Unraveling the phylogeny of Chaetopteridae (Annelida) through mitochondrial genome analysis

Xiu Wu, Xuechun Su, Yu Wang, Caifang He, Jian-Wen Qiu and Yanjie Zhang

1 School of Life and Health Sciences, Hainan University, Haikou, China, 2 Department of Biology, Hong Kong Baptist University, Hong Kong, Hong Kong SAR, China

Mitochondrial genomes serve as valuable markers for phylogenetic and evolutionary studies across diverse invertebrate taxa, but their application within Annelida remains limited. In this study, we report the mitochondrial genomes of seven species from four genera of Chaetopteridae (Annelida), obtained by high-throughput sequencing. Phylogenetic analysis was performed using COXI, 18S, 28S and all mitochondrial genes. Our results reveal Chaetopterus and Mesochaetopterus as well-supported monophyletic sister clades, while Phyllochaetopterus and Spiochaetopterus appear paraphyletic, with species from both genera in a mixed clade sister to Chaetopterus + Mesochaetopterus. While mitochondrial gene orders remain conserved within Chaetopteridae, they appear substantially different from those of the ancestral patterns in Annelida. All 13 protein-coding genes found in Chaetopteridae evolved under strong purifying selection, although Phyllochaetopterus exhibited the highest base-substitution rate for most of them, suggesting a more relaxed purifying selection. Overall, our study provides molecular resources for phylogenetic studies of Chaetopteridae, highlighting the necessity for a comprehensive revision of the family, particularly dealing with the paraphyletic Phyllochaetopterus and Spiochaetopterus.

1 Introduction

Chaetopteridae is a small family of the phylum Annelida, whose species live in self-secreted membranous tubes and are commonly found in different habitats from the intertidal to the deep sea (Moore et al., 2017; Britayev and Martin, 2019; Rouse et al., 2022). While most of them live buried in soft sediment, some are attached to rocks, either living alone or in groups, and one species is holoplanktonic (Blake, 1996;
analyses based on 18S rRNA of Spiochaetopterus Sars, 1856 (Read and Fauchald, 2023). In molecular phylogenetic studies, Trochochaetidae, among others, with Trochochaetidae as its sister family (Rouse and Fauchald, 1997). In molecular phylogenetic analyses based on 18S rRNA, Chaetopteridae variopedatus (Chaetopteridae) was recovered as sister to a clade containing two species of Brachiopoda and two of Phoronida, which altogether, were sister to Dendocarcin concharum (Spionidae) (McHugh, 2000).

Later studies with more genes (Colgan et al., 2006; Rosset et al., 2007), including transcriptomes (Struck, 2011; Weigert et al., 2014) or mitochondrial genomes (mtgenome) (Weigert et al., 2016; Struck et al., 2023), usually recovered Chaetopteridae as one of the basal groups of Annelida. However, the relationships among the basal annelid groups (i.e., Amphinomidae, Magelonidae, Myzostomida, Oweniidae, Spionidae, and Sipuncula) and between annelids and other invertebrates (i.e., Chaetopteridae, only one species each in two genera: Chaetopterus Cuvier, 1830, Mesochaetopterus Potts, 1914, Phyllochaetopterus Grube, 1863, and Spiochaetopterus Sars, 1856 (Read and Fauchald, 2023).

The phylogenetic position of Chaetopteridae in the tree of life of Annelida has not been stable over the last three decades. Based on morphology, they were considered a family within Spionida, together with the Spionidae, Poecilichaetidae, and Trochochaetidae, among others, with Trochochaetidae as its sister family (Rouse and Fauchald, 1997). In molecular phylogenetic analyses based on 18S rRNA, Chaetopterus variopedatus (Chaetopteridae) was recovered as sister to a clade containing two species of Brachiopoda and two of Phoronida, which altogether, were sister to Dendocarcin concharum (Spionidae) (McHugh, 2000).

Later studies with more genes (Colgan et al., 2006; Rosset et al., 2007), including transcriptomes (Struck, 2011; Weigert et al., 2014) or mitochondrial genomes (mtgenome) (Weigert et al., 2016; Struck et al., 2023), usually recovered Chaetopteridae as one of the basal groups of Annelida. However, the relationships among the basal annelid groups (i.e., Amphinomidae, Magelonidae, Myzostomida, Oweniidae, Spionidae, and Sipuncula) and between annelids and other invertebrates (i.e., Nemertea, Brachiopoda, Mollusca) remained unstable.

Within Chaetopteridae, there have also been inconsistencies in the monophyly of the four traditionally recognized genera, as well as in the phylogenetic relationships among them. Chaetopterus and Mesochaetopterus were paraphyletic based on cox1 only (Morineaux et al., 2010; Zhang et al., 2013), but monophyletic based on the combined dataset of cox1 and 18S (Martin et al., 2022) and cox1, 18S and 28S genes (Zhang et al., 2015; Moore et al., 2017), while the same authors found Phyllochaetopterus and Spiochaetopterus as paraphyletic, forming a sister clade to Mesochaetopterus + Chaetopterus (Osborn et al., 2007; Zhang et al., 2015). Therefore, more taxon sampling within each chaetopterid genus and more genetic data for each species are required to provide a well-resolved phylogenetic tree of Chaetopteridae and to properly place this family in the annelid tree of life.

Several studies have utilized transcriptomes and mtgenomes to reconstruct the phylogeny of annelids. However, within Chaetopteridae, only one species each in Chaetopterus, Phyllochaetopterus, and Mesochaetopterus have a sequenced mtgenome (Weigert et al., 2016; Yang et al., 2022), which limits assessing its utility both in annelid phylogeny or mtgenome evolution, such as gene order rearrangement, gene duplication and loss, codon usage and mutation rate (Bernt et al., 2013b; Halanych, 2016; Li et al., 2019). The gene order of mitogenomes was initially considered conserved in Annelida, including in the deep-sea tube worm Riftia pachyptila Jones, 1981, which shows remarkable body structure changes, such as loss of the digestive tract (Jennings and Halanych, 2005). However, more and more

### 2 Materials and methods

#### 2.1 Sample collection and genome sequencing

Specimens of Chaetopterus qiani Sun and Qiu, 2014, Chaetopterus sp., Phyllochaetopterus sp., Phyllochaetopterus hainanensis Wang and Li, 2017, Mesochaetopterus tingkokensis Zhang et al., 2015, Spiochaetopterus sp. A and Spiochaetopterus sp. B were collected from the intertidal and subtidal of Hong Kong and Hainan, China (Table 1, Figure 1) and preserved in 95% ethanol for DNA extraction or 4% formalin for morphological observations. Genomic DNA for each species was extracted using the CTAB method. Paired-end sequencing on the Illumina platform was performed at Novegene or Sangon to obtain 3.67 to 8.86 Gb bp data (Supplementary Table 1).

#### 2.2 Mtgenome assembly

Assembly for all species except P. hainanensis and Spiochaetopterus sp. B was performed using CLC Genomics Workbench v7.03 (CLCbio, Aarhus, Denmark). For P. hainanensis and Spiochaetopterus sp. B, SPAdes-3.15.4 (Prijibelski et al., 2020) was applied for contig assembly using multiple kmer settings (kmer = 21, 33, 55, 77, 99, 127), and NOVOPlasty v2.7.0 (Dierckxsens et al., 2017) and GetOrganelle (Jin et al., 2020) were run under default settings to build scaffolds, one of which was found to be the full mtgenome.

#### 2.3 Mtgenome annotation

The chaetopterid mitochondrial or mtgenome or 18S/28S rRNA were detected using BLAST V2.13.0 (Camacho and Madden, 2013). The mtgenomes were annotated on the MITOS web server using genetic code 05 for invertebrates (Bert et al., 2013c). The boundaries of protein-coding genes (PCGs) and tRNA genes were manually examined and adjusted based on alignment with published mtgenomes of this family. The GC-skew and AT-skew were defined according to Perna and Kocher (1995): AT skew = (A − T)/(A + T) and GC skew = (G − C)/(G + C).
2.4 Phylogenetic analyses

A three-gene dataset (mitochondrial cox1, nuclear 18S and 28S rRNA genes) contained 50 species, including 41 published in Moore et al. (2017) and Britayev et al. (2017), seven sequenced in this study, and two as outgroups, *Eurythoe complanata* (Pallas, 1766) and *Owenia fusiformis* Delle Chiaje, 1844, which were used in our phylogenetic analyses (Supplementary Table 2). The sequences of each gene were aligned using the MAFFT version 7 web server (Katoh et al., 2017), and concatenated using PhyloSuite (Zhang et al., 2020). Poorly arranged locations and very dispersive regions were removed using less stringent selection settings of Gblocks Server which include smaller final blocks, gap positions within the final blocks, and less strict flanking positions (Talavera and Castresana, 2007). The best nucleotide evolutionary model for each partition was selected based on the Akaike information criterion (AIC) (Darriba et al., 2012) of the PartitionFinder2 module in PhyloSuite. Phylogenetic analysis was performed using Maximum Likelihood (ML) and Bayesian inference (BI) respectively executed using the IQ-TREE module in PhyloSuite with 10,000 Ultrafast Bootstrap (UFBoot) replicates with the SH-aLRT test (Minh et al., 2013) and the MrBayes module in PhyloSuite with 1000 Sampling Freq replicates.

Phylogenetic analyses were also conducted using all mitochondrial genes (13 PCGs, 2 rRNAs, and 22 tRNAs) for all mtgenomes, except in *Spiochaetopterus* sp. A which contained only 3 PCGs and 11 tRNAs. *Eurythoe complanata* (Pallas, 1766) (accession number KT726962.1) and *Owenia fusiformis* (Delle Chiaje, 1844) (accession number NC_028712.1) were used as the outgroups. The saturation of genetic sequences was assessed using

![FIGURE 1](image-url)

*Species sequenced in this study: (A) Spiochaetopterus sp. A; (B) Spiochaetopterus sp. B; (C) Mesochaetopterus tingkokensis; (D) Phyllochaetopterus sp.; (E) Phyllochaetopterus hainanensis; (F) Chaetopterus qiani; (G) Chaetopterus sp.*

### TABLE 1 Collection information of specimens used in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>Collected date</th>
<th>Location</th>
<th>Coordinates</th>
<th>Deposited at*</th>
<th>No. of specimens</th>
<th>Preserved in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chaetopterus qiani</td>
<td>April, 2015</td>
<td>Hoi Ha Wan, Hong Kong</td>
<td>22.281°N, 114.202°E</td>
<td>Hainan University</td>
<td>&gt;20</td>
<td>Ethanol and formalin</td>
</tr>
<tr>
<td>Chaetopterus sp.³</td>
<td>September 08, 2014, April 20, 2015</td>
<td>Tai Tam, Hong Kong</td>
<td>22.243°N, 114.2235°E</td>
<td>Hainan University</td>
<td>10</td>
<td>Tissue in Ethanol and specimen in formalin</td>
</tr>
<tr>
<td>Mesochaetopterus tingkokensis</td>
<td>May 28, 2014</td>
<td>Ting Kok, Hong Kong</td>
<td>22.280°N, 114.124°E</td>
<td>Institute of Oceanology, Chinese Academy of Science (IOCAS)</td>
<td>14</td>
<td>Ethanol and formalin</td>
</tr>
<tr>
<td>Phyllochaetopterus sp.⁴</td>
<td>June 03, 2021</td>
<td>Off western Lamma Island, Hong Kong</td>
<td>22.122°N, 114.042°E</td>
<td>Hainan University</td>
<td>3</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Phyllochaetopterus hainanensis</td>
<td>March 03, 2022</td>
<td>Changshacun, Dandou, Hainan</td>
<td>19.899°N, 109.279°E</td>
<td>Hainan University</td>
<td>&gt;20</td>
<td>Ethanol and formalin</td>
</tr>
<tr>
<td>Spiochaetopterus sp. A⁴</td>
<td>June, 2015</td>
<td>Lantau Island, Hong Kong</td>
<td>22.274°N, 114.085°E</td>
<td>Hainan University</td>
<td>2</td>
<td>Ethanol</td>
</tr>
<tr>
<td>Spiochaetopterus sp. B⁴</td>
<td>December 26, 2023</td>
<td>Beigang Island, Haikou, Hainan</td>
<td>20.019°N, 110.569°E</td>
<td>Hainan University</td>
<td>4</td>
<td>Ethanol and formalin</td>
</tr>
</tbody>
</table>

*The specimens will be deposited at the Institute of Oceanology, Chinese Academy of Science (IOCAS), Qingdao, China.

*The species with incomplete names will be described as new species in another paper.
The methods for sequence alignment, concatenations, removal of poorly aligned locations, model selection and phylogenetic tree construction were the same as those mentioned above.

To assess the phylogenetic placement of Chaetopteridae in Annelida, we used all protein-coding amino acid datasets of 240 species in Struck et al. (2023) and the seven here sequenced, with *Terebratulina retusa* (Linnaeus, 1758) as outgroup (Supplementary Table 2). The methods for model selection and ML phylogenetic tree construction were the same as the above, while BI was executed with two chains for 5,000,000 generations using Mpi-MrBayes v3.2 (Ronquist et al., 2012). In ML, a bootstrap >90 was considered strong clade support, 70–90 as moderate, and < 70 as weak support (Krenz et al., 2005). For BI, posterior probabilities > 0.95 were considered strong support (Jacobsen et al., 2010).

2.5 Mitochondrial gene order rearrangement

CREx2 was used to assess the mitogenomic rearrangement (Bernt et al., 2007), and the TreeRex (Bernt and Middendorf, 2011) allowing deducing the ancestral gene order of the inner nodes for Chaetopteridae and predicting the evolution process of present species based on the given phylogenetic tree. The parameters were set following the software recommendations: -s (strong consistency method), -w (weak consistency method), and -W (parsimonious weak consistency method).

2.6 Genetic distance and base substitution rates of mitochondrial genes

The pairwise genetic distance analyses of amino acid sequences of each mitochondrial PCG were conducted using MEGA X (Kumar et al., 2018). Base-substitution rates for the 13 PCGs of all chaetopterid mtgenomes were calculated followed by Sun et al. (2021). In short, the sequences for each gene were aligned using default parameters of the Muscle module within MEGA X. The non-synonymous to synonymous rate ratio (Ka/Ks) was then calculated using the YN method implemented in KaKs Calculator 2.0, with Ka/Ks indicating the strength of the selective pressure as > 1 positive selection, = 1 neutral evolution and < 1, purifying selection (Yang and Nielsen, 2000). *Spiochaetopterus* sp. A was excluded from this analysis due to its incomplete mtgenome.

3 Results

3.1 Chaetopterid mtgenomes

Sequence assembly resulted in six complete or nearly complete mtgenomes ranging from 15,665 bp to 21,822 bp (Table 1) except for *Spiochaetopterus* sp. A which resulted in a mitochondrial contig with 3,388 bp with only three potential PCGs (*cox3, nad2, nad3*) and 11 tRNAs (Figure 2; Supplementary Table 3). The start gene protein-coding codon for the species of Chaetopteridae was ATG, ATA and ATT, while the stop codon was either TAA, TAG, or defective TA or T (Supplementary Table 4).

![Comparison of gene orders of chaetopterid mtgenomes. Conserved gene clusters of annelids are marked by different color bars. White circles represent non-coding regions with >100 bp sequences between genes. White blocks represent the location of control region in mtgenome. Triangles stand for a long non-coding region within a gene. Missing genes are indicated by black blocks.](image-url)
All chaetopterid mtgenomes exhibit a high base bias and are AT-rich (58.2-70.17%), with only P. hainanensis (59.45%) and Phyllochaetopterus sp. (58.12%) having lower values but also close to 60% A+T (Table 2; Supplementary Table 5). The 13 PCGs (63.65%) and the third codon (63.45%) were also AT-rich. The full mtgenome and the PCGs showed negative AT-skew (-0.213–0.018, -0.296–0.06) and negative GC-skew (-0.318–0.172, -0.465–0.121) in all chaetopterid species except for the full mtgenome of P. hainanensis whose AT-skew (0.018) was greater than GC-skew (-0.314) (Supplementary Table 5). The tRNAs and rRNAs GC-skew are also negative, while their AT-skew were positive in all species except Phyllochaetopterus sp. (Supplementary Table 5).

3.2 Mtgenomes gene order

Three conserved gene blocks are present in all chaetopterids except Phyllochaetopterus sp. B: rrnS-trnV-rrnL-trnI1-nad3-trnS1-nad2, trnI-trnG-atp6-trnQ-trnW, and cob-trnS2-nad4L-nad4-trnH-trnD, with those in Chaetopterus and Mesochaetopterus being identical except that the location of trnY and trnF (Figure 2). >100 bp non-coding regions between different genes occur in different mtgenome locations but show differences between genera (Figure 2; Supplementary Table 6): only 0–1 in Chaetopterus, 2–4 in Mesochaetopterus and Phyllochaetopterus, two in Spirochaetopterus sp. B and five in the incomplete mtgenome of Spirochaetopterus sp. A. In addition, the non-coding regions interrupted several PCGs: cox1 in most Mesochaetopterus and Chaetopterus spp., Phyllochaetopterus sp.; nad6 and nad2 in M. tingkokensis and M. japonicus.

3.3 Phylogenetic analysis based on cox1, 18S and 28S

Phylogenetic analysis with 48 ingroups was conducted based on concatenated fragments of cox1, 18S, and 28S with a total length of 3,179 bp, including 649 bp for cox1, 1,653 bp for 18S, and 877 bp for 28S (Supplementary Table 7). Chaetopteridae appears as a well-supported clade with ML/BI = 100/1 (Figure 3). Chaetopterus and Mesochaetopterus are both monophyletic sister groups, while Phyllochaetopterus and Spirochaetopterus form a well-supported clade (ML/BI: 100/1) with two moderately-supported sub-clades (ML/BI: 83/0.95, ML/BI: 85/#) including species from two genera so that both are paraphyletic.

3.4 Phylogeny tree based on all mtgenome genes

The phylogenetic analyses were based on a data matrix containing 16,939 characters, including 9,900 bp from 13 PCGs, 2,687 bp from two rRNA genes and 1,630 bp from 22 tRNA genes (Figure 4; Supplementary Table 7). Five out of 13 PCGs showed significant saturation but were included because they contained phylogenetic information (Supplementary Table 8). The best models selected for the phylogenetic analyses were shown in Supplementary Table 9 and they produced similar tree topologies to those based on the three-gene dataset (Figure 4). All species of Chaetopteridae form a well-supported clade (ML/BI: 100/1) with two sub-clades, one with three species of Chaetopterus and two of Mesochaetopterus, both well supported (ML/BI: 100/1) consistently with the three-gene dataset results, in which Chaetopterus and Mesochaetopterus are also monophyletic. Another well-supported sub-clade consisted of species from Phyllochaetopterus and Spirochaetopterus, but this showed the three species of Phyllochaetopterus as paraphyletic with species of Spirochaetopterus and Phyllochaetopterus from the same location clustered together.

3.5 Mitochondrial gene order rearrangements

There were seven gene order patterns in the chaetopterid mtgenomes, including three in Chaetopterus and Mesochaetopterus, three in Phyllochaetopterus and one in Spirochaetopterus (Figure 5). The gene order distance ranges from 1 to 7 (Supplementary Table 10). Those between Chaetopterus and Mesochaetopterus are similar (1 to 2) but they are substantially larger than with Phyllochaetopterus (4 to 7). The gene order evolutionary patterns were predicted by TreeRex based on the phylogenetic tree (Figure 5; Supplementary Table 11). Reversions were detected in the most recent common ancestor of two Phyllochaetopterus sp. (OR637234 and KT726961.1) (tronY and trnM), when compared with its most recent common ancestor with Chaetopterus. Two transpositions were detected in the most recent common ancestor of two Phyllochaetopterus sp. (OR637234 and KT726961.1) (tronN and trnK), and P. hainanensis and Spirochaetopterus sp. B (trnD and trnN), compared with its most recent hypothetical common ancestral gene order. One reversion (tronA-trnP and trnR-trnD) was detected in Phyllochaetopterus sp. (KT726961.1) and two transpositions (tronC and trnM-trnS) and one reversion (tronD and trnA) in Spirochaetopterus sp. B, compared with that of its most recent common ancestor.

3.6 Genetic divergence of chaetopterid mtgenome PCGs

Pairwise genetic distance analysis was performed using PCG amino acids. Cob, cox1, cox2 and cox3 exhibit low genetic divergence, with the lowest value in cox1 (0.014–0.283) and the highest in atp8 (0.255–0.804), followed by nad2 (0.363–0.712) (Figure 6). Compared to Chaetopterus and Mesochaetopterus, Phyllochaetopterus has a higher genetic divergence in several PCGs, including atp8, cob, cox1 and nad2.
<table>
<thead>
<tr>
<th>Species</th>
<th>GenBank accession number</th>
<th>Contig length</th>
<th>AT content (%)</th>
<th>GC content (%)</th>
<th>AT skew</th>
<th>GC skew</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Whole genome</td>
<td>Protein coding genes</td>
<td>Whole genome</td>
<td>Protein coding genes</td>
<td>Full genome</td>
</tr>
<tr>
<td>Chaetopterus qiiani</td>
<td>OR637230</td>
<td>15,903</td>
<td>67.67</td>
<td>66.79</td>
<td>32.33</td>
<td>33.21</td>
<td>-0.130</td>
</tr>
<tr>
<td>Chaetopterus sp.</td>
<td>OR637233</td>
<td>15,665</td>
<td>65.68</td>
<td>64.99</td>
<td>34.32</td>
<td>35.01</td>
<td>-0.140</td>
</tr>
<tr>
<td>Chaetopterus varioapedatus</td>
<td>NC_028710.1</td>
<td>16,142</td>
<td>64.35</td>
<td>62.47</td>
<td>35.64</td>
<td>37.53</td>
<td>-0.118</td>
</tr>
<tr>
<td>Mesochaetopterus tingkokensis</td>
<td>OR637231</td>
<td>19,522</td>
<td>70.13</td>
<td>68.52</td>
<td>29.87</td>
<td>31.48</td>
<td>-0.019</td>
</tr>
<tr>
<td>Mesochaetopterus japonicus</td>
<td>MZ921947.1</td>
<td>19,326</td>
<td>70.17</td>
<td>67.84</td>
<td>29.83</td>
<td>32.16</td>
<td>-0.021</td>
</tr>
<tr>
<td>Phyllochaetopterus sp.</td>
<td>OR637234</td>
<td>18,463</td>
<td>69.50</td>
<td>69.03</td>
<td>30.50</td>
<td>30.97</td>
<td>-0.213</td>
</tr>
<tr>
<td>Phyllochaetopterus hainanensis</td>
<td>OR637235</td>
<td>21,822</td>
<td>59.45</td>
<td>59.46</td>
<td>40.55</td>
<td>40.54</td>
<td>0.018</td>
</tr>
<tr>
<td>Phyllochaetopterus sp.</td>
<td>KT726961.1</td>
<td>16,087</td>
<td>58.12</td>
<td>56.81</td>
<td>41.88</td>
<td>43.19</td>
<td>-0.053</td>
</tr>
<tr>
<td>Spiochaetopterus sp. A</td>
<td>OR637232</td>
<td>3,388</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Spiochaetopterus sp. B</td>
<td>PP440187</td>
<td>15,510</td>
<td>68.11</td>
<td>67.34</td>
<td>31.89</td>
<td>32.66</td>
<td>-0.196</td>
</tr>
</tbody>
</table>

$^*$ '$—' indicates 'Not applicable' because of incomplete mitochondrial genome.
3.7 Purifying selection of 13 mitochondrial PCGs

The Ka/Ks of the 13 chaetopterid mitochondrial PCGs are all lower than 0.5, suggesting that they have undergone strong purifcation selection. Nevertheless, the Ka/Ks of \(\text{atp8} (0.378)\) and \(\text{nad4} (0.423)\) are higher than those of the other PCGs (Figure 7A), with the smallest in the complex IV (i.e. \(\text{cox1} = 0.018, \text{cox2} = 0.051, \text{cox3} = 0.035\) and \(\text{nad5} (0.070)\) (Figure 7A). Compared with Chaetopterus and Mesochaetopterus, Ka/Ks was higher in Phyllochaetopterus for all mitochondrial PCGs except \(\text{nad4}\) (Figure 7B), indicating a general relaxation of purifying selection in the mtgenomes of this genus.
FIGURE 5
Hypothetical ancestral gene order of mtgenomes for Chaetopteridae and gene order rearrangement scenarios. R= Reversion and T= transposition, with the related genes being marked by lines and triangles, respectively.

FIGURE 6
Pairwise genetic distances of mitochondrial protein-coding genes (amino acids) among species of Chaetopteridae, cha qia, Chaetopterus qian; cha sp HK, Chaetopterus sp.; cha var, Chaetopterus variopedatus; phy hai, Phyllochaetopterus hainanensis; phy sp HK, Phyllochaetopterus sp.; phy sp, Phyllochaetopterus sp.; mes tin, Mesochaetopterus tingkokensis; mes jap, Mesochaetopterus japonicus; spi sp, Spiochaetopterus sp. B; eur com, Eurythoe complanata.
### 3.8 Phylogenetic position of Chaetopteridae in the annelid tree of life

The ML/BI phylogenetic tree was constructed using the concatenated datasets of the amino acid sequences of 13 PCGs of mitochondrial genome, with 247 annelid terminals (Supplementary Table 2B, Supplementary Figure 1). Our results show the Oweniidae and Magelonidae as sister groups, forming a sister clade to Chaetopteridae + all other annelid families (ML/BI: 95%).

### 4 Discussion

#### 4.1 Structural features of the chaetopterid mtgenomes

The mitochondrial intergenic gaps of metazoans are usually composed of very short non-coding regions or very few overlapping bases (Zhong et al., 2008). Therefore, the length of metazoan mtgenome is generally stable, especially in vertebrates. Conversely, annelid mtgenomes vary substantially from 14kb (Seixas et al., 2017; Sun et al., 2021), although chaetopterid mtgenomes show a smaller range from 15kb (Chaeotopterus) to 21kb (Phyllochaetopterus). Longer annelid mtgenomes obey either an enormous control region (D-loop) (as in Spirorbranchus and Sibogilinus) or non-coding region within the cox1 gene (the group II intron) and many huge intergenic gaps (as in Glycera, Hydrodoides and Polynoidae). Longer mtgenomes generally contain more long non-coding regions and a non-conserved gene order, as in Hydrodoides spp. (Sun et al., 2021), Siboglinidae (Siboglinum fiordicum Webb, 1963, 19,502 bp; Li et al., 2015). However, there are some exceptions, such as among Glyceridae (Glycera unicornis Lamarck, 1818, 20,366 bp; Richter et al., 2015). In Phyllochaetopterus and Mesoachaetopterus, large mtgenomes were either due to the group II introns among the cox1, to the non-coding region within the nad2, or numerous huge intergenic non-coding regions. Long non-coding regions were produced probably during changes in gene structure in Chaetopteridae, since intergenic non-coding spaces have been suggested to facilitate the inception of replication (Pons et al., 2014), or as remnants of gene order change (such as TDRL and transpositions) (Seixas et al., 2017), strands of the template mispairing, or imprecise termination during replication (Bernt et al., 2013b; Aguado et al., 2016).

The mtgenome gene orders of Chaetopteridae greatly differ from those of the ancestral annelid and most other annelids, although it is relatively conserved and shows three conserved gene blocks in most species (Figure 2). Among the ten mtgenomes, those of Chaetopterus and Mesoachaetopterus show very similar gene orders, with only one reversion in the position of trnY and trnM or trnE and trnF. The Phyllochaetopterus and Spirochaetopterus mtgenomes are also very similar, with several unstable positions of rRNA (i.e. trnB, trnD, trnP, trnN, trnK, trnC) and a fragment of trnM-rmS. These results are consistent with the close phylogenetic relationships between Chaetopterus and Mesoachaetopterus, and between Spirochaetopterus and Phyllochaetopterus, respectively.

#### 4.2 Phylogenetic relationships and pairwise genetic distance analysis

Our study shows the Chaetopteridae as a sister family to most annelid families, being recovered as sisters to the Oweniidae/Magelonidae clade. Transcriptome and mtgenome studies have all recovered Chaetopteridae as one of the basal annelid families, although its relationships with other basal families are unstable (Roussel et al., 2007; Weigert et al., 2016; Struck et al., 2023). Our ML result (i.e. (Oweniidae + Magelonidae) + (Chaetopteridae + other annelid families)) agrees with the result in Struck et al. (2023), while the BI analyses did not favor these relationships (Struck et al., 2023). Our Chaetopteridae clade was well supported, in agreement with the previous results based on universal biomarker genes (cox1, 18S, 28S) (Martin et al., 2008; Osborn et al., 2007; Morineaux et al., 2010; Weigert et al., 2014; Andre et al., 2015; Zhang et al., 2015; Britayev et al., 2017; Moore et al., 2017; Martin et al., 2022; Yang et al., 2022), while differing from morphological studies based on larvae and adults, which classified Chaetopteridae as a family within Sabellida (Rouse and Fauchald, 1997; Brown et al., 1999; Rouse, 2009).
formed the clade of Chaetopteriformia based on phylogenetic analysis using transcriptomes (Helm et al., 2018), however, only Chaetopteridae was included in the later report on mtgenome evolution in Annelida (Struck et al., 2023). The mtgenomes of Apistobranchidae and Psammodrilidae are needed to further explore the ancestral mtgene order of Chaetopteridae considering their close evolutionary relationship with Chaetopteridae.

The monophyly of Chaetopterus and Mesochaetopterus was not supported by morphology and BI based on 28S (Osborn et al., 2007; Morineaux et al., 2010). In this study, Chaetopterus appears as a well-supported sister taxon to Mesochaetopterus, in agreement with Moore et al. (2017); Martin et al. (2022) and Zhang et al. (2015). The paraphyletic relationship of Phyllochaetopterus + Spiochaetopterus is consistent with the results of the three scholars above. Major revisions are needed to define more distinguishing characteristics of these two genera instead of the present/absent of notopodial cirrus of A1.

Among the PCGs, atp8 has the highest pairwise genetic distance (Figure 6), followed by nad2 and nad6, suggesting faster substitution rates compared to other PCGs, as previously shown in deep-sea Polynoidae (Zhang et al., 2018), in which atp8 is the fastest-evolving mitochondrial gene (Sun et al., 2021). Additionally, the PCG genetic distance in Phyllochaetopterus was generally greater than that of the other genera, suggesting different selective pressures and faster mutation rates in its mtgenomes. In addition, the branch length of the Phyllochaetopterus + Spiochaetopterus clade was longer than that of the Chaetopterus + Mesochaetopterus clade (Figures 3, 4), indicating a faster substitution rate in the former.

### 4.3 Mtgenome rearrangement

Chaetopteridae exhibits distinct mtgenome gene order arrangement patterns compared to other annelids. Gene orders were initially considered as conserved among annelid mtgenomes (Jennings and Halanych, 2005), which is still true for many families (Weigert et al., 2016). However, substantial gene order arrangements have been found among Syllidae (Aguado et al., 2015, 2016), deep-sea Polynoidae (Zhang et al., 2018) and Serpulidae (Sun et al., 2021). Chaetopteridae shows different mtgenomes than the ancestral annelids (excluding Oweniidae and Magelonidae), mainly in the atp6, nad1, nad6, cox3, cox1, and several tRNAs positions (Supplementary Figure 2). However, its gene order is relatively conserved, with three conserved regions in all species except Spiochaetopterus (Figure 2). The uncertain evolutionary relationships between Chaetopteridae and all other polychaetes, the gene order of its most recent ancestor cannot be determined. Therefore, it is not clear how the mitochondrial gene order of the common ancestor of Chaetopterus and Mesochaetopterus or Phyllochaetopterus and Spiochaetopterus evolved from that of the chaetopterid common ancestor.

### 4.4 Base-substitution rates of mtgenomes

The diversity of gene orders is positively related to the substitution rate (Bernt et al., 2013a, b; Luo et al., 2015). The conserved functions in the respiration of mitochondrial genes explain why they are undergoing purifying selection (Stewart et al., 2008). Our study shows the Ka/Ks for all chaetopterid protein-coding genes as being lower than 1, indicating purifying selection (Figure 7), while being quite fast evolving compared to other polychaetes (Zhang et al., 2018) or invertebrates (Ren et al., 2010). Phyllochaetopterus shows higher base-substitute rates for most mitochondrial PCGs than Chaetopterus or Mesochaetopterus, indicating faster mutation rate that is consistent with their longer branch length in phylogenetic trees. Different mutation rates have also been shown among Polynoidae, with the deep-sea genera Branchipolytone and Branchinotogluma, having high mutation rates that suggest adaptation values (Zhang et al., 2018).

Overall, our results show different mitochondrial gene orders in Chaetopteridae compared to the conserved Annelida pattern, despite most species in this family sharing three singular conserved regions. Chaetopterus and Mesochaetopterus are sister groups forming well-supported monophyletic clades, which, together, are sister to a paraphyletic Phyllochaetopterus clade with Spiochaetopterus as subclade. The mitochondrial PCG genes in all examined chaetopterids have undergone purifying selection, but in most cases, those of Phyllochaetopterus show a higher base-substitution rate than those of the other genera.

### Data availability statement

Sequence data of the mitochondrial genomes first reported in this study are available in the NCBI GenBank OR637230 - OR63235, PP440187.

### Ethics statement

The manuscript presents research on animals that do not require ethical approval for their study.

### Author contributions

XW: Writing – review & editing. CH: Writing – original draft, Visualization, Formal analysis, Data curation. XS: Writing – original draft, Visualization, Methodology. J-WQ: Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

### Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was financially supported by the National Natural Science Foundation of China (42106118), Hainan Province Science and
Acknowledgments

The authors would like to thank Prof. Jin Sun from Ocean University of China, and Deyuan Yang from Xiamen University for their guidance on sequence processing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References


Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2024.1382212/full#supplementary-material


