, five *Lepidocephalichthys* species and other five species (*K. pristes*, *A. choirohynchos*, *A. gracilentus*, *L. macrochir*, and *C. gongota*) were clustered into Clade II, among which the analyzed species of genus *Pangio* and *Lepidocephalichthys* formed two well-supported (pp = 1.00) monophyletic groups respectively. In addition, *Pangio* is the sister genus to *Lepidocephalichthys*.

The BI phylogenetic tree confirmed that *Cobitis* was a paraphyletic group, since *Misgurnus* clade A, *N. delicate*, *L. longicorpa*, and *K. multifasciata* shared the common ancestor with the all 15 *Cobitis* species analyzed in this study, with high posterior probability values (pp = 1.00). The species of *Misgurnus* were separated into two independent lineages: the majority of *M. anguillicaudatus* individuals (12/14) and *M. bipartitus* clustering with the *Cobitis* species (*Misgurnus* clade A), and two *M. anguillicaudatus* individuals, *M. mohoity*, and *M. nikolskyi* gathering with *P. dabryanus* and *K. naktongensis* (*Misgurnus* clade B).

### Divergence time estimation of Cobitinae fishes

The combination of strict clock model and Yule process tree prior provided the best fit to the data sets (Supplementary Table 4). The chronogram with divergence time of Cobitinae lineages was estimated based on the cytB mutation rate (0.68% per million years) (Fig. 9). The first split of Cobitinae lineages was estimated to have occurred in the late Eocene (42.11 Ma, 95% HPD: 36.35–47.86 Ma), then separated into clade I (northern clade) and clade II (southern lineages). *Cobitis-Ilsookimia-Kichulchoia-Niwaella* lineage diverged from the rest of northern clade lineage during the Oligocene (30.07 Ma, 95% HPD: 25.55–34.69 Ma), similar to the previous
described [35], then diversified and further radiated after 4.94 Ma. The mtDNA introgression between ancestral species of *Cobitis* and ancestral species of *Misgurnus* seems to have taken place in the Middle Miocene (14.40 Ma, 95% HPD: 12.30–16.54 Ma). *C. macrostigma* appeared about 0.36 Ma (95% HPD: 0.06–0.55 Ma) in the Pleistocene. *Pangio*-Lepidocephalichthys-other genera (southern lineages) might originate about 40.45 Ma. In
southern lineages, *Pangio* was estimated to have occurred about 20.14–29.88 Ma, and the divergence times of the four species analyzed in this study are congruent with the previous described dating [24].

**Discussion**

In this study, we conducted a comparative mitogenome analysis and revealed the conserved and unique characteristics of 58 Cobitinae mitogenomes. Cobitinae mitogenomes display highly conserved tRNA secondary structure, overlaps and non-coding intergenic spacers. Among the 22 tRNAs, tRNA_{ser(AGN)} (S1) is the only one that is not folded into the typical clover-leaf secondary structure (Fig. 2a). Loss of stem in S1 is common character among Cobitinae and other metazoan mitogenomes [57, 58]. Similarly, the widespread unmatched base pairs among Cobitinae tRNAs is also a conserved feature in the eukaryote mitogenome [59–61]. Although their functions are not clear in fish, the unmatched base pairs are considered as the current state of evolutionary and irreversible process, which might be caused by tRNA editing [62].

Like other cyprinid fishes [14, 63], two long overlaps and two long IGSs were found in Cobitinae mitogenomes. The motif “ATGCTAA” in *nd4l-nd4* was conserved in vertebrates, including fish, turtle and human.

**Fig. 3** Sequence logo of gene overlaps in *atp8-atp6* (a), non-coding intergenic spacers in *tRNA^{AGN}_{Asp}-cox2* (b) and short conserved motif in CR of Cobitinae mitogenomes (c).

**Fig. 4** Usage bias of start and stop codons of 13 PCGs in Cobitinae mitogenomes. Pie graphs show the use frequency of start and stop codons. Gene abbreviations are the same as Table 2.
Fig. 5 Codon distribution (a) and relative synonymous codon usage (b) of PCGs in *C. macrostigma* and other five representative species of Cobitinae. CDPₜ = codons per thousand codons.
However, in comparison with the conserved motif (ATGATAA) in other Cypriniformes fishes, there is a specific 3 bp insertion (GCA) in the atp8-atp6 overlap motif of Cobitinae and other loaches [67–69], indicating this insertion is a characteristic feature of loaches. IGSs are important for transcription and associated with gene rearrangement in insects [70–72]. It is commonly assumed that IGS had a rapid nucleotide substitution rate under relaxed selection [73]. Moreover, Cobitinae mitogenomes share highly conserved sequences in IGSs that are immediately adjacent to tRNAs, such as “CTTTCGCCG”, “AAGCGGGA” and “AGC”. Whether these conserved sequences have a function or not and how they act awaits further investigation. As the longest IGSs, CR plays an important role in controlling the transcription and replication of mtDNA molecule by...
several domains and motifs [74, 75]. Although significant length variation were found in CR of vertebrate [76], the three domains can also be recognized in Cobitinae mitogenomes. Furthermore, the AT-skew and GC-skew of CR might reflect the strand asymmetry [77–79]. In teleost, the skew inversion of CR was only found in the mitogenomes of *Albula glossodonta* and *Bathygadus antrode*, showing a reversed strand asymmetry [75]. The normal Cobitinae mitogenomes CR skewness indicates that the strand asymmetry is not reversed.

The phylogenetic analyses show the monophyly of the genus *Pangio* and *Lepidocephalichthys*, consistent with the previous study [35]. However, *Cobitis*, the biggest genus of Cobitinae [15], is a complex and controversial
paraphyletic group. Similar to the trees constructed by cyt b [25, 80, 81], Iksookimia, Kichulchoia and Niwaella species were nested within Cobitis, implying a close relationship among them. Perdices [37] proposed that these species of Iksookimia, Kichulchoia, and Niwaella might belong to genus Cobitis, as morphologically specialized species derived from a local Cobitis species. However, this assumption awaits more morphological, karyological and molecular investigation. In addition, our phylogenetic analysis confirmed the assumption that M. mizolepis and P. dabryanus are conspecific [33, 80] and the different lineages under the species name C. striata and C. takatsuensis might actually represent different species.

The species of Misgurnus were separated into two independent clade and clustered into Cobitis species and P. dabryanus-K. naktongensis, respectively. The same results were observed in the trees based on the cyt b [80] and 13 PCGs from 28 cobitidae species [47]. However, all Misgurnus and Koreocobitis species were grouped into a monophyletic clade when their phylogenetic relationships were constructed by nuclear gene rag-1 [80]. This incongruity between mitochondrial and nuclear gene trees was explained by the different evolutionary rate of markers, hybridization or introgression [82]. It is commonly believed that hybridization and subsequent mtDNA introgression might occur between ancestral species of Cobitis and ancestral species of Misgurnus [35, 80]. In this study, we collected 14 mitogenomes from M. anguillicaudatus, which were divided into two genetically divergent clades. The similar phenomenon has been reported by several previous studies, which is explained by hybridization and mtDNA introgression [34, 35, 47, 83, 84]. Considering that M. anguillicaudatus clustered into the clade of Misgurnus and Koreocobitis by nuclear analyses [80], we supposed that the 12 mitogenomes (No. 1–12) of M. anguillicaudatus in Misgurnus clade A could be considered as the introgressed mtDNA type because of their close relationship with Cobitis species, whereas the other two individuals in Misgurnus clade B retained the original M. anguillicaudatus mitogenomes. M. anguillicaudatus with introgressed mtDNA type spread over most of East Asia, including China, Japan and Korea. M. anguillicaudatus shows extensive ploidy variability in nature. Besides most common diploid individuals (2n = 50), triploid (3n = 75) and tetraploid (4n = 100) have been frequently recorded in some localities of China and Japan [21, 47, 85, 86]. Rare pentaploid (5n = 125) and even hexaploid (6n = 150) individuals were found in the Yangtze River basin [87]. All of M. anguillicaudatus polyploids analyzed in this study belonged to the introgressed mtDNA type. Since mtDNA is inherited maternally, these polyploids might have originated from the diploid M. anguillicaudatus with introgressed mtDNA. Further analyses are needed to confirm this hypothesis of inter-genus mtDNA introgression based on a large-scale sampling with quantitative morphological features, definite ploidy, and more genes from both mitochondria and nuclear genomes.

The first split of Cobitinae lineages was estimated to have occurred in the late Eocene (42.11 Ma, 95% HPD: 36.35–47.86 Ma), separating northern clade and southern lineages, consistent with reconstruction dates of the paleo-drainages of East Asia [35, 88]. Cobitinae fishes in Clade I and Clade II, nominated as “northern clade” and “southern lineages” respectively, show a distinct disjunctive distribution with a small area of sympatry in Vietnam [35]. Consistent with their locations, the northern clade
spread to most of East Asia, Siberia and Europe, while the southern lineages distribute across the Indian subcontinent and Southeast Asia after their isolation. The nodes within northern clade and southern lineages appear asynchronous, implying that some local dominant factors, rather than large-scale events, might shape the evolution within northern or southern lineage.

**Conclusions**

This study represents the first comparative mitogenome and phylogenetic analyses within Cobitinae. The conserved and unique characteristics of 58 Cobitinae mitogenomes were revealed. We observed distinct base compositions among different genus and identified a specific 3 bp insertion (GCA) in the atp8-atp6 overlap as a unique feature of loaches. ML and BI analyses both strongly support the paraphyly of Cobitis and polyphyly of Misgurnus. In addition, Cobitinae might have split into northern and southern lineages in the late Eocene (42.11 Ma), and a mtDNA introgression between Cobitis and Misgurnus might have occurred about 14.40 Ma. The current study provides new insights into the mitogenome features and evolution of Cobitinae fishes.

**Methods**

**Sampling, sequencing and assembly**

The *C. macrostigma* analyzed in this study was caught from the Yangtze River in Yibin City, Sichuan Province, China (N: 28°46′6.01″, E: 104°38′13.99″) in October 2018 and five individuals were transported to the laboratory (National Aquatic Biological Resource Center, NABRC) in oxygen-rich water. It possesses 5–9 large and round spot in the midline of lateral body side [89] (Fig. 1). Before sampling, they were reared in a square...
and glass recirculating freshwater tanks with a volume of about 100 L, at 22 °C on a 14 h (hour) light/10 h dark cycle for morphological identification. After deep and overdosed anesthesia with styrylpyridine (a common anesthetic used in fish, 30-50 mg/L; aladdin, China), one healthy one-year-old female fish, 7 cm in length and 1.8 g in weight, was euthanized by immediately cutting off the spinal cord adjacent to the head. Total DNA was extracted according to the Ezup Column Animal Genomic DNA Kit technical manual (Sangon, Shanghai, China). PCR primers were designed based on the conserved sequences between the mitogenomes of *C. granoei* (GenBank: NC_023473.1) and *C. sinensis* (GenBank: NC_007229.1). 742–2495 bp DNA were amplified by using High Fidelity DNA Polymerase (Yeasen, Shanghai) (Supplementary Table 1). To obtain accurate sequences, we chose a cloning strategy. According to manual, PCR amplicon was purified, ligated ESI-Blunt vector (Yeasen, Shanghai) and transfected into 5α Chemically Competent Cell (Tsingke Biological Technology, Beijing). The positive clones were sequenced by Quintara Biosciences (Wuhan, China). The segments, longer than 1500 bp,
were sequenced using the primer walking sequencing strategy. The resulting DNA sequences were assembled using DNAStar (DNASTAR Inc., USA) [90]. Other 57 Cobitinae mitogenomes were download from NCBI GenBank database [38–55].

**Gene annotation and bioinformatic analyses**

tRNA genes and their secondary structures were predicted with MITOS [91] and tRNAscan-SE 2.0 with default parameters [92]. All 13 PCGs and two rRNA genes were annotated by comparison with the sequences of other Cobitinae fishes in GenBank (https://blast.ncbi.nlm.nih.gov/). The mtDNA maps were drawn using CGView Server 1.0 [93]. The sequence logos of gene overlaps and non-coding IGSs were drawn using WebLogo 3.7.4 [94]. The base composition, codon distributions and relative synonymous codons usage were calculated using DNAStar (DNASTAR Inc., USA) [90], MEGA 7.0 [95] and Microsoft Excel 2010. Skewness was measured using the formulas: AT-skew = (A% - T%) / (A% + T%) and GC-skew = (G% - C%) / (G% + C%) [79]. The similarity of the sequences was calculated in MEGA 7.0 [95] under p-distance and NCBI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

**Phylogenetic analyses**

The phylogenetic analysis was performed based on 13 PCGs of 58 Cobitinae mitogenomes. *Sinorhodeus microlepis* and *Rhodeus shihtaiensis* were chosen as the outgroups (Table 1). Each of the 13 gene sequences was separately aligned using Muscle v3.8.31 [96] and concatenated into a sequence matrix by PhyloSuite v1.2.2 [97]. Then PartitionFinder2 [98] was used to find the best partitioning strategy and to calculate the best-fit evolutionary model for each subset. For the alignment, a scheme with eight partitions was selected and GTR + G + I was chosen as the best-fit evolutionary model for each partition. Phylogenetic trees were constructed by the maximum likelihood (ML) method and bayesian inference (BI). The ML method was implemented in RAxML v8.2.12 [99]. Each partition scheme was run with the GTR+GAMMA1 model, and 1000 rapid bootstrapping replications were set to evaluate the bootstrap support values and search for the best-scoring ML tree. The BI phylogeny was performed in MrBayes v3.1.2 [100] with the “unlink” and "prest ratepr = variable" model parameters. 10,000,000 generations were run in two independent runs of four independent Markov Chain Monte Carlo (MCMC) chains, and were sampled every 1000 generations. The convergence of the BI analyses was investigated using Tracer v1.7.1 software. The first 2500 trees were discarded as conservative burn-in, and the rests were used to generate a majority rule consensus tree.

In cobitid fishes, 0.680% (divergence per pairwise comparison per Ma) was calculated and suggested for the mutation rates of *cytb* gene [32]. In this study, BEAST v1.10.4 [101] was used to estimate the divergence time with the rate (0.68%). GTR + G + I was chosen as the best fit model by PartitionFinder2 [98]. The best-fit clock type and tree prior were selected from two clock models (strict clock and uncorrelated relaxed clock) and four tree priors (Yule process, Exponential growth, Constant size and Bayesian skyline) by comparing the marginal likelihood values estimated by path sampling [102]. The analyses were simultaneously run for 20,000,000 generations, with parameters sampled every 1000, then the first 25% of the trees were discarded as burn-in. Tracer v1.5 [103] and Figtree were used to assess the convergence and view trees, respectively.

**Supplementary Information**
The online version contains supplementary material available at https://doi.org/10.1186/s12864-020-07360-w.

**Additional file 1:** Table S1. List of primers used to amplify the mitogenome of *C. macrostigma*.

**Additional file 2:** Table S2. Length, base composition and skewness of Cobitinae fish mitogenomes.

**Additional file 3:** Table S3. Start and stop codons of 13 PCGs in Cobitinae mitogenomes.

**Additional file 4:** Table S4. Marginal likelihood values of different combinations of clock model and tree prior.

**Additional file 5:** Figure S1. Codon distribution (A) and relative synonymous codon usage (B) of PCGs in the 58 Cobitinae mitogenomes. CDpT = codons per thousand codons.

**Additional file 6:** Figure S2. Phylogenetic tree constructed by ML methods, based on 13 PCGs of 58 Cobitinae mitogenomes. *Sinorhodeus microlepis* and *Rhodeus shihtaiensis* were chosen as outgroups. Node numbers represent the bootstrap value.

**Abbreviations**

mitogenome: Mitochondrial genome; mtDNA: Mitochondrial DNA; Ma: Million years ago; tRNA: Transfer RNA; PCG: Protein coding gene; rRNA: Ribosomal RNA; O$_2$: Origin of L-strand replication; CR: Control region; cytb: Cytochrome b; rag-1: Recombination activating gene 1; bp: Base pair; nad1–6: NADH dehydrogenase subunit 1–6; nd4L: NADH dehydrogenase subunit 4 L; cox1–3: Cytochrome oxidase subunit I–III; atp6: ATPase subunit 6; atp8: ATPase subunit 8; DHU: Dihydrouracil; IGS: Non-coding intergenic spacer; RSCU: Relative synonymous codon usage; w or dN/dS: Non-synonymous and synonymous substitutions; HPD: Highest posterior density; ML: Maximum likelihood; BI: Bayesian inference

**Acknowledgments**

The research was supported by the Wuhan Branch, Supercomputing Centre, Chinese Academy of Sciences, China. The research was supported by the Wuhan Branch, Supercomputing Centre, Chinese Academy of Sciences, China.

**Authors’ contributions**

P Y and L Z conceived and designed the study. P Y, Y W, WT Y, LJ M, Z L and XJ Z collected the samples and analyzed the data. P Y and L Z wrote the draft manuscript, Y W and JF G revised the manuscript. All authors have read and approved the final manuscript.

**Funding**

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000) and the China Agriculture Research System (CARS-45-07).
Availability of data and materials
C. macrostigma mitochondrial genome has been deposited in GenBank under the accession numbers MT259034. The 59 mitogenomes from Cobitinae species, Sinorhodeus microlepis and Rhodoxus shihtaiensis were downloaded from GenBank. Their accession numbers and references were listed in Table 1. Other supporting results are included within the article and its additional files.

Ethics approval and consent to participate
The C. macrostigma analyzed in this study was caught from the Yangtze River in Yibin City, Sichuan Province, China and reared in the National Aquatic Biological Resource Center (NABRC). The acquisition of experimental fish complies with the laws of Fishery Administration of the Ministry of Agriculture and Rural Affairs of the People’s Republic of China. We confirm that C. macrostigma is not an endangered or protected species (http://www.iucnredlist.org). The protocol, including the research, question, key design features, and analysis plan, was provided to the Animal Care and Use Committee of Institute of Hydrobiology, Chinese Academy of Sciences before the study, and all procedures in this research performed with the approval of the Committee. No ethics approval was required for the public sequence data used in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
1State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, the Innovation Academy of Seed Design, Chinese Academy of Sciences, Wuhan 430072, China. 2University of Chinese Academy of Sciences, Beijing 100049, China.

Received: 28 May 2020 Accepted: 29 December 2020
Published online: 14 January 2021

References
1. Satoh TP, Miya M, Mabuchi K, Nishida M. Structure and variation of the mitochondrial genome of fishes. BMC Genomics. 2016;17:719.
2. Cuppari A, Fernandez-Millan P, Battistini F, Tares-Sole A, Lyonnais S, Iruela G, et al. DNA specificities modulate the binding of human transcription factor a to mitochondrial DNA control region. Nucleic Acids Res. 2019; 47(12):6519–37.
3. Shadel GS, Clayton DA. Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem. 1997;66:409–35.
4. Lin CP, Danforth BN. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol Phylogenet Evol. 2004;30(3):686–702.
5. Elson JL, Lightowlers RN. Mitochondrial DNA cloning in the dock: can surveillance swing the case? Trends Genet. 2006;22(11):603–7.
6. Curole JP, Kocher TD. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol Evol. 1999;14(10):394–8.
7. Cole TL, Ksepka DT, Mitchell KJ, Tennyson AJD, Thomas DB, Pan HL, et al. Mitogenomes uncover extinct penguin taxa and Reveal Island formation as a key driver of speciation. Mol Biol Evol. 2019;36(4):784–97.
8. Baron S, van der Merwe NA, Maritz-Olivier C. The genetic relationship between R. microplus and R. decoloratus ticks in South Africa and their population structure. Mol Phylogenet Evol. 2018;129:60–9.
9. Min-Shan Ko A, Zhang Y, Yang MA, Hu Y, Cao P, Feng X, et al. Mitochondrial genome of a 22,000-year-old giant panda from southern China reveals a new panda lineage. Curr Biol. 2018;28(12):6934–8.
10. Wei M, Yu P, Yang Y, Wan Q. The complete mitochondrial genome of Leptobotia taiyangi (Cypriniformes: Cobitidae). Mitochondrial DNA part a, DNA mapping, sequencing, and analysis. Mitochondrial DNA A DNA Mapp Seq Anal. 2016;27(6):1707–8.
11. Luo H, Hong X, Chen S, Shi W. Mechanisms of gene rearrangement in 13 borthids based on comparison with a newly completed mitogenome of the three-spotted flounder, Grammatothoys polyophthalmus (Pleuronectiformes: Bothidae). BMC Genomics. 2019;20(1):792.
12. Dong S, Zhao C, Zhang S, Zhang L, Wu H, Liu H, et al. Mitochondrial genomes of the early land plant lineage liverworts (Marchantiophyta): conserved genome structure, and ongoing low frequency recombination. BMC Genomics. 2019;20(1):953.
13. Ponts N, Gautier C, Gouzy J, Pinson-Gadais L, Foulongne-Oriol M, Ducos C, et al. Evolution of Fusarium tricinctum and Fusarium avenaceum mitochondrial genomes is driven by mobility of introns and of a new type of palindromic microsatellite repeats. BMC Genomics. 2020;21(1):358.
14. Yu P, Zhou L, Zhou XY, Yang WT, Zhang J, Zhang XJ, et al. Unusual AT-skew of Sinorhodeus microlepis mitogenome provides new insights into mitogenome features and phylogenetic implications of baffling fishes. Int J Biol Macromol. 2019;129:35–50.
15. Froese R, Pauly D. FishBase: World wide web electronic publication. Wwfishbaseorg: 2019. version 04.
16. Gao J, Koshio S, Nguyen BT, Wang WM, Cao XJ. Comparative studies on lipid profiles and amino acid composition of wild and cultured dojo loach Misgurnus anguillicaudatus obtained from southern Japan. Fish Sci. 2012; 78(6):1331–6.
17. Chen J. Chen L. Effects of chitosan-supplemented diets on the growth performance, nonspecific immunity and health of loach fish (Misgurnus anguillicaudatus). Carbohydr Poly. 2019;225:115227.
18. Gao Y, He J, He Z, Li Z, Zhao B, Mu Y, et al. Effects of fulvic acid on growth performance and intestinal health of juvenile loach Paramisgurnus dabryanus (Sauvage). Fish Shellfish Immunol. 2017;72:427–56.
19. Juchno D, Jablonska O, Boron A, Kujawa R, Leska A, Grabowska A, et al. Ploidy-dependent survival of progeny arising from crosses between natural allotriploid Cobitis females and diploid C-taenia males (Pisces, Cobitidae). Genetica. 2014;142(4):351–9.
20. Morishima K, Yoshikawa H, Ara K. Diploid clone produces unreduced diploid gametes but tetraploid clone generates reduced diploid gametes in the Misgurnus loach. Biol Reprod. 2012;86(2):333.
21. Zhou L, Gai J. Natural and artificial polyploidy in aquaculture. Aquaculture. 2017;323:103–11.
22. Janko K, Culling MA, Rab P, Kotlik P. Ice age clonogen—comparison of the quaternary evolutionary histories of sexual and clonal forms of spiny loaches (Cobitis; Teleostei) using the analysis of mitochondrial DNA variation. Mol Ecol. 2005;14(10):2991–3004.
23. Janko K, Böhlen J, Lamatsch F, Flajshans M, Epple JB, Rab P, et al. The gynogenetic reproduction of diploid and triploid hybrid spined loaches (Cobitis; Teleostei), and their ability to establish successful clonal lineages—on the evolution of polyploidy in asexual vertebrates. Genetica. 2007;131(2): 185–94.
24. Böhlen J, Sterchová V, Tan HH, Britz R. Phylogeny of the southeast Asian freshwater fish genus Pangio (Cypriniformes, Cobitidae). Mol Phylogenet Evol. 2011;61(3):854–65.
25. Chen YX, Chen YF. Three new species of cobitid fish (Teleostei, Cobitidae) from the river Xinjiang and the river Le’anjiang, tributaries of Lake Poyang of China, with remarks on their classification. Folia Zool. 2013;62(2):883–93.
26. Peng G, Zhu B, Yang D, Su L, Shi H, Li D. Microplastics in sediments of the Changjiang estuary, China. Environ Pollut. 2017;225:283–90.
27. Wang J, Wang M, Ru S, Liu X. High levels of microplastic pollution in the sediments and benthic organisms of the South Yellow Sea, China. Sci Total Environ. 2019;651(Pt 2):1661–9.
28. Bui I, Caleta M, Marcic Z, Sanda R, Vukic J, Mrakovic M. Different histories, different Distinctiesm mappings of evolutionary history and population genetic structure on extinction risk of the Adriatic Spined loaches (genus Cobitis; Cypriniformes, Apogoniphilus). PLoS One. 2015;10(7):e0131580.
29. Zhao S, Feng C, Quan W, Chen X, Niu J, Shen Z. Role of living environments in the accumulation characteristics of heavy metals in fishes and crabs in the Yangtze River estuary, China. Mar Pollut Bull. 2012;64(6):1163–71.
30. Allert AL, Fairchild JF, Schmitt CJ, Besser JM, Brumbaugh WG, Olson SJ. Effects of mining-derived metals on riffle-dwelling benthic fishes in Southeast Missouri, USA. Ecotoxicol Environ Saf. 2009;72(6):1642–51.
31. Perdices A, Böhlen J, Doadrio I. The molecular diversity of adriatic spined loaches (Teleostei, Cobitidae). Mol Phylogenet Evol. 2008;46(1):35–46.
32. Doadrio I, Perdices A, Bohlen J. The molecular diversity of adriatic spined loaches (Teleostei, Cobitidae) based on cytochrome b sequence data. Mol Phylogenet Evol. 2005;37(2):484–93.
33. Stechová V, Böhlen J, Tan HH, Families of Cobitoidea (Teleostei; Cypriniformes) as revealed from nuclear genetic data and the position of
