The Mitochondrial Genome of *Erpobdella octoculata* (Hirudinida: Arhynchobdellida), as the First Representative for the Suborder Erpobdelliformes

Yun-Ling Xu 1,2, Ling Xiao 1 and Jing Nie 1 *

1Hubei Institute for Food and Drug Control, Wuhan, Hubei Province, 430064, PR China
2College of Pharmaceutical Sciences, Beijing Chinese Medical University, Beijing, 100000, PR China

Abstract

*Erpobdella octoculata* Linnaeus, 1758 (Hirudinea: Arhynchobdellida: Erpobdellidae) is a very common and morphologically variable macrophagous predators of aquatic invertebrates. Here we determined the complete mitochondrial DNA (mtDNA) sequence of this species, as the first representative of the suborder Erpobdelliformes. This genome is 14,407 bp in length with an A+T content of 71.55%, containing 37 typical animal mitochondrial genes and a non-coding region (NCR). All genes were arranged in the same order as most reported annelid. A remarkable strand bias found for all protein coding genes (PCGs) was positive AT-skew and negative GC-skew. The models of secondary structures for the two RNA genes of this species were predicted. rml consisted of six domains and 46 helices, while three domains and 34 helices for rrnS. The secondary structures of some transfer RNA genes were relatively special. Bayesian phylogenetic and maximum likelihood analyses, based on concatenated nucleotide sequences of 12 PCGs, placed *E. octoculata* within the suborder Hirudiniformes, and the relationships depicted within that order were not quite in agreement with previous morphological classifications. The complete mtDNA sequence of *E. octoculata* provides important genetic markers for identification, ecological and evolutionary studies of leeches.

Keywords: *Erpobdella octoculata*; Mitochondrial DNA; Phylogenetic analyses

Introduction

Leeches (Clitellata: Hirudinida) are hermaphroditic worms and abundant predators or ecto-parasites in various freshwater habitats. Many leeches have a significant adverse impact on freshwater culture and animal health [1-3]. *Erpobdella octoculata* Linnaeus, 1758 (Hirudinea: Arhynchobdellida: Erpobdelliformes) is one of the most common leech species in freshwaters [4]. Its assemblage typically dominates the lowlands in lake, streams and rivers as well as in urban park ponds [5-7]. Despite widely distributed, the current genomic knowledge of leeches is scant and taxon diversity is poorly constructed [8,9].

Animal mitochondrial DNA (mtDNA) is a single circular duplex molecule typically coding for 13 protein-coding genes, 2 rRNA genes and 22 tRNA genes [10]. The mitochondrial genome has been extensively used to study the phylogenetic relationships at several taxonomic levels [11,12]. Compared to other metazoan animals, only five complete mt genome sequences of leech species for the suborder Hirudiniformes have been sequenced and deposited in GenBank to date, and no mt genome for the suborder Erpobdelliformes has been reported [13,14]. In this study, we sequenced the complete mitochondrial (mt) genomes of *E. octoculata*, analyzed its mitochondrial gene arrangements, gene compositions, translation and initiation codons and codon usage as well as the structure of rRNA and tRNA genes. In addition, we used the currently available nucleotide sequences to reconstruct the phylogenetic relationship among the included annelids. The new mt genome sequence may provide novel and useful mtDNA markers for investigating genetic composition of suborder Erpobdelliformes, as it is the first reported complete mtDNA sequence in this suborder.

Materials and Methods

Samples and DNA extraction

Specimens of *E. octoculata* were sampled from Hubei province, China. After morphological identification according to existing keys and descriptions [4], they were fixed in 95% (v/v) ethanol and stored at -20°C. Total genomic DNA was extracted from this specimen using standard phenol/chloroform extraction [15].

PCR amplification and sequencing

The mt genome of *E. octoculata* was generated using amplification of overlapping PCR fragments with each fragment overlapped upstream (5') by ~ 200 bp. First, 1-2kb fragments were amplified using primers designed from conserved region of published sequences of *W. pigra* and *H. nipponia* mt genome. Then, specifically designed primers based on the known sequence were used for the secondary PCRs (primer sequences list see Table S1). PCRs were performed with TransStart FastPfu DNA Polymerase (TransGen Co., Beijing, China) under the following conditions: initial denaturation cycle of 95°C for 2 min, followed by 30 cycles of 95°C for 20 s, annealing of 50-60°C for 15 s, and extension at 72°C for 20-40 s, with a final elongation at 72°C for 5 min after the last cycle. After end repairing, all of the amplified products were inserted into T-vector pMD18-T (TaKaRa Co., Dalian, China) and sequenced in both directions. Individual clone sequences were analyzed with the DNASTAR software package (http://www. DNASTAR.com).

Sequence analysis and annotation

Sequences were assembled manually and aligned against the complete mt genome sequence of *W. pigra* and *H. nipponia* using the

*Corresponding author: Jing Nie, Hubei Institute for Food and Drug Control, Wuhan, Hubei Province 430064, PR China, Tel: +86 27 8727 0452; E-mail: niejingwh@hotmail.com*

Received August 10, 2015; Accepted September 14, 2015; Published September 21, 2015


Copyright: © 2015 Xu YL, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
computer program Clustal X 1.83 [16] to identify gene boundaries. Protein coding regions were identified via ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) with invertebrate mitochondrial genetic codes. Nucleotide composition, amino acid composition, and Relative Synonymous Codon Usage (RSCU) values were analyzed with MEGAS 5.0 [17]. AT- and GC-skew were calculated using the following formula: AT skew=([A-T]/[A+T], GC skew=[G-C]/[G+C]) [18]. The two rRNA genes were predicted by comparison with those of other metazoan animals previously reported [13,14] and the secondary structure drawn by the software of XRNA 1.2.0b. The secondary structures of two rRNA genes were inferred based on models developed for other metazoan animals [19-22]. To infer the rRNA gene secondary structures, we used a commonly accepted comparative approach to correct for unusual pairings with RNA-editing mechanisms that were well known in metazoan mt genomes [23,24]. For analyzing tRNA genes, putative secondary structures of tRNA genes were identified using tRNAscan-SE [25] and the ARWEN program [26]. For tRNAscan-SE, the following parameters were used: Source=Mito/Chloroplast, Search Mode=tRNA scan-only, Genetic Code=Invertebrate Mito, Cove score cutoff=1. The remaining tRNA genes were identified by recognizing potential secondary structures and anticodon sequences by eye.

Phylogenetic analyses

All the currently available 14 Annelida complete mt genomes data from this study were collected for phylogenetic analyses. The mt genome of Phascolosoma esculenta and Sipunculus nudus [27] served as out-groups. The nucleotide sequences for all 20 PCGs (excluding atp8) were aligned using MAFFT7 [28] with the default settings and concatenation. The concatenated set of amino acid sequences from all 12 PCGs was then used for phylogenetic analyses. Bayesian inferences (BI) and maximum likelihood (ML) analysis performed with MrBayes 3.2 [29] and PHYML online web server [30] with optimal model determined by Mrmodeltest 2.3, respectively. According to the Akaile information criterion, the GTR+I+G model with optimal model determined by Mrmodeltest 2.3 was used for BI analysis. The BI analyses were conducted under the following conditions: 1,000,000 generations, four chains (one cold chain and three hot chains), a burn-in step for the first 25% (2,500) generations, and a thinning interval of 10. The remaining trees were used to calculate Bayesian posterior probabilities (BPP). For ML phylogeny construction, the confidences were estimated by 1000 bootstrap replicates.

Results and Discussion

General features of the mt genome of E. octoculata

Gene composition and gene arrangement of E. octoculata are summarized in Figure 1 and Table 1, and the sequences have been deposited in GenBank with the accession No. KC-688270. The complete mt genome of E. octoculata is 14,407 bp in length, which is relatively smaller than those reported annelids mt genomes (e.g., Lumbricus terrestris, Platynereis dumerilii). This difference is mainly due to the total length of non-coding sequence. This circular mt genome (Figure 1) is typical of annelid mt genomes, including 13 PCGs (atp6, atp8, cox1-3, cytb, nad4L and nad6), two ribosomal RNA genes (rrnL and rrnS), 22 transfer RNA genes and a large non-coding region (Table 1). All mt genes of E. octoculata are transcribed in the same direction and the gene arrangement is consistent with those of the W. pigra and H. nipponia. In this species, protein-coding gene nad3 overlaps tRNA-SerAGN, tRNA-SerAGN overlaps nad2 by 15 nucleotides, respectively (Table 1, Figure 1). Nad4L and nad4 share an overlap by 7 nucleotides. Such overlaps are also found in reported Hirudinea mtDNA genomes [13,14] and common to annelids [31-34]. The mt genome of E. octoculata contains 54 bp of intergenic sequences which spread over 10 locations, ranging from 1 to 15 bp in size. The longest intergenic regions are located between the rrnL and tRNA-SerUCN, tRNA-Arg and tRNA-Gln genes. The Hirudinea leech mtDNA of W. pigra and H. nipponia also contained intergenic regions (1-61 bp in length) in 12 locations.

Protein-coding genes

In this mt genome, 7 genes (cox1, atp6, atp8, nad2, nad4, nad6 and cytb) start with ATG, 4 genes (cox2, nad3, nad5 and nad4L) start with GTG, 1 gene (cox3) start with ATA and 1 gene (nad1) start with ATT (Table 1). Only 4 genes (atp6, nad3, nad5 and nad4L) contain the full termination codon TAA as termination codon, while other genes end with abbreviated stop codon T (Table 1). And the latter are presumably converted to TAA by post-transcriptional polyadenylation [35].

A total of 3,660 amino acids are encoded by the E. octoculata mt genome. Similar to other leech species, the nucleotide composition of E. octoculata mt genome is also present a clear bias, with the base contents of A, T, G, C in the entire mt genome are 4,456(30.93%), 5,852(40.62%), 1,910 (15.88%) and 1,807 (12.54%), respectively. Among the 13 PCGs, the lowest A+T content is 65.36% in cox1, while the highest is 78.29% in nad6 (Table S2). The AT-skew and GC-skew across 13PCGs and rRNA are shown in Figure 2. Within 13PCGs, the AT-skews ranging from -0.26 in nad3 to -0.07 in cox2, and the average AT-skew in mt genomes is -0.14. This kind of base bias of nucleotides composition has been generally related to asymmetric mutational constraints in the process of replication [36]. The nucleotide bias toward AT was also reflected in the codon usage (Table 2). The relative synonymous codon usage (RSCU) values also showed a biased use of A and T nucleotides in the codon usage (Table 2). The high incidence of anticodons NNA and NNU indicated partiality for AT in the codon of PCGs. Leu2 (11.31%), Phe (8.57%), Ile (8.28%) and Met (7.74%) were the most frequent amino acids in E. octoculata mtDNA PCGs.

Ribosomal RNA genes

The rrnL (165 ribosomal RNA) and rrnS (125 ribosomal RNA) genes of E. octoculata were identified by sequence comparison with other annelids. These two genes are separated by tRNA-Val. The sizes of the rrnL and rrnS genes are 1145 and 732 bp respectively, which


Gene | Position | Size (bp) | Codon | Anticodon | Intergenic nucleotide^b^ | Notes
--- | --- | --- | --- | --- | --- | ---
cox1 | 1 | 1534 | 1534 | 511 | ATG | T- | 0
rrnL | 1535 | 1596 | 62 | GTT | 0
rrnS | 1597 | 2275 | 679 | 226 | GTG | T- | 0
atp8 | 2276 | 2347 | 72 | GTC | -8
atp6 | 2340 | 2490 | 151 | 50 | ATG | T- | 0
nad6 | 2491 | 2549 | 59 | TCC | 0
cytb | 2550 | 2819 | 70 | GTA | 2
rrnL | 2622 | 3420 | 799 | 266 | ATA | T- | 0
rrnS | 3421 | 3489 | 69 | TTG | 0
rrnL | 3490 | 3946 | 457 | 152 | ATG | T- | 0
rrnS | 3947 | 5081 | 1135 | 378 | ATG | T- | 0
nad5 | 5082 | 5144 | 63 | TCA | 0
nad4 | 5145 | 5846 | 702 | 233 | ATG | TAA | 0
rrnL | 5847 | 5909 | 63 | TCG | 0
rrnS | 5910 | 6005 | 96 | GTG | 0
rrnL | 6006 | 6070 | 65 | TGT | 2
nad4 | 6071 | 7806 | 1736 | 578 | GTG | TAA | 0
nad3 | 7807 | 7868 | 62 | GAA | 0
nad5 | 7869 | 7929 | 61 | TTC | -3
nad6 | 7927 | 7987 | 61 | TGG | 0
nad4 | 7990 | 8048 | 59 | TGT | 2
nad5 | 8049 | 8336 | 288 | 95 | GTG | TAA | -7
nad6 | 8330 | 9659 | 1330 | 443 | ATG | T- | 12
nad4 | 9672 | 9732 | 61 | GCA | 0
nad5 | 9733 | 9795 | 63 | CAT | 0
nad6 | 9796 | 10527 | 732 | TCA | -1
rrnL | 10528 | 10583 | 56 | TCT | 0
rrnS | 10582 | 111726 | 1145 | ATG | -1
rrnL | 111727 | 111787 | 61 | TAG | -1
rrnS | 111787 | 111854 | 68 | TGA | 0
rrnL | 111855 | 111914 | 60 | TCT | -1
rrnS | 111914 | 111974 | 61 | TAA | 12
rrnL | 111987 | 12860 | 874 | 291 | ATT | T- | 33
rrnS | 12860 | 12955 | 62 | GAT | 1
rrnL | 12957 | 13020 | 64 | TTT | 1
rrnS | 13022 | 13368 | 345 | 114 | GTG | TAA | -15
rrnS | 13352 | 13434 | 83 | TCT | -15
rrnL | 13420 | 14401 | 982 | 327 | ATG | T- | 6

Note: a. T-represent incomplete stop codons.
b. b. indicates length of intergenic gap (positive value) or overlap (negative value) between two adjacent genes.
c. c. the structure of rrnL may be three-armed with a DHU-replacement loop (positions13365-13421, 53 bp) or cloverleaf form (13365-13434, 83 bp).

d. d. the secondary structure of rrnS may be three-armed with a DHU-replacement loop (positions13365-13421, 53 bp) or cloverleaf form (13365-13434, 83 bp).

e. e. the structure of rrnS may be three-armed with a DHU-replacement loop (positions13365-13421, 53 bp) or cloverleaf form (13365-13434, 83 bp).

Table 1: Gene organization and features of E. octoculata mt genome (starting from cox1).

Figure 2: Scatter plot of AT- and GC-skews in E. octoculata. Values were calculated for the protein and rRNA genes. AT-skew= (A-T) (A+T); GC-skew= (G-C) (G+C).

Figure 2: Scatter plot of AT- and GC-skews in E. octoculata. Values were calculated for the protein and rRNA genes. AT-skew= (A-T) (A+T); GC-skew= (G-C) (G+C).

are the lowest among the leeches studied to date (Tables 1and S2), and their A+T contents are 74.79% and 70.86%, respectively. The secondary structure of rrnL and rrnS, which was inferred according to the models proposed for these genes in other species. The secondary structure of E. octoculata rrnL consisted of six structural domains ((labeled I, II, III, IV, V and VI) (Figure-3) and included 46 helices. Overall, the highest level of invariable positions was located on domain V-IV, while lowest level was on domains I–II compared with other leech species. Some helices (e.g., H25, H27, H30 and H37) were highly conserved in both sequence and secondary structure among the leech mtDNA, and only few nucleotides differences were found either in the terminal loops or in the terminal couplets of helices. The secondary structure of rrnS contained three domains (labeled I, II, III) and 34 helices (Figure 4). Conservative sites were mainly in domain III and some helices (e.g., H10, H11 and H16) in domain II.
Transfer RNA (tRNA) genes

A total of 22 tRNA sequences (ranging from 59 to 83 nucleotides in length) were identified in the *E. octoculata* mt genome. Except for tRNA-Gly (TCC) and tRNA-Pro (TGG) which lack DHU arms, most tRNA genes appeared to exhibit the standard cloverleaf structure. The predicted secondary structures of glycine and proline tRNA contained the DHU arm but lack the TψC arm, which was not common in previously reported annelids. For the serine tRNA, a four-armed structure with variable loop formed a 6-bp "Extra"arm" was feasible scanned by tRNAscan-SE, while a three-armed structure with a DHU-replacement loop was also possible (Figure 5) when using the ARWEN program.

<table>
<thead>
<tr>
<th>AA</th>
<th>codon</th>
<th>N</th>
<th>%</th>
<th>RSCU</th>
<th>Codon</th>
<th>N</th>
<th>%</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>GCU</td>
<td>62</td>
<td>1.69</td>
<td>1.81</td>
<td>CCU</td>
<td>49</td>
<td>1.34</td>
<td>1.35</td>
</tr>
<tr>
<td>GCC</td>
<td>23</td>
<td>0.63</td>
<td>0.67</td>
<td>CCC</td>
<td>12</td>
<td>0.33</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>GCA</td>
<td>44</td>
<td>1.2</td>
<td>1.28</td>
<td>CCA</td>
<td>68</td>
<td>1.66</td>
<td>1.88</td>
<td></td>
</tr>
<tr>
<td>GCG</td>
<td>8</td>
<td>0.22</td>
<td>0.23</td>
<td>CCG</td>
<td>16</td>
<td>0.44</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Cys (C)</td>
<td>UGU</td>
<td>45</td>
<td>1.23</td>
<td>1.55</td>
<td>CAA</td>
<td>43</td>
<td>1.17</td>
<td>1.48</td>
</tr>
<tr>
<td>UGC</td>
<td>13</td>
<td>0.36</td>
<td>0.45</td>
<td>CAG</td>
<td>15</td>
<td>0.41</td>
<td>0.52</td>
<td></td>
</tr>
<tr>
<td>Asp (D)</td>
<td>GAU</td>
<td>70</td>
<td>1.91</td>
<td>1.59</td>
<td>Arg (R)</td>
<td>CGU</td>
<td>18</td>
<td>0.49</td>
</tr>
<tr>
<td>GAC</td>
<td>18</td>
<td>0.49</td>
<td>0.41</td>
<td>CGC</td>
<td>5</td>
<td>0.14</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Glu (E)</td>
<td>GAA</td>
<td>50</td>
<td>1.37</td>
<td>1.32</td>
<td>CGA</td>
<td>28</td>
<td>0.77</td>
<td>2</td>
</tr>
<tr>
<td>GAG</td>
<td>26</td>
<td>0.71</td>
<td>0.68</td>
<td>CGG</td>
<td>5</td>
<td>0.14</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Phe (F)</td>
<td>UUU</td>
<td>256</td>
<td>6.99</td>
<td>1.63</td>
<td>Ser1 (S1)</td>
<td>AGU</td>
<td>48</td>
<td>1.31</td>
</tr>
<tr>
<td>UUC</td>
<td>58</td>
<td>1.58</td>
<td>0.37</td>
<td>AGC</td>
<td>5</td>
<td>0.14</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Gly (G)</td>
<td>GGU</td>
<td>63</td>
<td>1.72</td>
<td>1.36</td>
<td>AGA</td>
<td>60</td>
<td>1.64</td>
<td>1.26</td>
</tr>
<tr>
<td>GCC</td>
<td>19</td>
<td>0.52</td>
<td>0.41</td>
<td>AGG</td>
<td>32</td>
<td>0.87</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>GCA</td>
<td>44</td>
<td>1.2</td>
<td>0.95</td>
<td>Ser2 (S2)</td>
<td>UCU</td>
<td>87</td>
<td>2.38</td>
<td>1.83</td>
</tr>
<tr>
<td>GGG</td>
<td>59</td>
<td>1.61</td>
<td>1.28</td>
<td>UCC</td>
<td>24</td>
<td>0.66</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>His (H)</td>
<td>CAU</td>
<td>59</td>
<td>1.61</td>
<td>1.79</td>
<td>UCA</td>
<td>105</td>
<td>2.87</td>
<td>2.2</td>
</tr>
<tr>
<td>CAC</td>
<td>7</td>
<td>0.19</td>
<td>0.21</td>
<td>UCG</td>
<td>20</td>
<td>0.55</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Ile (I)</td>
<td>AUU</td>
<td>264</td>
<td>7.21</td>
<td>1.74</td>
<td>Thr (T)</td>
<td>ACU</td>
<td>70</td>
<td>1.91</td>
</tr>
<tr>
<td>AUC</td>
<td>39</td>
<td>1.07</td>
<td>0.26</td>
<td>ACC</td>
<td>14</td>
<td>0.38</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Lys (K)</td>
<td>AAA</td>
<td>88</td>
<td>2.4</td>
<td>1.54</td>
<td>ACA</td>
<td>49</td>
<td>1.34</td>
<td>1.37</td>
</tr>
<tr>
<td>AAG</td>
<td>26</td>
<td>0.71</td>
<td>0.46</td>
<td>ACG</td>
<td>10</td>
<td>0.27</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>Leu1 (L1)</td>
<td>CUU</td>
<td>58</td>
<td>1.58</td>
<td>0.62</td>
<td>Val (V)</td>
<td>GUU</td>
<td>113</td>
<td>3.09</td>
</tr>
<tr>
<td>CUC</td>
<td>13</td>
<td>0.36</td>
<td>0.14</td>
<td>GUC</td>
<td>18</td>
<td>0.49</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>CUA</td>
<td>61</td>
<td>1.67</td>
<td>0.65</td>
<td>GUA</td>
<td>122</td>
<td>3.33</td>
<td>1.72</td>
<td></td>
</tr>
<tr>
<td>CUG</td>
<td>13</td>
<td>0.36</td>
<td>0.14</td>
<td>GUG</td>
<td>31</td>
<td>0.85</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Leu2 (L2)</td>
<td>UUA</td>
<td>322</td>
<td>8.8</td>
<td>3.46</td>
<td>Trp (W)</td>
<td>UGA</td>
<td>79</td>
<td>2.16</td>
</tr>
<tr>
<td>UUG</td>
<td>92</td>
<td>2.51</td>
<td>0.99</td>
<td>UGG</td>
<td>25</td>
<td>0.68</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Met (M)</td>
<td>AUA</td>
<td>218</td>
<td>5.96</td>
<td>1.54</td>
<td>Try (Y)</td>
<td>UAU</td>
<td>136</td>
<td>3.72</td>
</tr>
<tr>
<td>AUG</td>
<td>65</td>
<td>1.78</td>
<td>0.46</td>
<td>UAC</td>
<td>35</td>
<td>0.96</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>Asn (N)</td>
<td>AAU</td>
<td>113</td>
<td>3.09</td>
<td>1.67</td>
<td>Stop</td>
<td>UAA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAC</td>
<td>22</td>
<td>0.6</td>
<td>0.33</td>
<td>UAG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Codon usage of 13 PCGs of *E. octoculata* mt genome.

**Transfer RNA (tRNA) genes**

A total of 22 tRNA sequences (ranging from 59 to 83 nucleotides in length) were identified in the *E. octoculata* mt genome. Except for tRNA-Gly (TCC) and tRNA-Pro (TGG) which lack DHU arms, most tRNA genes appeared to exhibit the standard cloverleaf structure. The predicted secondary structures of glycine and proline tRNA contained the DHU arm but lack the TψC arm, which was not common in previously reported annelids. For the serine tRNA, a four-armed structure with variable loop formed a 6-bp—"Extra"arm" was feasible scanned by tRNAscan-SE, while a three-armed structure with a DHU-replacement loop was also possible (Figure 5) when using the ARWEN program.

Figure 3: Predicted secondary structure of the rrnL gene in E. octoculata. Roman numerals denote the conserved domain structure. Inferred Watson-Crick bonds are illustrated by lines, GT bonds by dots.

Phylogenetic analyses

In this study, the nucleotide sequences of the 12 PCGs were concatenated to construct phylogenetic relationships, which might result in a more complete analysis than analyzing each sequence separately. We incorporated species from Echiura, Oligochaeta, Polychaeta and Hirudinea in the phylogenetic analysis. Sipuncula species Phascolosoma esculenta and Sipunculus nudus were included as out-groups. Analysis using BI and ML resulted in two trees with the identical or similar topology except for some variation in node confidence values (Figure 6). Numbers at each node indicate bootstrap support, percentages of Bayesian posterior probabilities (left value) and ML bootstrap support values (right value). In both cases, the mtDNA data supported the hypothesis that Sipunculans and Annelids (including echiurans) formed a monophyletic group. While in Annelids, phylogenetic tree indicated that echiurans were derived from annelids, the species from Oligochaeta and Hirudinea formed a monophyletic group Clitellata, while the class Polychaeta showed polyphyletic based on the mt genome (Figure 6). The above results were consistent with previous research based on morphological classifications and some molecular studies [37-44]. Within the Hirudinea clade, the statistical data revealed that E. octoculata (Arhynchobdellida: Erpobdelliformes) was more closely related to W. pigra (Arhynchobdellida: Hirudiniformes) than to H. nipponia (Arhynchobdellida: Hirudiniformes), indicating that family Erpobdellidae was more closely related to family Haemopidae. The above result based on the nucleotide sequences of the 12PCGs was not consistent with previous classification based on their morphological features and using LSU and SSU rDNA and cox1 sequences [45,46]. Apparently, more taxon sampling and additional markers for further phylogenetic inference are required to resolve the controversy on the phylogeny of the Arhynchobdellida.

Conclusion

In the present study, the complete mtDNA sequences of E. octoculata was the first sequenced high quality mt genome of non-blood feeding leech of the suborder Erpobdelliformes. The availability of the E. octoculata mtDNA sequences can provide novel and useful genetic markers for investigating genetic composition, population genetics studies and phylogeographics of leech species. Phylogenetic
Figure 4: Predicted secondary structure of the rrnS gene in *E. octoculata*. Roman numerals denote the conserved domain structure. Inferred Watson-Crick bonds are illustrated by lines, GT bonds by dots.

Figure 5: Two possible structures of tRNA-Ser2 (cloverleaf structure or D-loop).

Figure 6: Phylogenetic relationships among the sequenced annelids. Phylogenetic analyses were carried out for the 15 annelid based on all 12 protein-coding genes from their respective mt genomes. Phascolosoma esculenta and Sipunculus nudus was used as out groups. Numbers at the nodes are Bayesian posterior probabilities (left) and ML bootstrap values (right).

analyses using concatenated nucleotide sequences of 12 mt PCGs indicated that *E. octoculata* (Erpobdellidae) was more closely related to the Haemopidae than to Hirudinida families.

Acknowledgements

Special grateful to Prof. Pin Nie (Institute of Hydrobiology, Chinese Academy of Sciences) for his technical assistance. Thanks go to Dr. Ran Jiao (Institute of Hydrobiology, Chinese Academy of Sciences) for her comments on this experiment. This study was supported by the National Science and Technology Major Project "Creation of Major New Drugs"of China (No. 2014ZX09304307-002).

References


