

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

Yun-Ling Xu^{1,2}, Ling Xiao¹ 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 rRNA genes of this species were predicted. *rnrL* consisted of six domains and 46 helices, while three domains and 34 helices for *rnrS*. 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 hermaphrodite 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 2min, 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 0482; E-mail: niejingwh@hotmail.com

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

Citation: Xu YL, Xiao L, Nie J (2015) The Mitochondrial Genome of *Erpobdella octoculata* (Hirudinida:Arhynchobdellida), as the First Representative for the Suborder Erpobdelliformes. J Phylogen Evolution Biol 3: 157. doi:10.4172/2329-9002.1000157

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 MEGA5.0 [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 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 and the sequences from this study were collected for phylogenomic analyses. The mt genome of *Phascolosoma esculenta* and *Sipunculus nudus* [27] served as out-groups. The nucleotide sequences for all 12 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 Akaike information criterion, the GTR+I+G model was selected for 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) trees and 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*

Genome 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*, *nad1-6* and *nad4L*), 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 *rrnS* 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 *incox2*, 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 *E. octoculata* (Table-2). The high incidence of anticodons NNA and NNU indicated partiality for AT in the anticodon 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* (16S ribosomal RNA) and *rrnS* (12S 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

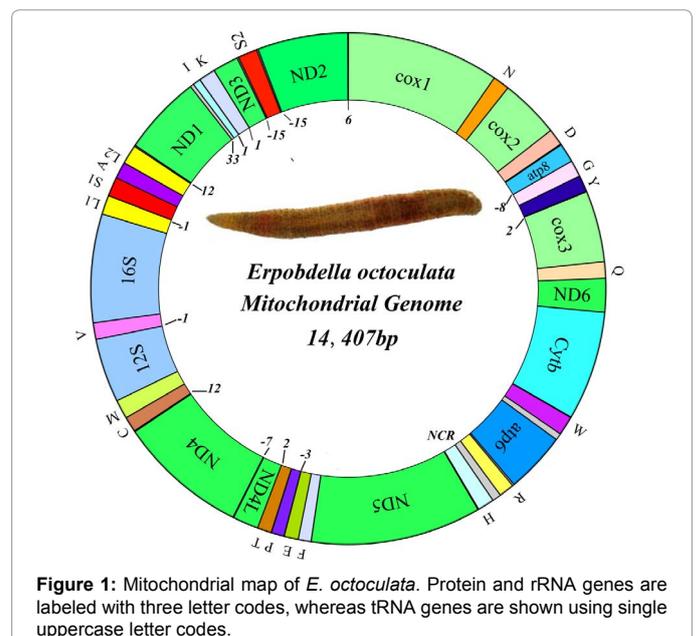


Figure 1: Mitochondrial map of *E. octoculata*. Protein and rRNA genes are labeled with three letter codes, whereas tRNA genes are shown using single uppercase letter codes.

Gene	Position		Size (bp)		Codon		Anticodon	Intergenic nucleotide ^b
	From	To	Nucleotide	Amino acid	Initiation	stop ^a		
<i>cox1</i>	1	1534	1534	511	ATG	T-		0
<i>tRNA^{Asn}</i>	1535	1596	62				GTT	0
<i>cox 2</i>	1597	2275	679	226	GTG	T-		0
<i>tRNA^{Asp}</i>	2276	2347	72				GTC	-8
<i>atp8</i>	2340	2490	151	50	ATG	T-		0
<i>tRNA^{Gly}</i>	2491	2549	59				TCC	0
<i>tRNA^{Tyr}</i>	2550	2619	70				GTA	2
<i>cox 3</i>	2622	3420	799	266	ATA	T-		0
<i>tRNA^{Glu}</i>	3421	3489	69				TTG	0
<i>nad6</i>	3490	3946	457	152	ATG	T-		0
<i>cyt b</i>	3947	5081	1135	378	ATG	T-		0
<i>tRNA^{Trp}</i>	5082	5144	63				TCA	0
<i>atp6</i>	5145	5846	702	233	ATG	TAA		0
<i>tRNA^{Arg}</i>	5847	5909	63				TCG	0
<i>NCR</i>	5910	6005	96					0
<i>tRNA^{His}</i>	6006	6070	65				GTG	0
<i>nad5</i>	6071	7806	1736	578	GTG	TAA		0
<i>tRNA^{Phe}</i>	7807	7868	62				GAA	0
<i>tRNA^{Glu}</i>	7869	7929	61				TTC	-3
<i>tRNA^{Pro}</i>	7927	7987	61				TGG	0
<i>tRNA^{Thr}</i>	7990	8048	59				TGT	2
<i>nad4l</i>	8049	8336	288	95	GTG	TAA		-7
<i>nad4</i>	8330	9659	1330	443	ATG	T-		12
<i>tRNA^{Cys}</i>	9672	9732	61				GCA	0
<i>tRNA^{Met}</i>	9733	9795	63				CAT	0
<i>srRNA</i>	9796	10527	732					0
<i>tRNA^{Val}</i>	10528	10583	56				TCA	-1
<i>lrRNA</i>	10582	11726	1145					0
<i>tRNA^{Leu(CUN)}</i>	11727	11787	61				TAG	-1
<i>tRNA^{Ser(UCN)}</i>	11787	11854	68				TGA	0
<i>tRNA^{Ala}</i>	11855	11914	60					-1
<i>tRNA^{leu(UUR)}</i>	11914	11974	61				TAA	12
<i>nad1</i>	11987	12860	874	291	ATT	T-		33
<i>tRNA^{Ile}</i>	12894	12955	62				GAT	1
<i>tRNA^{Lys}</i>	12957	13020	64				TTT	1
<i>nad3</i>	13022	13366	345	114	GTG	TAA		-15
<i>tRNA^{Ser(AGN) c}</i>	13352	13434	83				TCT	-15
<i>nad2</i>	13420	14401	982	327	ATG	T-		6

Note: a. T-represent incomplete stop codons.

b. indicates length of intergenic gap (positive value) or overlap (negative value) between two adjacent genes.

c. the structure of *tRNA^{Ser(AGN)}* 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*).

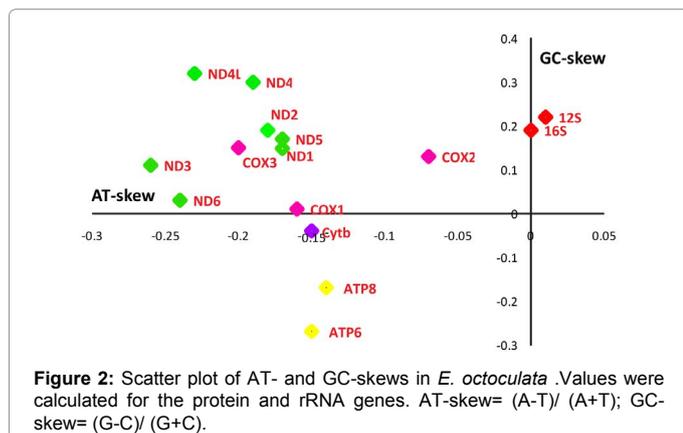


Figure 2: Scatter plot of AT- and GC-skews in *E. octoculata*. Values were calculated for the protein and rRNA genes. AT-skew= (A-T)/(A+T); GC-skew= (G-C)/(G+C).

are the lowest among the leeches studied to date (Tables 1 and 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.

AA	codon *	N	%	RSCU		Codon	N	%	RSCU
Ala (A)	GCU	62	1.69	1.81		CCU	49	1.34	1.35
	GCC	23	0.63	0.67		CCC	12	0.33	0.33
	GCA	44	1.2	1.28		CCA	68	1.86	1.88
	GCG	8	0.22	0.23		CCG	16	0.44	0.44
Cys (C)	UGU	45	1.23	1.55		CAA	43	1.17	1.48
	UGC	13	0.36	0.45		CAG	15	0.41	0.52
Asp (D)	GAU	70	1.91	1.59	Arg (R)	CGU	18	0.49	1.29
	GAC	18	0.49	0.41		CGC	5	0.14	0.36
Glu (E)	GAA	50	1.37	1.32		CGA	28	0.77	2
	GAG	26	0.71	0.68		CGG	5	0.14	0.36
Phe (F)	UUU	256	6.99	1.63	Ser1 (S1)	AGU	48	1.31	1.01
	UUC	58	1.58	0.37		AGC	5	0.14	0.1
Gly (G)	GGU	63	1.72	1.36		AGA	60	1.64	1.26
	GGC	19	0.52	0.41		AGG	32	0.87	0.67
	GGA	44	1.2	0.95	Ser2 (S2)	UCU	87	2.38	1.83
	GGG	59	1.61	1.28		UCC	24	0.66	0.5
His (H)	CAU	59	1.61	1.79		UCA	105	2.87	2.2
	CAC	7	0.19	0.21		UCG	20	0.55	0.42
Ile (I)	AUU	264	7.21	1.74	Thr (T)	ACU	70	1.91	1.96
	AUC	39	1.07	0.26		ACC	14	0.38	0.39
Lys (K)	AAA	88	2.4	1.54		ACA	49	1.34	1.37
	AAG	26	0.71	0.46		ACG	10	0.27	0.28
Leu1 (L1)	CUU	58	1.58	0.62	Val (V)	GUU	113	3.09	1.59
	CUC	13	0.36	0.14		GUC	18	0.49	0.25
	CUA	61	1.67	0.65		GUA	122	3.33	1.72
	CUG	13	0.36	0.14		GUG	31	0.85	0.44
Leu2 (L2)	UUA	322	8.8	3.46	Trp (W)	UGA	79	2.16	1.52
	UUG	92	2.51	0.99		UGG	25	0.68	0.48
Met (M)	AUA	218	5.96	1.54	Try (Y)	UAU	136	3.72	1.59
	AUG	65	1.78	0.46		UAC	35	0.96	0.41
Asn (N)	AAU	113	3.09	1.67	Stop	UAA	0	0	0
	AAC	22	0.6	0.33		UAG	0	0	0

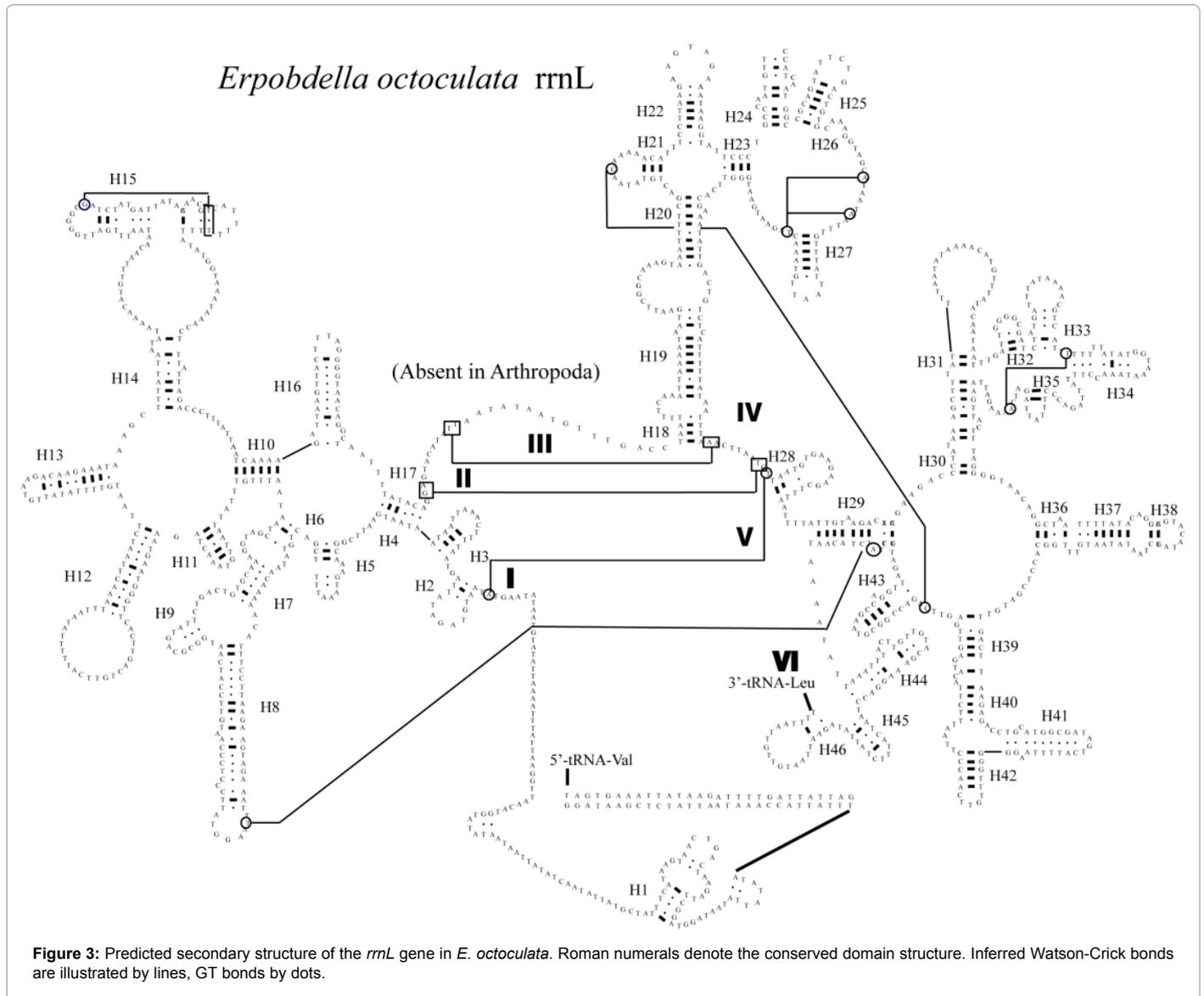
“**”: A total of 3660 codons from *Erpobdella octoculata* were analyzed, incomplete termination codons was excluded. AA, amino acid; RSCU, Relative synonymous codon usage; N=number of copies; %=n/3660.

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 T ψ C 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.



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

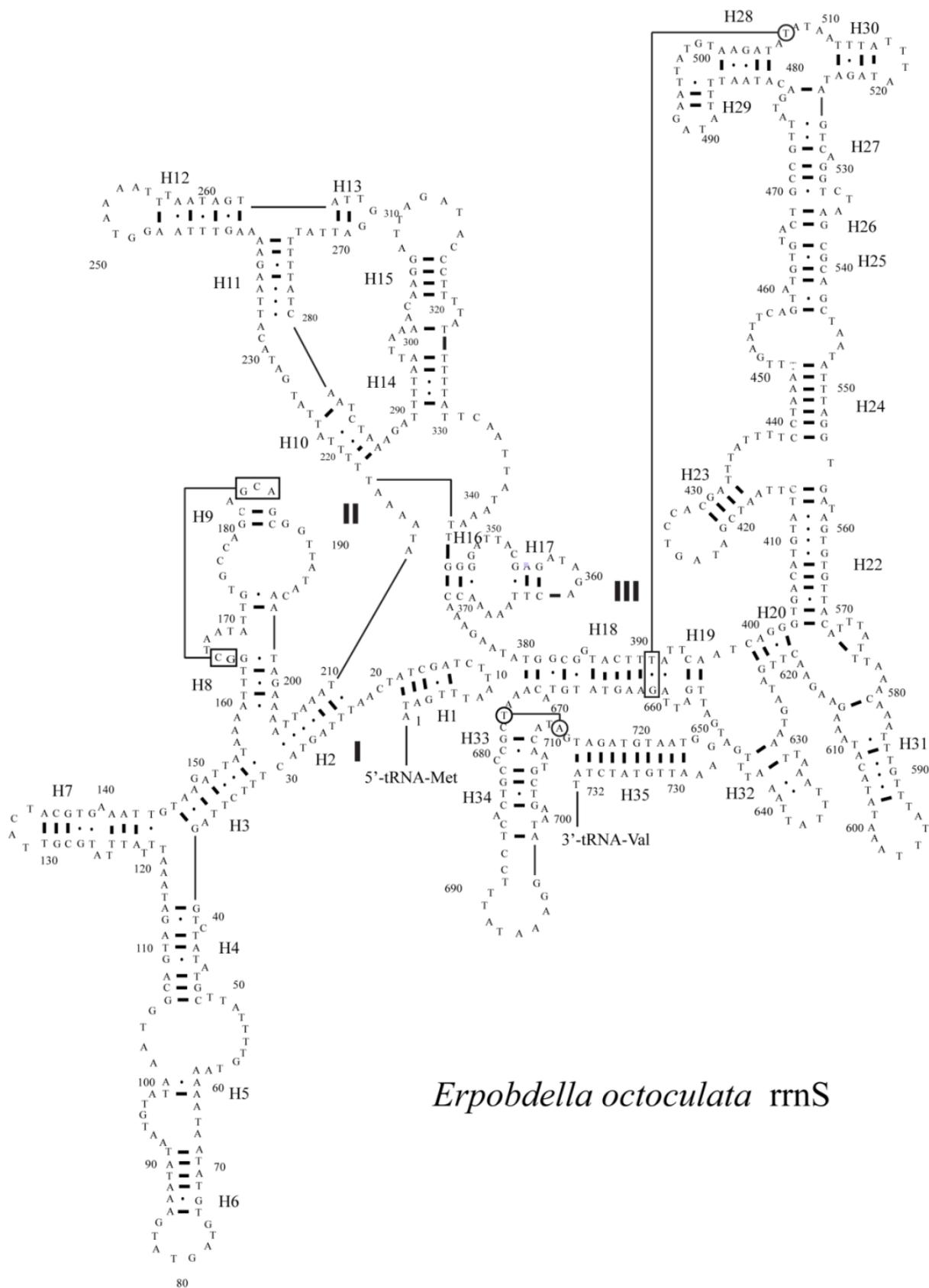


Figure 4: Predicted secondary structure of the *rrnS* gene in *E. octoculata*. Roman numerals denote the conserved domain structure. Inferred Watson-Crick bonds are illustrated by lines, GT bonds by dots.

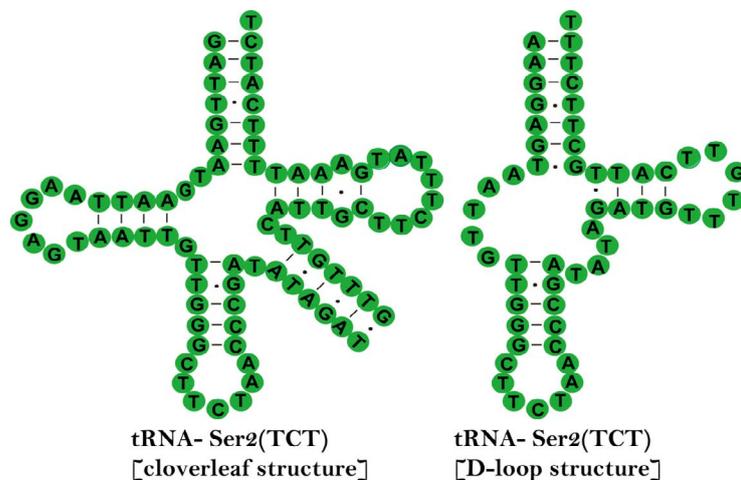


Figure 5: Two possible structures of tRNA-Ser2 (cloverleaf structure or D-loop).

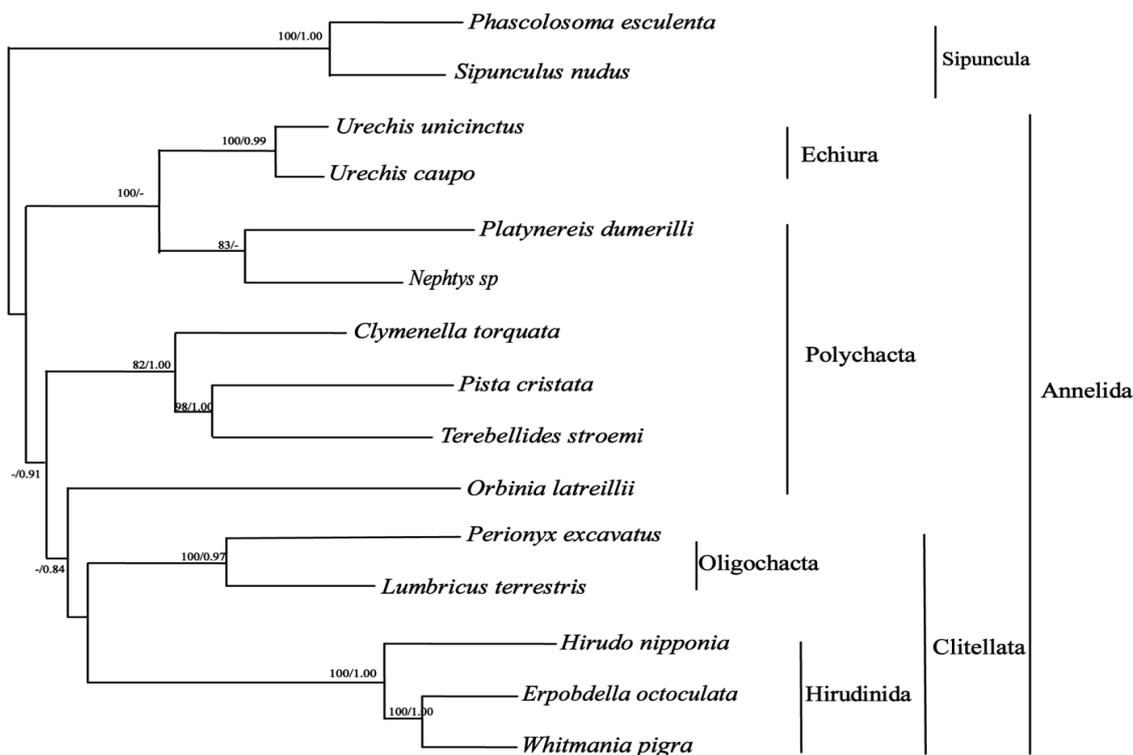


Figure 6: Phylogenetic relationships among the sequenced annelids. Phylogenetic analyses were carried out for the 15 annelid based on all 12 protein-coding genes from their respective mt genomes. *Phascolosoma esculenta* and *Sipunculus nudus* was used as out groups. Numbers at the nodes are Bayesian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

- Schulz CA (2009) Leeches in the Lake Erie Watershed: The Identification and Potential Role in Disease Transmission. ProQuest.
- Raj SM, Radzi M, Tee MH (2000) Severe rectal bleeding due to leech bite. Am J Gastroenterol 95: 1607.
- Ikizceli I, Avsarogullari L, Sözüer E, Yürümez Y, Akdur O (2005) Bleeding due to a medicinal leech bite. Emerg Med J 22: 458-460.
- Yang T (1996) Fauna Sinica, Annelida: Hirudinea. Science Press, Beijing, China 117-129.

5. Bhat SU, Pandit AK (2010) Ecological study of benthic communities in three limnocene freshwater springs of Kashmir Himalaya. Journal of Natural Sciences and Mathematics 3: 89-96.
6. Koperski P (2006) Relative importance of factors determining diversity and composition of freshwater leech assemblages (Hirudinea; Clitellata): A meta-analysis. Archiv für Hydrobiologie 166: 325-341.
7. Mann KH (1953) The life history of *Erpobdella octoculata* (Linnaeus, 1758). The Journal of Animal Ecology 22: 199-207.
8. Elliott JM, Kutschera U (2011) Medicinal leeches: historical use, ecology, genetics and conservation. Freshwater Reviews 4: 21-41.
9. Phillips AJ, Siddall ME (2009) Poly-paraphyly of Hirudinidae: Many lineages of medicinal leeches. BMC Evol Biol 9: 246.
10. Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27: 1767-1780.
11. McKay BD, Zink RM (2010) The causes of mitochondrial DNA gene tree paraphyly in birds. Mol Phylogenet Evol 54: 647-650.
12. Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: A case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Molecular Biology and Evolution, 18: 1993-2009.
13. Shen X, Wu Z, Sun M, Ren J, Liu B (2011) The complete mitochondrial genome sequence of *Whitmania pigra* (Annelida, Hirudinea): the first representative from the class Hirudinea. Comp Biochem Physiol Part D Genomics Proteomics 6: 133-138.
14. Xu Y, Nie J, Hou J, Xiao L, Lv P (2014) Complete mitochondrial genome of *Hirudo nipponia* (Annelida, Hirudinea). Mitochondrial DNA.
15. Sambrook J, Russell DW (2006) Purification of nucleic acids by extraction with phenol:chloroform. CSH Protoc 2006.
16. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic acids research 25: 4876-4882.
17. 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. Molecular biology and evolution 28: 2731-2739.
18. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41: 353-358.
19. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, et al. (2002) The comparative RNA web (CRW) site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3: 2.
20. Gillespie JJ, Johnston JS, Cannone JJ, Gutell RR (2006) Characteristics of the nuclear (18S, 5.8 S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retro transposable elements. Insect molecular biology 15: 657-686.
21. Winterton SL, Wiegmann BM, Schlinger EI (2007) Phylogeny and Bayesian divergence time estimations of small-headed flies (Diptera: Acroceridae) using multiple molecular markers. Molecular Phylogenetics and Evolution 43: 808-832.
22. Negrisol E, Babbucci M, Patarnello T (2011) The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondrial genomics of neuropterid insects. BMC Genomics 12: 221.
23. Gutell RR, Larsen N, Woese CR (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol Rev 58: 10-26.
24. Buckley TR, Simon C, Flook PK, Misof B (2000) Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. Insect Molecular Biology 9: 565-580.
25. 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.
26. Laslett D, Canbäck B (2008) ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24: 172-175.
27. Mwinyi A, Meyer A, Bleidorn C, Lieb B, Bartolomaeus T, et al. (2009) Mitochondrial genome sequence and gene order of *Sipunculus nudus* give additional support for an inclusion of Sipuncula into Annelida. BMC Genomics 10: 27.
28. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30: 772-780.
29. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61: 539-542.
30. Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33: W557-559.
31. Shen X, Ma X, Ren JF, Zhao F (2009) A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of *Phascolosoma esculenta*. BMC genomics 10: 136.
32. Vallès Y, Halanych KM, Boore JL (2008) Group II introns break new boundaries: Presence in a bilaterian's genome. PLoS One 3: e1488.
33. Zhong M, Struck TH, Halanych KM (2008) Phylogenetic information from three mitochondrial genomes of Terebelliformia (Annelida) worms and duplication of the methionine tRNA. Gene 416: 11-21.
34. Jennings RM, Halanych KM (2005) Mitochondrial genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): Evidence for conserved gene order in annelida. Mol Biol Evol 22: 210-222.
35. Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470-474.
36. Hassanin A, Léger N, Deutsch J (2005) Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. Syst Biol 54: 277-298.
37. Rousset V, Pleijel F, Rouse GW, Erseus C, Siddall ME (2007) A molecular phylogeny of annelids. Cladistics 23: 41-63.
38. Struck TH, Hessling R, Purschke G (2002) The phylogenetic position of the Aeolosomatidae and Parergodrilidae, two enigmatic oligochaetae like taxa of the systematic Polychaeta, based on molecular data from 18S rDNA sequences. Journal of Zoological Systematics and Evolutionary Research 40: 155-163.
39. Struck TH, Purschke G (2005) The sister group relationship of Aeolosomatidae and Potamodrilidae a molecular phylogenetic approach based on 18S rDNA and cytochrome oxidase I. Zoologischer Anzeiger-A Journal of Comparative Zoology 243: 281-293.
40. Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, et al. (2007) Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol Biol 7: 57.
41. Purschke G, Hessling R, Westheide W (2000) The phylogenetic position of the Clitellata and the Echiura on the problematic assessment of absent characters. Journal of Zoological Systematics and Evolutionary Research 38: 165-173.
42. Purschke G (2003) Is *Hrabeiella periglandulata* (Annelida, "Polychaeta") the sister group of Clitellata? Evidence from an ultrastructural analysis of the dorsal pharynxin *H. periglandulata* and *Enchytraeus minutus* (Annelida, Clitellata). Zoomorphology 122: 55-66.
43. McHugh D (1997) Molecular evidence that echiurans and pogonophorans are derived annelids. Proc Natl Acad Sci U S A 94: 8006-8009.
44. McHugh D (2000) Molecular phylogeny of the Annelida. Canadian Journal of Zoology 78: 1873-1884.
45. Ocegüera-Figueroa A, LeÁn-RÁgagnon V, Siddall ME (2005) Filogenia y revisiÁn de losErpobdelliformes (Annelida, Arhynchobdellida) de MÃxico, con base en secuencias de ADNnuclear y mitochondrial. Revista mexicana de biodiversidad, 76: 191-198.
46. Borda E, Siddall ME (2004) Arhynchobdellida (Annelida: Oligochaeta : Hirudinida): phylogenetic relationships and evolution. Molecular phylogenetics and evolution 30: 213-225.