Complete mitochondrial genome of *Whitmania laevis* (Annelida, Hirudinea) and comparative analyses within *Whitmania* mitochondrial genomes

Fei Ye, Ting Liu, Wenbo Zhu & Ping You*

Co-Innovation Center for Qinba regions’ sustainable development, College of Life Science, Shaanxi Normal University, Xi’an, 710062, China.

* Corresponding author: E-mail: youping@snnu.edu.cn

**ABSTRACT.** The complete mitochondrial genome of *Whitmania laevis* is 14,442 bp in length and contains 37 genes including 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, and two ribosomal RNA (rRNA) genes. The almost-complete mitochondrial genome of *Whitmania acranulata*, consisting of 13,494 bp, contains 35 genes including 13 PCGs, 20 tRNA genes, and two rRNA genes. COI phylogenetic analyses showed that the samples reported in GenBank and analysed as *Hirudo nipponia* KC667144, *Hirudinaria manillensis* KC688268 and *Erpobdella octoculata* KC688270 are not the named species and they should belong to *Whitmania*. We compared and analyzed the characteristics of nucleotide composition, codon usage, and secondary structures of 22 tRNAs and two rRNAs from *Whitmania* taxa. Moreover, we analyzed phylogenetic relationships of Annelida using maximum likelihood (ML) and Bayesian inference (BI) methods, based on 11 mitochondrial genes. Our results reveal that *W. laevis* has a close relationship with *W. pigra*.

**KEY WORDS:** *Whitmania laevis, Whitmania acranulata*, mitochondrial genome, comparative analyses, phylogenetics

---

**INTRODUCTION**

The typical metazoan mitochondrial genome is a double-stranded circular DNA molecule, varying in length from 14 to 20 kb, usually composed of 36–37 genes including 12–13 protein-coding genes (PCGs), two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes (Boore, 1999). The mitochondrial genome is becoming increasingly important for phylogenetic reconstruction, due to its rapid evolutionary rate, low recombination and maternal inheritance (Elson & Lightowers, 2006; Gissi et al., 2008). The mitochondrial genome can also provide genome-level characters, such as gene order, RNA secondary structures and conserved motif for replication and transcriptional control (Boore, 2006). These useful features can be utilized by comparative genomics for phylogenetic analysis, biological identification and population studies.

Leeches are clitellate annelids with the synapomorphies of a glandular clitellum, unique sperm morphology, hermaphroditism and direct development (Rouse & Fauchald, 1995). Due to the remarkable diversity in habitats that range from terrestrial to aquatic (both marine and freshwater) environments and important role for these ecosystems, leeches have been used as environmental stress indicators (Grantham & Hann, 1994). Nonsanguivorous leeches have been used as model organisms in neurobiological and developmental studies (Ferrier, 2012; Marrec-Croq et al., 2013). Additionally, the powerful anticoagulant (hirudin) in leech salivary secretions has been of interest to the field of medicine. Some species of leeches are also used in Traditional Chinese Medicine, including *Whitmania pigra*, *W. acranulata* and *Hirudo nipponia* (Zhang et al., 2013). The morphologies of *W. pigra* and *W. laevis* are similar, and the geographical ranges of *W. laevis*,
W. pigra and W. acranulata overlap broadly in central China (TAN, 2007). A clear phylogenetic framework and correct identification are helpful to the development and conservation of these diverse leeches. Existing information in GenBank regarding Hirudinea mitochondrial genomes is inadequate for phylogenetic studies of leeches and deep understanding of evolution and characteristics of the hirudinean mitochondrial genomes.

In this study, we present the complete and nearly complete mitochondrial genome sequences of Whitmania laevis and Whitmania acranulata respectively and describe both genome features. Then, we emphasize comparative analyses among all the complete mitochondrial genomes from Whitmania and highlight unique features and shared characteristics. Finally, we analyze phylogenetic relationships among Annelida.

MATERIALS AND METHODS

Specimen collection and DNA extraction

Specimens of Whitmania laevis (WLSX) and W. acranulata (WASX) were collected at Hanbin district (32°43'N, 108°46'E), Ankang, Shaanxi, China, and preserved in 95% ethanol at 4°C. DNA was extracted from the caudal sucker muscle tissue of single individuals using a TIANamp Micro DNA Kit (tiangen Biotech, Beijing, China) according to the manufacturer’s protocol.

PCR and sequencing

Mitochondrial genomes of W. laevis (WLSX) and W. acranulata (WASX) were amplified with the primers listed in Table 1. PCR reactions were performed in a total volume of 25 μl, containing 2.5 mM MgCl₂, 2.5 μl 10 × LA PCR Buffer II (Mg²⁺ free), 0.4 mM of each dNTP, 1.25 U LA Taq polymerase, 0.4 μM of each primer, 45 ng gDNA. Cycling conditions were: an initial denaturation for 1 min at 93°C, followed by 40 cycles of 10 sec at 92°C, 30 sec at 46–57°C, 2–5 min at 68°C, and final extension of 10 min at 68°C. For nearly complete mitochondrial genome of W. acranulata, we were unable to amplify part of ATP6 and ND5 genes and the region between them with highly variable sequence and potential secondary structures. PCR products were purified with PCR Purification Kit (Sangon Biotech, Shanghai, China) and directly sequenced with the PCR primers and internal primers to complete sequences by primer walking.

Sequence analysis and Phylogenetic analyses

Contiguous sequence fragments were assembled using Staden Package v1.7.0 (STADEN et al., 2000). Protein-coding and ribosomal RNA genes were initially identified using BLAST (Basic Local Alignment Search Tool) searches on GenBank, then by alignment with the published mitochondrial genome of W. pigra GenBank no. EU304459 (WP59). The secondary structure of the two rRNA genes was determined mainly by comparison with the published rRNA secondary structures of Paragryrodactylus variegatus, Drosophila melanogaster and D. virilis (CANNONE et al., 2002; YE et al., 2014). The program tRNAscan-SE v1.21 was used to identify tRNA genes and their potential cloverleaf structures (LOWE & EDDY, 1997). The tRNAs, which were not detected by tRNA scan-SE v1.21, were identified by comparison with W. pigra. The base composition and codon usage were calculated with MEGA v5.1 (TAMURA et al., 2011). AT and GC skew were calculated according to the formulae: AT skew = (fA − fT) / (fA + fT) and GC skew = (fG − fC) / (fG + fC). To detect regions of highest variability, sliding window analyses were performed using DnaSP v5 (LIBRADO & ROZAS, 2009). A sliding window of 500 bp (in 25 bp overlapping steps) was used to estimate nucleotide diversity Pi (π) across the alignment of WLSX, WP59, W. acranulata GenBank no. KC688271 (WA71), W. laevis GenBank no. KC688269 (WL69), Hirudo nipponia GenBank no. KC667144 (HN44), Hirudinaria manillensis GenBank no.
KC688268 (HM68) and Erpobdella octoculata GenBank no. KC688270 (EO70) mitochondrial genomes. MrBayes ver.3.1.2 (RONQUIST & HUELSENBECK, 2003) and RAxML ver.7.2.8 (STAMATAKIS et al., 2005) were used to draw a maximum likelihood (ML) and bayesian inference (BI) phylogeny based on part COI gene for leeches identification, and nine concatenated PCGs (COI, COII, COIII, CYTB, ND1, ND2, ND3, ND4, ND5) and two rRNA genes (ZHONG et al., 2008) for phylogenetic relationships of Annelida. Piscicola geometra, and [Terebratalia transversa and Laqueus rubellus] were specified as the outgroups respectively. The best-fit model (GTR+Γ+I) for both datasets was estimated by ModelTest (POSADA & CRANDALL, 1998). For ML analyses, bootstrap analysis was performed with 1,000 replicates. For BI analyses, two sets of four chains were allowed to run simultaneously for 1,000,000 generations. Each set was sampled every 100 generations with a burn-in of 25%. Stationarity was considered to be reached when the average standard deviation of split frequencies was less than 0.01.

**RESULTS AND DISCUSSION**

**COI analysis of used species**

COI gene is used as a standard DNA barcoding for many animal taxa. COI gene was also confirmed as a suitable marker for biological identification, and inter- and intraspecific relationships in leeches (KOPERSKI et al., 2011; KAYGORODOVA & MANDZYAK, 2014). To evaluate the validity of species used for comparative analyses of mitochondrial genomes, the COI phylogenetic...
analysis based on all the relevant species data from GenBank was established. Both ML and BI trees showed a stable topology, which is similar to the findings of Phillips & Siddall (2009), and major internal nodes were well-supported by bootstrap values and posterior probabilities (Fig. 1). All the representatives of Hirudo nipponia, Hirudinaria manillensis and Erpobdella octoculata are clustered together respectively, except for HN44, HM68 and EO70. These three last-listed specimens lie within the cluster formed by Whitmania species. This result suggests that these three individuals may have been erroneously identified. For the genus Whitmania, the different samples from W. laevis and W. acranulata are also not found in the same branches respectively. Thus, for comparative analyses of mitochondrial genomes, we employed all the Whitmania mitochondrial genome data from GenBank including HN44, HM68 and EO70.

Genome organization and base composition

The complete mitochondrial genome of W. laevis (WLSX) (GenBank no. KM655839) is 14,442 bp in length and contains 13 PCGs, 22 tRNA genes, and two rRNA genes (Fig. 2). The nearly complete mitochondrial genome of W. acranulata (WASX) (GenBank no. KM655838) has 13,494 bp, consisting of 13 PCGs, 20 tRNA genes, and two rRNA genes. The gene order of these genes in WLSX and WASX is identical to published Whitmania mitochondrial genomes, and all the genes are transcribed from the same strand in these leeches.

The overall A + T contents of WLSX and WASX are 73.0% and 72.4% respectively, which are similar to sequenced Whitmania spp. (Table 2). Statistically, nucleotide composition can be reflected by AT skew and GC skew (Perna & Kocher, 1995). The AT skew values

Fig. 1. – Phylogenetic reconstructions based on 40 COI gene sequences of leeches. First values at the branches correspond to Bayesian posterior probabilities while the second values indicate ML bootstrap support in percentages (ML bootstrap values < 50% are not shown).
### TABLE 2

Nucleotide composition of *Whitmania* spp. mitochondrial genomes.

| Feature               | AT%          |
|-----------------------|--------------|
|                       | WLSX | WL69 | WASX | WA71 | WP59 | HN44 | HM68 | EO70 |
| Whole genome          | 73.0 | 71.9 | 72.4 | 71.6 | 72.2 | 72.6 | 72.0 | 71.6 |
| Protein-coding genes  | 72.5 | 71.1 | 71.7 | 70.8 | 71.4 | 71.7 | 71.0 | 70.7 |
| *rrnL* genes          | 73.5 | 73.2 | 73.6 | 74.1 | 73.0 | 74.5 | 75.1 | 73.0 |
| *rrnS* genes          | 72.6 | 72.3 | 72.9 | 71.3 | 72.1 | 75.1 | 75.4 | 72.3 |
| rRNA genes            | 73.1 | 72.8 | 73.3 | 73.0 | 72.7 | 74.7 | 75.2 | 72.7 |
| tRNA genes            | 75.5 | 75.9 | 76.4 | 74.6 | 75.5 | 74.5 | 74.2 | 75.2 |

| Feature               | AT−skew     |
|-----------------------|-------------|
|                       | WLSX | WL69 | WASX | WA71 | WP59 | HN44 | HM68 | EO70 |
| Whole genome          | -0.148 | -0.144 | -0.140 | -0.140 | -0.145 | -0.127 | -0.129 | -0.135 |
| Protein-coding genes  | -0.192 | -0.185 | -0.182 | -0.182 | -0.191 | -0.164 | -0.168 | -0.174 |
| *rrnL* genes          | -0.001 | -0.002 | -0.010 | 0.013 | -0.001 | -0.021 | -0.010 | 0.009 |
| *rrnS* genes          | 0.011 | 0.015 | 0.027 | 0.018 | 0.015 | 0.002 | 0.018 | 0.009 |
| rRNA genes            | 0.004 | 0.004 | 0.004 | 0.015 | 0.005 | -0.012 | 0.001 | 0.004 |
| tRNA genes            | -0.008 | -0.034 | 0.035 | -0.012 | 0.002 | -0.006 | -0.021 |     |

| Feature               | GC−skew     |
|-----------------------|-------------|
|                       | WLSX | WL69 | WASX | WA71 | WP59 | HN44 | HM68 | EO70 |
| Whole genome          | 0.180 | 0.142 | 0.128 | 0.148 | 0.155 | 0.128 | 0.126 | 0.117 |
| Protein-coding genes  | 0.168 | 0.123 | 0.109 | 0.140 | 0.144 | 0.117 | 0.108 | 0.095 |
| *rrnL* genes          | 0.205 | 0.190 | 0.211 | 0.190 | 0.179 | 0.145 | 0.172 | 0.191 |
| *rrnS* genes          | 0.228 | 0.216 | 0.168 | 0.206 | 0.216 | 0.217 | 0.198 | 0.222 |
| rRNA genes            | 0.214 | 0.200 | 0.194 | 0.197 | 0.194 | 0.173 | 0.182 | 0.203 |
| tRNA genes            | 0.232 | 0.211 | 0.204 | 0.177 | 0.215 | 0.197 | 0.199 | 0.191 |

**Note:** WLSX: *Whitmania laevis* KM655839, WL69: *Whitmania laevis* KC688269, WASX: *Whitmania acranulata* KM655838, WA71: *Whitmania acranulata* KC688271, WP59: *Whitmania pigra* EU304459, HN44: *Hirudo nipponia* KC667144, HM68: *Hirudinaria manillensis* KC688268 and EO70: *Erpobdella octoculata* KC688270.

---

**Fig. 2.** – The gene map for the mitochondrial genome of *Whitmania laevis* and *W. acranulata*. The incomplete region of *W. acranulata* is in grey.
for the encoding strand of *Whitmania* spp. mitochondrial genomes are moderate T-skew, and GC skew values are moderate G-skew. These trends of AT and GC skew are also found in PCGs. In the rRNA genes, the bias of these leeches is moderate G-skew and weak A-skew, except for HN44 with weak T-skew (-0.012). The tRNAs show moderate G-skew and weak A-skew, except for HN44 (0.002, AT skew) and WASX (0, AT skew).

**Protein-coding genes and codon usage**

Four start codons are used in the PCGs of *Whitmania* mitochondrial genomes. GTG is found in all COII except for HN44 and WASX, in all ND4L except for HN44, WA71 and WASX, in all ND5 except for WA71 and HM68, in all ND1 except for WLSX and HM68, and in ND3 for HN44, HM68, EO70 and WP59; TTG in all COIII, and the most frequent start codon is ATG in the other genes for *Whitmania* spp. In all *Whitmania* spp., five of 13 PCGs terminate with TAA (ND5, ND4L, ATP6, ND3 and CYTB, expect for ATP6 in HN44, ND5 in HN44 and WP59); ND2 terminate with incomplete-stop codons TA, and the remaining genes use the incomplete-stop codon T.

The average A + T content of PCGs for WLSX and WASX are 72.5% and 71.7%, respectively. It is similar to that of other *Whitmania* spp. (Table 2). This significant AT-richness is reflected in codon usage for mitochondrial proteins, which is similar to that observed in some other annelids (Boore, 2000; Zhong et al., 2008). In *Whitmania* mitochondrial genomes, all 64 codons in the mitochondrial genetic code table are used except for stop codon TAG in WLSX, WASX, WA71, HN44, HM68, EO70 and WP59. The most frequent amino acids in the PCGs are as follows: Leucine (15.06–15.96%), Serine (10.12–10.64%), Isoleucine (7.50–8.79%), Phenylalanine (8.00–8.60%), and Methionine (7.67–8.57%). UUA (Leucine), AUU (Isoleucine), UUU (Phenylalanine) and AUA (Methionine) are the most frequently used codons (Table 4).

**Transfer RNA and ribosomal RNA genes**

The length of large ribosomal subunit (*rrnL*) is 1,139 bp in WLSX and 1,133 bp in WASX, with an A + T content of 73.5% and 73.6%, respectively. The small ribosomal subunit (*rrnS*) is 736 bp in WLSX and 726 bp in WASX, and the A + T content is 72.6% and 72.9% for WLSX and WASX, respectively. The predicted secondary structure of *rrnL* and *rrnS* of WLSX is shown in Fig. 3 and Fig. 4, respectively. The secondary structure of *rrnL* contains six domains and 43 helices. But domain III is absent, which was reported in secondary structure of other invertebrate *rrnL* (Domes et al., 2008; Liu & Huang, 2010; Li et al., 2013). Among *Whitmania* spp. mitochondrial genomes, domains IV and V are more conserved than domains I, II, and VI. Overall, some helices (H235, H533, H589, H671, H687, H837, H946, H1057, H1196, H1648, H2023, H2347, H2675, and H2735) are greatly variable regions. The secondary structure of *rrnS* contains three domains and 27 helices. The domain III is more conserved than domains I and II. In domains I and II, conservative sites are mainly in helices H9, H367, H511, H769, H885 and loop of H673.

All of the 22 tRNA genes typical of metazoan mitochondrial genomes were identified in WLSX mitochondrial genome, while 20 tRNA genes were identified in WASX. All present tRNAs can be folded into the typical cloverleaf structure with the exception of *tRNA-Pro* and *tRNA-Gly* (Fig. 5). In *tRNA-Pro* and *tRNA-Gly*, the TψC arm simply forms a loop. In addition, the TψC arm of other five tRNAs (*tRNA-Ala*, *tRNA-Met*, *tRNA-Trp*, *tRNA-Tyr* and *tRNA-Val*) is short with only one complementary base pair. The level of nucleotide conservation in tRNA genes is markedly different. The highest levels of nucleotide conservation occur in *tRNA-Pro*, *tRNA-Ala(tur)*, *tRNA-Fpr*, *tRNA-Fyb* and *tRNA-Val*. However, *tRNA-Arg*, *tRNA-His* and *tRNA-Thr* show low levels of identical nucleotides among *Whitmania* spp.
### TABLE 3

Annotation of the mitochondrial genomes of *Whitmania laevis* and *W. acranulata* (continued on next page).

| Gene         | From (bp) | To (bp) | Size (bp) | Start Codon | Stop Codon | Anticodon |
|--------------|-----------|---------|-----------|-------------|------------|-----------|
| **Whitmania laevis** |           |         |           |             |            |           |
| COI          | 1         | 1534    | 1534      | ATG         | T          | GTT       |
| tRNA-Asn (N) | 1535      | 1596    | 62        |             |            |           |
| COII         | 1597      | 2275    | 679       | GTG         | T          | GTC       |
| tRNA-Asp (D) | 2276      | 2339    | 64        |             |            |           |
| ATP8         | 2340      | 2490    | 151       | ATG         | T          | TCC       |
| tRNA-Gly (G) | 2491      | 2549    | 59        |             |            |           |
| tRNA-Tyr (Y) | 2550      | 2610    | 61        |             |            | GTA       |
| COIII        | 2622      | 3402    | 781       |             |            |           |
| tRNA-Glu (Q) | 3403      | 3471    | 69        |             |            |           |
| ND6          | 3472      | 3928    | 457       | ATG         | T          | TTG       |
| CYTB         | 3929      | 5074    | 1146      | ATG         | TAA        |           |
| tRNA-Trp (W) | 5080      | 5140    | 61        |             |            | TCA       |
| ATP6         | 5204      | 5908    | 705       | ATG         | TAA        |           |
| tRNA-Arg (R) | 5908      | 5970    | 63        |             |            | TCG       |
| tRNA-His (H) | 6079      | 6139    | 61        |             |            | GTG       |
| ND5          | 6140      | 7835    | 1696      | GTG         | T          |           |
| tRNA-Phe (F) | 7836      | 7897    | 62        |             |            | GAA       |
| tRNA-Glu (E) | 7898      | 7958    | 61        |             |            | TTC       |
| tRNA-Pro (P) | 7956      | 8016    | 61        |             |            | TGG       |
| tRNA-Thr (T) | 8019      | 8078    | 60        |             |            | TGT       |
| ND4L         | 8079      | 8366    | 288       | GTG         | TAA        |           |
| ND4          | 8360      | 9029    | 1333      | ATG         | T          |           |
| tRNA-Cys (C) | 9702      | 9762    | 61        |             |            | GCA       |
| tRNA-Met (M) | 9763      | 9825    | 63        |             |            | CAT       |
| rrnS (12S)   | 9826      | 10561   | 736       |             |            |           |
| tRNA-Val (V) | 10562     | 10623   | 62        |             |            | TAC       |
| rrnL (16S)   | 10624     | 11762   | 1139      |             |            |           |
| tRNA-Leu<sup>CCU</sup> (L1) | 11763     | 11823   | 61        |             |            | TAG       |
| tRNA-Ser<sup>CCU</sup> (S2) | 11823     | 11890   | 68        |             |            | TGA       |
| tRNA-Ala (A) | 11891     | 11950   | 60        |             |            | TGC       |
| tRNA-Leu<sup>CCU</sup> (L2) | 11951     | 12011   | 61        |             |            | TAA       |
| ND1          | 12012     | 12930   | 919       | ATG         | T          |           |
| tRNA-Ile (I) | 12931     | 12992   | 62        |             |            | GAT       |
| tRNA-Lys (K) | 12994     | 13055   | 62        |             |            | TTT       |
| ND3          | 13057     | 13401   | 345       | ATG         | TAA        |           |
| tRNA-Ser<sup>AGU</sup> (S1) | 13388     | 13454   | 67        |             |            | TCT       |
| ND2          | 13455     | 14437   | 983       | ATG         | TA         |           |
| **Whitmania acranulata** |           |         |           |             |            |           |
| tRNA-Phe (F) | 1260      | 1321    | 62        |             |            | GAA       |
| tRNA-Glu (E) | 1322      | 1380    | 59        |             |            | TTC       |
| tRNA-Pro (P) | 1378      | 1436    | 59        |             |            | TGG       |
| tRNA-Thr (T) | 1438      | 1497    | 60        |             |            | TGT       |
| ND4L         | 1498      | 1785    | 288       | ATG         | TAA        |           |
| ND4          | 1779      | 3111    | 1333      | ATG         | T          |           |
| tRNA-Cys (C) | 3121      | 3181    | 61        |             |            | GCA       |
| tRNA-Met (M) | 3182      | 3243    | 62        |             |            | CAT       |
| rrnS (12S)   | 3244      | 3969    | 726       |             |            | TAC       |
| tRNA-Val (V) | 3970      | 4035    | 66        |             |            |           |
| rrnL (16S)   | 4036      | 5168    | 1133      |             |            |           |
| tRNA-Leu<sup>CCU</sup> (L1) | 5172      | 5231    | 60        |             |            | TAG       |
| tRNA-Ser<sup>CCU</sup> (S2) | 5231      | 5298    | 68        |             |            | TGA       |
| tRNA-Ala (A) | 5299      | 5358    | 60        |             |            | TGC       |
| tRNA-Leu<sup>CCU</sup> (L2) | 5359      | 5419    | 61        |             |            | TAA       |
### Table 2: Gene Locations and Characteristics

| Gene     | From (bp) | To (bp) | Size (bp) | Start Codon | Stop Codon | Anticodon |
|----------|-----------|---------|-----------|-------------|------------|-----------|
| ND1      | 5420      | 6338    | 919       | GTG         | T          | GAT       |
| tRNA-Ile (I) | 6339    | 6400    | 62        |             |            | GAT       |
| tRNA-Lys (K) | 6401     | 6462    | 62        |             |            | TTT       |
| ND3      | 6464      | 6808    | 345       | ATG         | TAA        | TCT       |
| tRNA-Ser (AGN) (S1) | 6795    | 6861    | 67        |             |            |           |
| ND2      | 6862      | 7844    | 983       | ATG         | TA         | T         |
| COI      | 7850      | 9383    | 1534      | ATG         | T          |           |
| tRNA-Asn (N) | 9384    | 9445    | 62        |             |            | GTT       |
| COII     | 9446      | 10127   | 682       | ATG         | T          | GTC       |
| tRNA-Asp (D) | 10128    | 10193   | 66        |             |            |           |
| ATP8     | 10194     | 10341   | 148       | ATG         | T          |           |
| tRNA-Gly (G) | 10345    | 10404   | 60        |             |            | TCC       |
| tRNA-Tyr (Y) | 10405   | 10464   | 60        |             |            | GTA       |
| COIII    | 10450     | 11245   | 796       | TTG         | T          | TTG       |
| tRNA-Gln (Q) | 11246    | 11314   | 69        |             |            |           |
| ND6      | 11314     | 11775   | 462       | ATG         | T          |           |
| CYTB     | 11776     | 12921   | 1146      | ATG         | TAA        | TCA       |
| tRNA-Trp (W) | 12925    | 12984   | 60        |             |            |           |
| ATP6     | 13040     | 13494   | 455       | ATG         |            |           |

### Non-coding regions

*Whitmania* spp. mitochondrial genomes are highly compacted in genome size as in other animals (Boore, 1999). A total of 7 short non-coding regions were identified ranging from 1 bp to 11 bp in the mitochondrial genome of WLSX (Table 3). There are two major non-coding regions (NCR1 and NCR2) in the same positions of HN44, WL69 and WP59 mitochondrial genome, while the remaining ones have one non-coding region (NCR2). NCR1 and NCR2 are located between tRNA_{Trp} and ATP6, and tRNA_{Arg} and tRNA_{His}, respectively. The NCR1 and NCR2 are too variable for alignments, but the sequence similarity of NCR1 between WLSX and WP59 is 63.4% and the NCR2 has 53.2% sequence similarity. The NCR1 contains two stem-loop

---

**Fig. 3.** Inferred secondary structure of the mitochondrial *rrnL* gene for *Whitmania laevis*. Conserved nucleotides of all *Whitmania* taxa are labelled in grey.
| Codon (AA) | WLSX   | WL69   | WASX   | WA71   | WP59   | HN44   | HM68   | EO70   |
|-----------|--------|--------|--------|--------|--------|--------|--------|--------|
| UUU(U)    | 277    | 263    | 264    | 262    | 265    | 279    | 257    | 257    |
| UUC(U)    | 31     | 46     | 28     | 37     | 38     | 32     | 37     | 38     |
| UUG(U)    | 84     | 91     | 71     | 92     | 101    | 73     | 97     | 92     |
| UCU(U)    | 47     | 52     | 38     | 51     | 50     | 45     | 48     | 58     |
| UCA(U)    | 4      | 10     | 6      | 11     | 12     | 16     | 13     |        |
| UCU(U)    | 40     | 57     | 51     | 52     | 55     | 61     | 55     | 61     |
| UCC(U)    | 10     | 10     | 14     | 9      | 10     | 11     | 13     |        |
| AUC(U)    | 292    | 264    | 263    | 243    | 270    | 261    | 268    | 265    |
| AUC(U)    | 29     | 37     | 36     | 31     | 38     | 35     | 38     | 39     |
| AUC(U)    | 235    | 211    | 232    | 220    | 223    | 230    | 226    | 216    |
| AUC(U)    | 71     | 63     | 79     | 74     | 70     | 72     | 73     | 72     |
| GCC(U)    | 9      | 9      | 15     | 14     | 11     | 9      | 17     | 16     |
| GCC(U)    | 108    | 112    | 97     | 123    | 115    | 107    | 117    | 103    |
| GCU(U)    | 16     | 17     | 16     | 15     | 17     | 13     | 17     | 18     |
| GUA(U)    | 119    | 121    | 111    | 129    | 133    | 120    | 131    | 122    |
| GUC(U)    | 39     | 30     | 31     | 30     | 35     | 27     | 27     | 32     |
| UCU(U)    | 99     | 95     | 103    | 106    | 99     | 98     | 95     | 97     |
| UCC(U)    | 25     | 14     | 21     | 19     | 20     | 21     | 24     |        |
| UAC(A)    | 101    | 90     | 85     | 93     | 95     | 93     | 105    |        |
| AUG(U)    | 9      | 4      | 45     | 39     | 17     | 17     | 17     |        |
| ACA(U)    | 55     | 56     | 51     | 50     | 51     | 53     | 55     | 49     |
| ACC(U)    | 3      | 10     | 12     | 10     | 11     | 4      | 17     | 12     |
| ACC(U)    | 66     | 66     | 65     | 61     | 59     | 64     | 68     | 64     |
| CCU(P)    | 11     | 18     | 7      | 18     | 18     | 17     | 16     |        |
| AGU(U)    | 64     | 52     | 70     | 69     | 76     | 68     | 71     | 71     |
| AGC(U)    | 9      | 15     | 7      | 14     | 11     | 9      | 17     | 14     |
| AAC(A)    | 46     | 45     | 46     | 43     | 43     | 55     | 49     | 59     |
| ACA(A)    | 11     | 7      | 12     | 10     | 11     | 10     | 10     |        |
| AGC(A)    | 5      | 60     | 71     | 63     | 61     | 64     | 69     | 62     |
| AGG(S)    | 9      | 15     | 24     | 20     | 22     | 16     | 22     | 23     |
| AGA(S)    | 50     | 46     | 47     | 43     | 47     | 42     | 44     |        |
| UAA(*)    | 8      | 9      | 7      | 10     | 8      | 9      | 9      |        |
| UAG(*)    | 148    | 148    | 138    | 148    | 132    | 132    | 135    |        |
| CUG(A)    | 22     | 29     | 37     | 35     | 34     | 35     | 35     |        |
| CUA(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUU(F)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUC(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUG(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUU(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUC(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUG(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |
| CUG(L)    | 0      | 0      | 0      | 0      | 0      | 0      | 0      |        |

**Notes:** WLSX: *Whitmania* KM655839, WL69: *Whitmania laevis* KC688269, WASX: *Whitmania acranulata* KM655838, WA71: *Whitmania acranulata* KC 688271, WP59: *Whitmania pigra* EU304459, HN44: *Hirudo nipponia* KC667144, HM68: *Hirudinaria manillensis* KC688268, EO70: *Erpobdella octoculata* KC688270 and AA: amino acid.
structures at positions 4-21 bp and 27-45 bp in WLSX. Two stem-loop structures were also found in NCR2. The conserved sequences of both NCR1 and NCR2 between WLSX and WP59 mainly occur in the stem-loop structures. Tandem repeat sequences commonly observed in other invertebrate lineages (Zhang & Hewitt, 1997) were not found in NCR1 and NCR2 for *Whitmania* mitochondrial genomes.

**Sliding window analyses and nucleotide diversity**

Sliding window analysis was performed to estimate nucleotide diversity $\pi$ for the mitochondrial genome of *Whitmania*. Not unexpectedly, the most variable regions were found in the major non-coding regions (Fig. 6). The sliding window indicated that the most

---

Fig. 4. – Inferred secondary structure of the mitochondrial *rrnS* gene for *Whitmania laevis*. Conserved nucleotides of all *Whitmania* taxa are labelled in grey.
Fig. 5. – The inferred secondary structures of mitochondrial tRNA genes of *Whitmania laevis*.

Fig. 6. – Sliding window analyses of the alignment among *Whitmania* spp. mitochondrial genomes. The line shows the value of nucleotide diversity (\(\pi\)) in a sliding window analysis of window size 500 bp with step size 25; the value is inserted at its mid-point.
variable coding regions were within the genes ATP6 and 5’ part of ND5 (Fig. 6). Amongst PCGs the most conserved gene fragments are the 3’ end of COII, ND6 and 5’ part of CYTB. By contrast, the most variable regions in ATP6, ND5 and ND4 genes can be used as effective markers to investigate relationships of populations and the closely related species.

Phylogenetic analyses

Annelida, the segmented worms, traditionally includes two taxonomic groups, namely clitellates and polychaetes. Recently, analyses of molecular data indicate Annelida may contain several other phyla (STRUCK et al., 2007; ZrZAVÝ et al., 2009), but the evolution and phylogeny of Annelida is still controversial. In Euhirudinea, although the relationships within Hirudiniformes have been extensively investigated (APAKUPAKUL et al., 1999; BORDA & SIDDALL, 2004; BORDA et al., 2008; PHILLIPS & SIDDALL, 2009), few relationships of closely related species within *Whitmania* have as yet been clearly elucidated. In order to infer phylogenetic relationships of annelids, especially for these closely related species within *Whitmania*, the nucleotide dataset of concatenated nine PCGs and two rRNA genes were employed for phylogenetic analysis. Both ML and BI analysis showed similar tree topologies (Fig. 7). The results of the *Whitmania* branch revealed that *W. laevis* and *W. pigra* were closely related with high statistical support without considering the uncertain species HN44, HM68, EO70. Our results of *Whitmania* (*W. acranulata* + (*W. laevis* + *W. pigra*)) differ from the results of XU et al. (2013) based on only three mitochondrial genes. Compared with reported molecular phylogenies (ROUSSET et al., 2007; STRUCK et al., 2007; SHEN et al., 2009), Clitellata appears consistently as a monophyletic group; Sipunculans form a sister group of annelids (including echiurans); *Clymenella torquata* (Capitellida) clusters with two Terebellida species. With greater numbers of species in mitochondrial genomic analyses, the phylogenetic positions of Echiurida and some groups within

![Phylogenetic tree inferred from nine PCGs and two rRNA genes using BI and ML analysis. First values at the branches correspond to Bayesian posterior probabilities while the second values indicate ML bootstrap support in percentages (ML bootstrap values < 50% are not shown).](image-url)
Polychaeta appear quite different (ZHONG et al., 2008; SHEN et al., 2011). The Echiurida and Clitellata cluster together as a sister clade and the branch consists of the cluster Maldanidae/Terebellida, Marphysa sanguinea (Eunicidae), Orbitia latreillii (Orbiniidae) and Nephtys sp. (Nephtyidae) with low nodal support suggesting that their relationships still need to be investigated with a broader taxonomic sample. Furthermore, differing topologies derived from nuclear and mitochondrial data sets indicate the need for more investigation of the “symplesiomorphy trap” in Annelida (ZHONG et al., 2011).

CONCLUSIONS

The mitochondrial genomes of *W. laevis* and *W. acramulata* display identical genome organization and gene order to previously reported *Whitmania* mitochondrial genomes. Comparative analyses of *Whitmania* mitochondrial genomes reveal: (i) the nucleotide composition is significantly biased toward A and T; (ii) the significant AT-richness is reflected in codon usage with frequent UUA, AUA, UUU, and AUU; (iii) the TyC arm of five tRNAs (tRNA\(^{Ala}\), tRNA\(^{Val}\), tRNA\(^{Tyr}\), tRNA\(^{Trp}\) and tRNA\(^{Val}\)) is short with only one complementary base pair; (iv) domain III in rrnS and domains IV and V in rrnL are the most conserved parts. The sliding window analysis reveals that ND4, ND5 and ATP6 genes may serve as useful markers to investigate relationships of population and of closely related species. The phylogenetic analysis based on nine PCGs and two rRNA genes confirms *W. laevis* and *W. pigra* are closely related with high statistical support. The comparative analyses of *Whitmania* mitochondrial genomes could provide more information for understanding of the characteristics and evolution of the *Whitmania* mitochondrial genomes.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Yuan Huang (College of Life Science, Shaanxi Normal University) for assistance with the phylogenetic analyses and Dr Zhongmin Dong (Department of Biology, Saint Mary’s University) for editorial comments on an early draft of the manuscript. This research was supported by the National Natural Science Foundation of China (31372158).

REFERENCES

APAKUPAKUL K, SIDDALL ME & BURRESON EM (1999). Higher level relationships of leeches (Annelida: Clitellata: Euhirudinea) based on morphology and gene sequences. Molecular Phylogenetics and Evolution, 12: 350–359.

BOORE JL (1999). Animal mitochondrial genomes. Nucleic Acids Research, 27: 1767–1780.

BOORE JL (2006). The use of genome-level characters for phylogenetic reconstruction. Trends in Ecology and Evolution, 21: 39–446.

BOORE JL & BROWN WM (2000). Mitochondrial genomes of Galathealinum, Helobdella, and Platyneurius: Sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Molecular Biology and Evolution, 17: 87–106.

BORDA E, OCEGUERA-FIGUEROA A & SIDDALL ME (2008). On the classification, evolution and biogeography of terrestrial haemadipsoeid leeches (Hirudinida: Arhynchobdellida: Hirudiniformes). Molecular Phylogenetics and Evolution, 46: 142–154.

BORDA E & SIDDALL ME (2004). Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. Molecular Phylogenetics and Evolution, 30: 213–225.

CANNONE JJ, SUBRAMANIAN S, SCHNARE MN, COLLETT JR, D’SOUZA LM, DU Y, FENG B, LIN N, MADABUSI LV, MÜLLER KM, Pande N, SHANG Z, YU N & GUTELL RR (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.

DOMES K, MARAUN M, SCHEU S & CAMERON SL (2008). The complete mitochondrial genome of the sexual oribatid mite Steganacarus magnus: genome rearrangements and loss of tRNAs. BMC Genomics, 9: 532.
ELSON JL & LIGHTOWLERS RN (2006). Mitochondrial DNA clonality in the dock: can surveillance swing the case? Trends in Genetics, 22: 603–607.

FERRIER DE (2012). Evolutionary crossroads in developmental biology: annelids. Development, 139: 2643–2653.

GISSI C, IANNELLI F & PESOLE G (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity, 101: 301–320.

GRANTHAM BA & HANN BJ (1994). Leeches (Annelida: Hirudinea) in the experimental lakes area, Northwestern Ontario, Canada: Patterns of species composition in relation to environment. Canadian Journal of Fisheries and Aquatic Sciences, 51: 1600–1607.

JENNINGS RM & HALANYCH KM (2005). Mitochondrial genomes of Climenella torquata (Maldanidae) and Rifta pachyprila (Siboglinidae): evidence for conserved gene order in Annelida. Molecular Biology and Evolution, 22: 210–222.

KAYGORODOVA IA & MANDZYAK NB (2014). Molecular phylogeny of siberian Glossiphoniidae (Hirudinea). Molecular Biology, 48: 452–455.

KOPERSKI P, MILANOWSKI R & KRZYK A (2011). Searching for cryptic species in Erpobdella octoculata (L.) (Hirudinea: Clitellata): discordance between the results of genetic analysis and cross-breeding experiments. Contributions to Zoology, 80: 85–94.

LI T, GAO C, CUI Y, XIE Q & BU W (2013). The Complete Mitochondrial Genome of the Stalk-Eyed Bug Chauliops fallax Scott, and the Monophyly of Malciade (Hemiptera: Heteroptera). PLoS ONE, 8: e55381.

LIBRADO P & ROZAS J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25: 1451–1452.

LIU N & HUANG Y (2010). Complete mitochondrial genome sequence of Acrida cinerea (Acrididae: Orthoptera) and comparative analysis of mitochondrial genomes in Orthoptera. Comparative and Functional Genomics, 2010: 319486.

LOWE TM & EDDY SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research, 25: 955–964.

MARREC-CROQ FL, DRAGO F, VIZIOLI J, SAUTIÈRE PE & LEFEBVRE C (2013). The leech nervous system: A valuable model to study the microglia involvement in regenerative processes. Clinical and Developmental Immunology, 2013: 274019.

PERNA NT & KOCHER TD (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution, 41: 353–358.

PHILLIPS AJ & SIDDALL ME (2009). Poly-paraphyly of Hirudinidae: many lineages of medicinal leeches. BMC Evolutionary Biology, 9: 246.

POSADA D & CRANDALL KA (1998). Modeltest: Testing the model of DNA substitution. Bioinformatics, 14: 817–818.

RONQUIST F & HUELSENBECK JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19: 1572–1574.

ROUSE GW & FAUCHALD K (1995). The articulation of annelids. Zoologica Scripta, 4: 269–301.

ROUSSET V, PLEIJEL F, ROUSE GW, ERSEUS C & SIDDALL ME (2007). A molecular phylogeny of annelids. Cladistics, 23: 41–63.

SAKAI M & SAKAIZUMI M (2012). The complete mitochondrial genome of Dugesia japonica (Platyhelminthes; Order Tricladida). Zoological Science, 29: 672–680.

SHEN X, MA X, REN J & ZHAO F (2009). A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of Phascosoma esculenta. BMC Genomics, 10: 136.

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. Comparative Biochemistry and Physiology D, 6: 133–138.

SIMON C, BUCKLEY TR, FRATI F, STEWART JB & BECKENBACH AT (2006). Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annual Review of Ecology Evolution and Systematics, 37: 545–579.

STADEN R, BEAL KF & BONFIELD JK (2000). The Staden package, 1998. Methods in Molecular Biology, 132: 115–130.

STAMATAKIS A, LUDWIG T & MEIER H (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics, 21: 456–463.
STRUCK TH, SCHULT N, KUSEN T, HICKMAN E, BLEIDORN C, MCHUGH D & HALANYCH KM (2007). Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evolutionary Biology, 7: 57.

TAMURA K, PETERSON D, PETERSON N, STECHER G, NEI M & KUMAR S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28: 2731–2739.

TAN N (2007). Zoogeography of Hirudinidae in China. Acta Scientiarum Naturalium Universitatis Sunyatseni, 46: 100–104.

XU Y, NIE J & XIAO L (2013). Molecular evolution analysis of COI, 12S rRNA and 16S rRNA gene in six species of leech. Journal of Biology, 30: 10–13.

YE F, KING SD, CONE DK & YOU P (2014). The mitochondrial genome of Paragyradactylus variegatus (Platyhelminthes: Monogenea): differences in major non-coding region and gene order compared to Gyrodactylus. Parasites & Vectors, 7: 377.

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.

ZHONG M, HANSEN B, NESNIDAL M, GOLOMBEK A, HALANYCH KM & STRUCK TH (2011). Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. BMC Evolutionary Biology, 11: 369.

ZHANG W, ZHANG R, LI J, LIANG F & QIAN Z (2013). Species study on Chinese medicine leech and discussion on its resource sustainable utilization. China Journal of Chinese material medica, 38: 914–918.

ZHANG DX & HEWITT GM (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics and Ecology, 25: 99–120.

ZRZAVÝ J, RÍHA P, PÍAŁEK L & JANOUŠKOVEC J (2009). Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. BMC Evolutionary Biology, 9: 189.

Received: September 29th, 2014
Accepted: March 24th, 2015
Branch editor: Kurt Jordaens