Molecular Genetics and Genomics (2019) 294:715–727
https://doi.org/10.1007/s00438-019-01543-1

ORIGINAL ARTICLE

Comparative analysis of the mitochondrial genomes of Orthonectida: insights into the evolution of an invertebrate parasite species

N. Bondarenko[1] · A. Bondarenko[2] · V. Starunov[1] · G. Slyusarev[1]

Received: 14 November 2018 / Accepted: 26 February 2019 / Published online: 8 March 2019

Abstract

Among invertebrates, only a few groups still have uncertain phylogenetic position, Orthonectida, a small group of rare multi-cellular parasites of marine invertebrates, being one of them. Recent molecular and morphological findings suggest that orthonectids belong to Lophotrochozoa and are close to Annelida. Nevertheless, phylogenetic relationships between orthonectids and annelids are unclear, and the phylogeny within the group itself has never been studied. Sequencing of mitochondrial genomes is used here to clarify this issue. Complete mt genomes of the orthonectids *Intoshia variabili* and *Rhopalura litoralis* were characterized and compared with *Intoshia linei* mt genome. Our results show that Orthonectida mt genomes have undergone reduction and gene loss, and that they have complicated organization revealed in strand asymmetry in nucleotide composition, in some features of intergenic non-coding regions, tRNA duplication and folding. Moreover, all species of Orthonectida have a unique gene order with complicated rearrangement landscape. Significant differences in mitochondrial genomes in the three orthonectid species could be explained by the fact that their host species belong to different taxa (flat worms, nemertines and gastropods). Among the analyzed mt genomes of Orthonectida, *I. linei* possesses the closest gene order to the ancestral genome. All Orthonectida species are monophyletic, and in the phylogenetic tree are close to Pleistoannelida, and specifically, to Clitellata.

Keywords Orthonectida · Mitochondrial genome · Gene order · Mitochondrial DNA · Mitochondrion genome rearrangements

Abbreviations

cox1, cox2, cox3	Cytochrome oxidase subunit I, II, and III protein genes
cob	Cytochrome *b* gene
atp6, atp8	ATP synthase subunit 6 and 8 gene
nad2, nad4, nad4L, nad5, nad6, nad11	NADH dehydrogenase subunit 1–6, 11 and 4L genes
$tRNA^{Trp}$, $tRNA^{Ile}$, RNA^{Lys}, $tRNA^{Met}$, $tRNA^{Val}$, $tRNA^{Cys}$, $tRNA^{Met}$, $tRNA^{Gln}$, $tRNA^{Leu1}$, $tRNA^{Leu2}$, $tRNA^{Phe}$, $tRNA^{Arg}$, $tRNA^{Tyr}$, $tRNA^{Ser1}$, $tRNA^{Ser2}$, $tRNA^{Asp1}$, $tRNA^{Asp2}$, $tRNA^{Asn}$	Transfer RNA genes
rrnL, rrnS	Ribosomal RNA genes
ORF	Open reading frames
PCG	Protein-coding gene

Introduction

Orthonectida is probably one of the most enigmatic groups of so-called lower Metazoa, whose phylogenetic relationships remain uncertain (Slusarev 2018). Representatives of this group are parasites of a wide range of marine invertebrates. Their life cycle is rather simple and consists of two alternating generations: a free-living sexual stage and a parasitic stage commonly referred to as plasmodium. The

Communicated by B. Franz Lang.

✉ N. Bondarenko
n.bondarenko@spbu.ru

[1] Department of Invertebrate Zoology, Faculty of Biology, Saint-Petersburg State University, Universitetskaja nab. 7/9, 199034 St. Petersburg, Russia

[2] Faculty of Physics, Saint-Petersburg State University, Universitetskaja nab. 7/9, 199034 St. Petersburg, Russia

phylum is not divided into classes or orders and consists of about 25 known species (Kozloff 1992; Slusarev 2018). The systematics of the orthonectids is still unresolved (Slusarev 2018). Recent molecular and morphological findings suggest that they belong to Lophotrochozoa and are close to the Annelida (Schiffer et al. 2018; Slusarev 2018). Only one mitochondrial genome from Orthonectida is available to date (Schiffer et al. 2018). There is a lack of information concerning the mt genomes for other species of Orthonectida which is needed to derive insights into their taxonomy, phylogeny, and possible molecular markers. Nothing is known about the rearrangement events within Orthonectida mt genomes and how and when these occurred. More detailed knowledge about these processes could help to improve the understanding of the mechanisms of Orthonectida species evolution.

Due to the abundance of mitochondrial genomes in tissues, their faster rate of evolution compared with the nuclear genome and the availability of a large number of informative data sets, they are widely used for phylogenetic analyses across broad taxonomic levels (Rota-Stabelli et al. 2010; Bernt et al. 2013). During evolution, mitochondrial genomes are modified by large-scale structural events, such as rearrangements, deletions or insertions of DNA blocks. Differences in gene arrangements and genetic codes may be used as valuable phylogenetic markers (Bernt et al. 2013). Also, gene loss and duplications, strand asymmetry in nucleotide composition, length, and structure of the control region, features of intergenic non-coding regions, codon usage, variation in gene length, variation in start and stop codons, gene diversity levels, mutation rates, may be of interest to comparative mitogenomics (Curole and Kocher 1999; Gissi et al. 2008; Plazzi et al. 2016).

Here we present the mitochondrial genomes of two orthonectid species *Intoshia variabili* and *Rhopalura litoralis*. These genomes were sequenced, assembled, and annotated as circular DNA molecules. We compared the *I. variabili* and *R. litoralis* mt genomes with the *I. linei* mt genome and reconstructed ancestral gene order for the better understanding of the evolution of these genomes. Additionally, we performed a phylogenomic investigation to assess relationships of Orthonectida with Annelida.

Materials and methods

Material collection and DNA extraction

The orthonectid *I. variabili* (Alexandrov and Sljusarev 1992) occurs in the turbellarian *Macrorhynchus crocea* Graff, 1882 (Platyhelminthes: Rhabditophora, suborder Calyptorhynchia). The orthonectid *R. litoralis* (Shtein 1953) is a parasite of the gastropod *Onoba aculeus* (Gastropoda: order Littorinimorpha, superfamily Rissooidea). The turbellarians and gastropods were collected in August 2017, in the Barents Sea at the marine biological station Dalnie Zelentsi (69°07′N, 36°05′E). The hosts were collected at the low tide and maintained in Petri dishes with filtered seawater. Free-living stages of *I. variabili* and *R. litoralis* were collected as described in detail elsewhere (Slyusarev 1994; Slyusarev and Ferraguti 2002). For genomic sequencing, the collected samples of *I. variabili* and *R. litoralis* were fixed using ethanol. Total genomic DNA was isolated using the PicoPure DNA Extraction Kit (Thermo Fisher) according to the recommendations of the manufacturer. DNA quantity assessments were performed with Qubit fluorometric quantification (Life Technologies).

Mitochondrial genome sequencing, assembly, and annotation

The genomic DNA libraries of *I. variabili* and *R. litoralis* were constructed using TruSeq library preparation protocol (Illumina) and sequenced on a HiSeq 2000 sequencing system. 62.8M 100 bp paired-end reads were obtained. Quality control check on raw sequence data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed November 2017). De novo assemblies were implemented using SPAdes assembler (Bankevich et al. 2012). The quality of assemblies was evaluated using QUAST (Gurevich et al. 2013). The mitochondrial scaffold was identified in the assembly using BLAST (Altschul et al. 1997). Initial annotation was performed using the MITOS web server under the code for invertebrate mitochondria (Bernt et al. 2013). However, the MITOS server did not correctly identify the start and stop codons, so all protein-coding genes were further adjusted and corrected manually using mt genome of *I. linei* (Schiffer et al. 2018). Putative tRNA and rRNA genes were identified using the MITOS web server. To check prediction and analyze duplication events of all tRNAs, the program tRNAcan-SE Search Server v.1.21 (Lowe and Eddy 1997) was applied. The physical map was generated by our script written in Python (script available into GitHub, https://github.com/nilannik). The mitochondrial genomes have been deposited in GenBank under the accession numbers MG893580 (*I. variabili*) and MG917727 (*R. litoralis*).

Genetic code and nucleotide composition

AT and GC skew were calculated using the formulae: AT skew = [A − T]/[A + T] and GC skew = [G − C]/[G + C], for the strand encoding the majority of the protein-coding genes (Perna and Kocher 1995). Genetic code and codon usage were analyzed by GenDecoder v1.6 web tool (Abascal et al. 2006).

Ancestral genome reconstruction and gene order analyses

To show similarities in gene clusters alignment of Orthonectida mt genomes was performed with the Mauve (Darling et al. 2004). This program allows visualizing major rearrangements and inverted regions. For ancestral mt genome reconstruction, GRIMM program was used (Tesler 2002). For pairwise comparisons of the mitochondrial gene order and determining the most parsimonious genome rearrangement scenario CREx program was used (Bernt et al. 2007). The analysis was performed by applying the common intervals parameter for distance measurement.

Phylogenetic analysis

The protein-coding gene sequences of *I. variabili* and *R. litoralis* were translated using the "Invertebrate mitochondrial" genetic code per each taxon using the software Translate tool (https://web.expasy.org/translate). *Intoshia linei* and 28 Annelida species were selected from the NCBI's organelles genome database and included in the phylogenetic analysis. The amino acid sequences were concatenated to the alignment using our own script written in Python. Multiple sequence alignment was implemented using muscle program integrated to SeaView (version 4; Gouy et al. 2010) with the default parameters for the protein-coding genes. Variable regions were recognized and excluded using the program Gblocks (Castresana 2000). Near 25% of the most varying sites were excluded from alignment. Concatenation of partitions for the combined data set was conducted with SeaView. Phylogenetic inference was performed by PhyloBayes (Lartillot and Philippe 2004). The PhyloBayes with two independent Monte Carlo Markov chains running for 20,000 cycles under the MtZoa model (Rota-Stabelli et al. 2009). The analysis was considered to have reached stationarity when the average standard deviation of split frequencies decreased to 0.01. Stationarity for each run was assessed by importing the parameter files into Tracer v. 1.5 (Rambaut and Drummond 2009). 36,002 trees were summed after removing first 1999 as burn-in. The resulting Bayesian tree was visualized in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/, last accessed March 2018).

Results and discussion

Genome organization and nucleotide composition

BLAST searches identified the mt genomes of *I. variabili* and *R. litoralis* as single contigs. The complete mt genome of *I. variabili* is 13.989 bp (Fig. 1) and the one of *R. litoralis* is 14.299 (Fig. 2).

These two genomes are smaller than *I. linei* mt genome which has 15.217 bp in length (Schiffer et al. 2018). The difference in length is explained by the gene content and the size of the non-coding regions (Fig. 3; Tables 1, 2).

Compared with some other metazoan groups, the size variation in Orthonectida mt genomes is relatively small (approx. 14 kb ± 1 kb) (Gissi et al. 2008). The size of the two orthonectid mt genomes seems to be slightly smaller than typical metazoan mt genomes which vary in size from 15 to 20 kbp in length (Bernt et al. 2013; Boore 1999).

The genomes of *I. variabili* and *R. litoralis* have low GC content (17.38% and 21.28%, respectively) (Table 2). A is the most common base (41.76% in *I. variabili* and 39.72% *R. litoralis*), while G is the least common (8.85% and 10.67%, respectively). Both mt genomes have a positive AT and GC skew (Table 2) which differs from *I. linei* characterized by a negative AT and positive GC skew (Schiffer et al. 2018). These results indicated a low degree of strand asymmetry of the base composition in the *I. variabili* and *R. litoralis* mt genomes. Positive AT skew can be explained by approximately equal frequencies of A and T bases within each strand of *I. variabili* and *R. litoralis* mt DNA (a slight prevalence of A over T is characteristic of both species) (Table 3).

Some genes partially overlapped in both species. Overlaps were detected in the genes located in major and minor strands. *I. variabili* mt genome has ten gene overlaps. The largest overlap, 28 bp long, is located between *trnL2* and *nad6*. In *R. litoralis* mt genome, genes have 13 overlaps and the largest overlap of 77 bp is also located between *nad6* and *trnL2* genes. Coding gene overlapping seems to be a common feature in Orthonectida since the *I. linei* showed this feature in its genome too (Schiffer et al. 2018). Several non-coding regions are dispersed throughout the whole mt genome in both species. The non-coding regions in the mt genomes of *I. variabili* and *R. litoralis* are 1645 bp and 486 bp in total, respectively. *R. litoralis* has two major non-coding regions of 227 and 239 bp located between *trnN* and *nad1* and between *trnK* and *nad5*, respectively. *I. variabili* mt genome, like that of *I. linei*, has the largest non-coding region with a size of 1662 bp located between the *nad5* and the t*rnG* genes.

Protein-coding genes

Typical metazoan mt genomes are considered to be relatively constant in gene content and order and to consist of a single circular DNA which codes 13 protein-coding genes (PCGs), the 2 subunits of the rRNA, and 22 tRNAs with 2 copies of serine and leucine (Bernt et al. 2013; Boore 1999). The two mt genomes contain the same gene sets as found in most metazoan mt genomes except some variations. They include two rRNA genes and 21 tRNA genes in both species, ten PCGs (*atp8, nad2* and *cox3* are missing) in *I.*

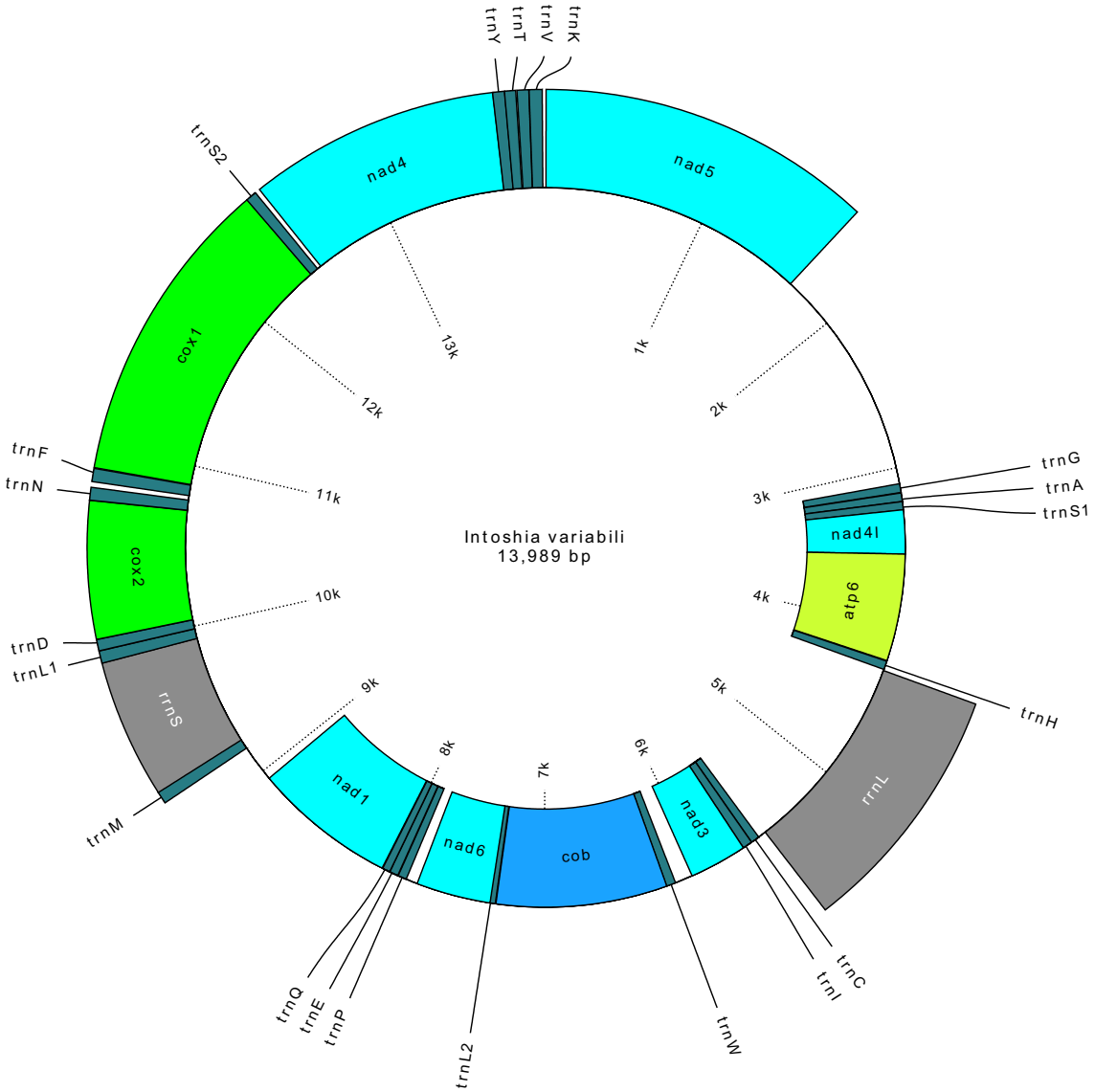

Fig. 1 *Intoshia variabili* mitochondrial genome map. The tRNA genes are labeled based on the IUPACIUB single letter amino acid codes

variabili (Fig. 1; Table 1), and 12 PCGs (*atp8* is missing) in *R. litoralis* (Fig. 2; Table 2). Loss of *atp8* is also reported from several other, phylogenetically distant taxa, such as Platyhelminthes, Chaetognatha, and Nematoda (Gissi et al. 2008). In both genomes, some of the genes are transcribed from the major strand (17 genes for *I. variabili*; 22 genes for *R. litoralis*), others from the minor strand (16 genes for *I. variabili*; 13 genes for *R. litoralis*). By comparison, *I. linei* has only eight genes located in the minor strand (Schiffer et al. 2018). The cumulative length of PCGs, excluding all termination codons, is 8841 bp encoding 2947 amino acid residues for ten genes of *I. variabili* and 10.578 bp encoding 3526 amino acid residues for 12 genes of *R. litoralis*.

Codon usage

Both Othonectida species use standard invertebrate mitochondrial code (GenBank translation Table 5) and do not use TGA as the stop codon (TGA specifies tryptophan). The majority of *I. variabili* and *R. litoralis* PCGs have ATA as the start codon (Table 1, 2). Only *nad1* and *nad4l* genes in *I. variabili* mt genome have ATG and ATT as start codons, respectively. As for *R. litoralis, nad1* also has ATG as a start codon, *nad6* and *nad4* genes begin with the start codon ATT. *I. variabili* and *R. litoralis* PCGs have two stop codons TAA and TAG (Table 1, 2), and TAA is preferred over TAG. Diverse start and stop codons have been also found in other Orthonectida (Schiffer et al. 2018).

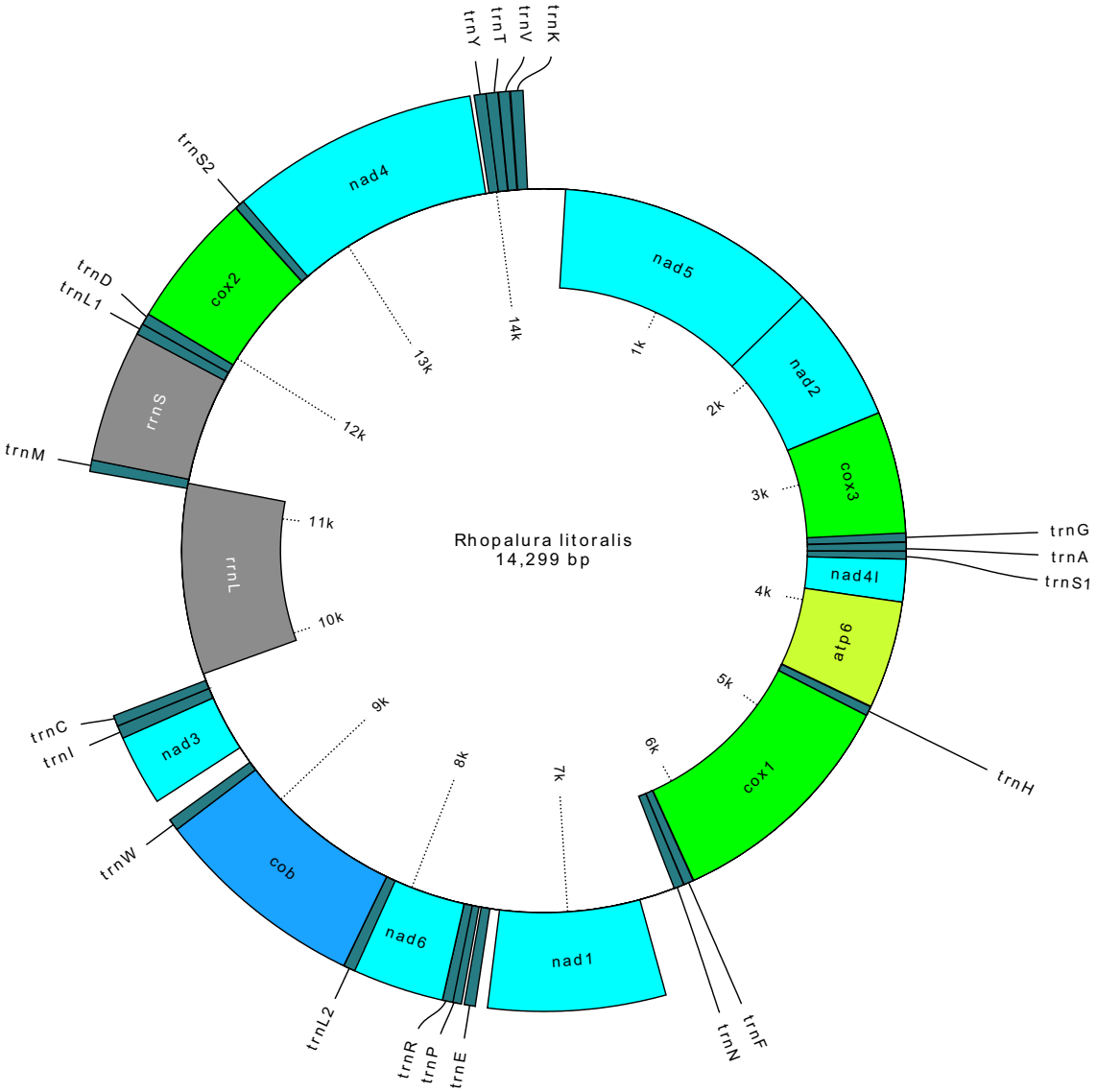

Fig. 2 *Rhopalura litoralis* mitochondrial genome map. The tRNA genes are labeled based on the IUPACIUB single letter amino acid codes

The mitochondrial PCGs of *I. variabili* lack six of the 64 possible codons: GCC for alanine, GCG, and CGG for arginine, CTC for leucine, AGC for serine, and TAG stop codon (Table 4). The PCGs of *R. litoralis* lack only two of the 64 possible codons: GCG for arginine and CCG for proline (Table 5). AT richness of both mt genomes affects the amino acid composition bias in PCGs towards the amino acids coded by the AT-rich codons (Tables 4, 5). There is also a codon usage bias in the two genomes (Table 4, 5).

In general, NNA and NNT codons are the most common codon types, while NNG and NNC codons are the least used. Thus, the eight most frequently used codons were TTT, TTA, ATT, ATA, TCT, TAT, AAT, and AAA in *I. variabili* (Table 4) and *R. litoralis* mt genomes (Table 5). These AT-rich codons account for 55.2% and 53.2% in *I. variabili* and *R. litoralis*, respectively. This is consistent with the high percentage of A + T content in the nucleotide composition of PCGs as in another mt genome for Orthonectida.

Transfer and ribosomal RNA genes

The majority of metazoans mt genomes typically encode 22 tRNAs with a single tRNA for each of 18 amino acids and 2 tRNA genes for serine and leucine. A total of 21 tRNA

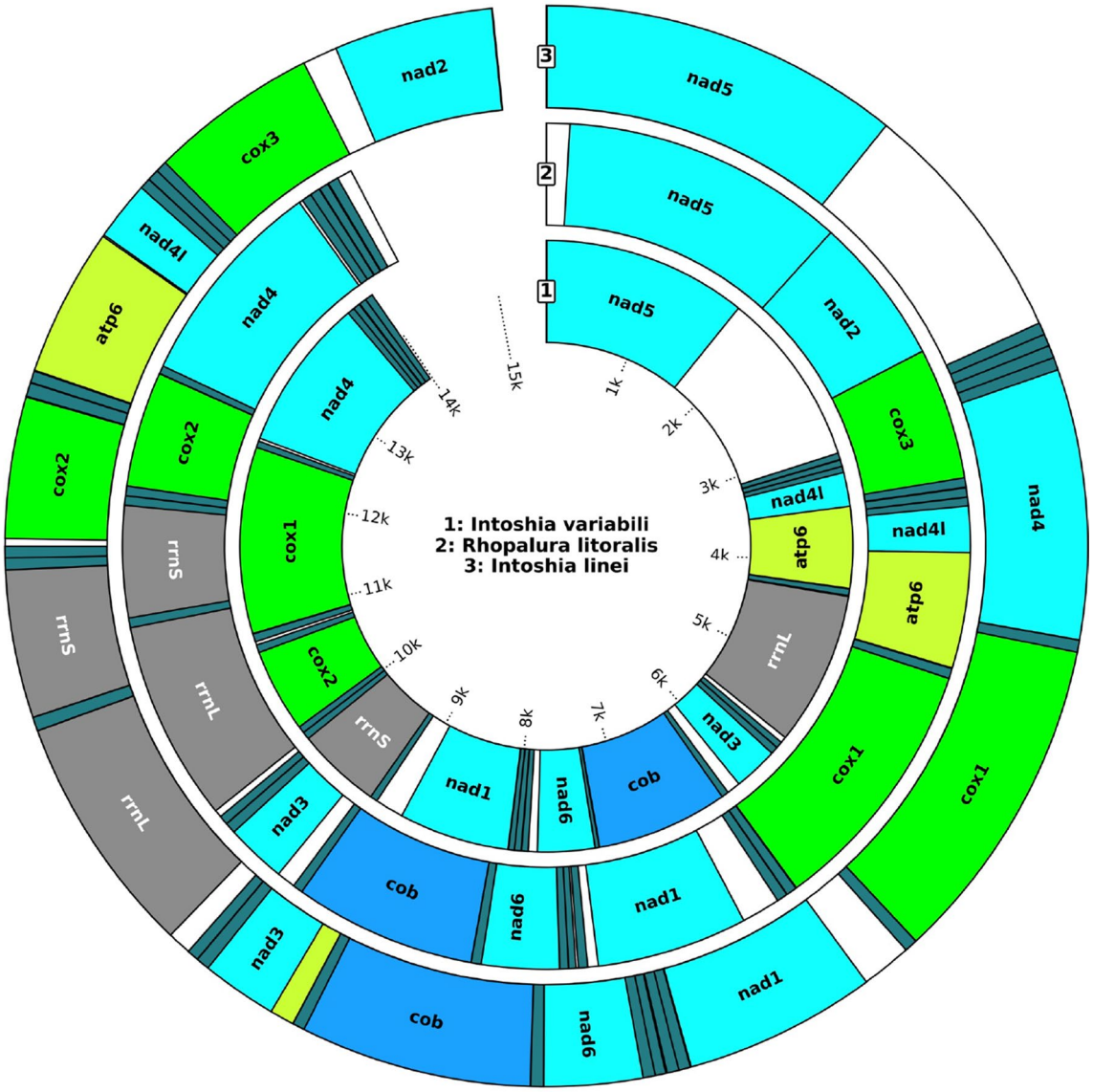

Fig. 3 Comparison map of the three mitochondrial genomes of Orthonectida

genes were found interspersed in the mt genomes of *I. variabili* and *R.* (Tables 1, 2). Besides two $tRNA^{Ser}$ and $tRNA^{Leu}$ genes, the mt genome of *I. variabili* encodes two aspartic acid tRNA genes ($tRNA^{Asp}$) (Table 1; Fig. 1). The $tRNA^{Asp1}$ gene is located after $tRNA^{Leu1}$, and the $tRNA^{Asp2}$ gene before *cox1* gene. A small number of tRNAs of both mt genomes possess the common cloverleaf structure. Eight tRNAs of *I. variabili* and 11 tRNAs of *R. litoralis* have the TΨC arm simplified to a loop. Another three tRNAs of *I. variabili* and *R. litoralis* have the dihydrouridine arm simplified to a loop.

The length of 16S rRNA gene of both *I. variabili* and *R. litoralis* mt genomes is in the range of other metazoan mt LSU rDNA (from 1 to 1.5 kb), though 12S rRNA is slightly smaller (660 and 690 bp, while it is between 700 and 1.5 kb for other metazoans) (Wey-Fabrizius et al. 2013).

Ancestral genome reconstruction and gene order analysis

So far, the order of genes and their rearrangements within Orthonectida remained unknown because only one mt genome belonging to *I. variabili* is available (Schiffer et al. 2018). To show similarities in gene clusters and initial view to rearrangements the three Orthonectida mt genomes were compared using Mauve. The mt genome of *I. linei* was chosen as a reference. Significant homology was observed for the six regions represented by colored blocks connected between genomes by lines (Fig. 4).

As shown in Fig. 4, several blocks are shifted downward relative to the reference genome, which indicates that such blocks are in the reverse complement orientation. Regions

Table 1 *Intoshia variabili* genome organization

Gene	Strain	Position (start–stop)	Length (bp)	Intergenic space (bp)	Start codon	Stop codon
tRNA^{Gly}	−	121–176	56	2		
tRNA^{Ala}	−	179–232	54	0		
tRNA^{Ser1}	−	233–285	53	−2		
nad4l	−	284–559	276	−1	ATT	TAA
atp6	−	559–1230	672	3	ATA	TAA
t tRNA^{His}	−		55	6		
rrnL	+	1295–2551	1257	49		
tRNA^{Cys}	−	2601–2655	55	0		
tRNA^{Ile}	−	2656–2717	62	−1		
nad3	−	2717–3079	363	10	ATA	TAA
tRNA^{Trp}	−	3190–3249	60	−2		
cob	−	3248–4321	1074	3	ATA	TAA
tRNA^{Leu1}	−	4325–4385	61	−28		
nad6	−	4358–4825	468	72	ATA	TAG
tRNA^{Pro}	−	4898–4952	55	2		
tRNA^{Glu}	−	4955–5009	55	0		
ttRNA^{Gln}	−	5010–5058	49	3		
nad1	−	5062–5961	900	32	ATG	TAA
tRNA^{Met}	+	6194–6259	66	−6		
rrnS	+	6254–6943	690	−2		
tRNA^{Leu2}	+	6942–7003	62	0		
tRNA^{Asp1}	+	7004–7060	57	0		
cox2	+	7061–7741	681	0	ATA	TAA
tRNA^{Asn}	+	7742–7805	64	27		
tRNA^{Asp2}	+	7833–7898	66	3		
cox1	+	7902–9431	1530	−2	ATA	TAA
tRNA^{Ser2}	+	9430–9484	55	24		
nad4	+	9509–10,753	1245	−2	ATA	TAA
tRNA^{Tyr}	+	10,752–10,811	60	−2		
tRNA^{Thr}	+	10,810–10,865	56	5		
tRNA^{Val}	+	10,871–10,929	59	0		
tRNA^{Lys}	+	10,930–10,994	65	21		
nad5	+	11,016–12,677	1662	1431	ATA	TAA

containing *cox2* and *nad3* gene, are not defined by Mauve program probably because the mt genome of *I. variabili* lacks such genes. Also, there are regions located outside the blocks, which are too divergent in genomes and contain lineage-specific sequences (Fig. 4). As shown in Fig. 4, the mt genome of *I. linei* can be transformed into that of *I. variabili* by a reverse transposition of regions number 2, 3 and 6 (genes located in regions are represented in Table 6).

As for *R. litoralis*, mt genome of *I. linei* can be transformed into that by a reverse transposition of regions number 1, 2 and 6. As seen from the analysis, the rearrangement landscape of Orthonectida mt genomes is rather complicated and it is unclear which of the gene order is more ancestral. For the better understanding of Orthonectida mt genome evolution, ancestral gene order was reconstructed (Fig. 5).

Ancestral gene order is more similar to *I. linei* mt genome (Fig. 5). To estimate the number of rearrangements between the ancestral and Orthonectida mt genomes the CREx analysis was implemented (Table 4). According to the analysis, *I. linei* mt genome had only two reversal events compared with the ancestral genome. (Table 7; Fig. 6a). In the course of evolution *I. variabili* lost *nad2, atp8, cox3, trnR*, and *trnF* genes, duplicated *trnD* gene, and had two reversals of two gene blocks (Fig. 6b). As for *R. litoralis*, this mt genome lost *atp8* and *trnQ* genes and had three reverse transposition and one reversal events (Fig. 6c).

Table 2 *Rhopalura litoralis* genome organization

Gene	Strain	Position (start–stop)	Length (bp)	Intergenic space (bp)	Start codon	Stop codon
nad5	−	137–1813	1677	−4	ATA	TAA
nad2	−	1810–2685	876	0	ATA	TAA
cox3	−	2686–3468	783	−6	ATA	TAA
tRNA^{Gly}	−	3463–3519	57	2		
tRNA^{Ala}	−	3522–3577	56	0		
tRNA^{Ser1}	−	3578–3629	52	−2		
nad4l	−	3628–3900	273	−1	ATA	TAA
atp6	−	3900–4586	687	3	ATA	TAA
tRNA^{His}	−	4590–4649	60	0		
cox1	−	4650–6179	1530	3	ATA	TAA
tRNA^{Phe}	−	6183–6245	63	1		
tRNA^{Asn}	−	6247–6311	65	227		
nad1	+	6539–7429	891	60	ATG	TAA
tRNA^{Glu}	+	7490–7545	56	13		
tRNA^{Pro}	+	7559–7609	51	−8		
tRNA^{Arg}	+	7602–7654	53	0		
nad6	+	7655–8188	534	−77	ATT	TAA
tRNA^{Leu2}	+	8112–8171	60	0		
cob	+	8172–9245	1074	−1	ATA	TAA
tRNA^{Trp}	+	9245–9304	60	10		
nad3	+	9415–9771	357	−1	ATA	TAG
tRNA^{Ile}	+	9771–9835	65	0		
tRNA^{Cys}	+	9836–9890	55	43		
rrnL	−	9934–11,151	1217	−34		
tRNA^{Met}	+	11,118–11,177	60	−2		
rrnS	+	11,176–11,840	665	−2	ATA	TAA
tRNA^{Leu1}	+	11,839–11,894	56	0		
tRNA^{Asp}	+	11,895–11,951	57	0		
cox2	+	11,952–12,632	681	1	ATT	TAG
tRNA^{Ser2}	+	12,634–12,687	54	−6		
nad4	+	12,682–13,932	1251	20	ATT	TAA
tRNA^{Tyr}	+	13,953–14,013	61	−2		
tRNA^{Thr}	+	14,012–14,071	60	2		
tRNA^{Val}	+	14,074–14,129	56	4		
tRNA^{Lys}	+	14,134–14,196	63	239		

Table 3 Nucleotide composition characteristics of *I. variabili* and *R. litoralis* mitochondrial genomes

Species	GC%	A%	T%	G%	C%	AT skew	GC skew
I. variabili	17.38	41.76	40.85	8.85	8.54	0.011	0.018
R. litoralis	21.28	39.72	38.98	10.67	10.63	0.009	0.001

The arrangement observed for *I. linei* can be obtained with a minimum number of changes only when starting from that of the ancestral genome (Table 7). *I. linei* mt genome has minimum rearrangement events and poses the closest gene order to the ancestral genome (Table 7). Mt genome of *I. variabili* has the largest number of rearrangement events when starting from that of the ancestral genome. Rearrangements are represented by two reversals of significant gene blocks, deletions of five genes (*nad2, atp8, cox3, trnR*, and *trnF)*, and one duplication event (Fig. 6). *R. litoralis*, compared with *I. variabili*, has lower number of rearrangements: three reverse transpositions of different gene blocks,

Table 4 The codon usage in *Intoshia variabili*

Amino acid	Codon	*Intoshia variabili*	Amino acid	Codon	*Intoshia variabili*
A	GCA	16	F	TTC	32
	GCC	–		TTT	281
	GCG	1	L	TTA	353
	GCT	28		TTG	22
R	CGA	26		CTA	33
	CGC	–		CTC	–
	CGG	–		CTG	3
	CGT	4		CTT	35
Y	TAT	95	I	ATC	19
	TAC	19		ATT	290
N	AAT	96	V	GTA	57
	AAC	16		GTC	3
D	GAT	28		GTG	12
	GAC	4		GTT	50
C	TGT	20	S	TCA	65
	TGC	3		TCC	9
E	GAA	54		TCG	4
	GAG	5		TCT	106
P	CCA	37		AGT	31
	CCC	4		AGC	–
	CCG	3		AGA	56
	CCT	34		AGG	4
T	ACA	61	H	CAT	29
	ACC	5		CAC	5
	ACG	2	K	AAA	97
	ACT	54		AAG	6
Q	CAA	30	M	ATG	21
	CAG	3		ATA	310
G	GGA	29	W	TGA	42
	GGC	2		TGG	1
	GGG	6	Stop codons	TAA	9
	GGT	30		TAG	–

Table 5 The codon usage in *Rhopalura litoralis*

Amino acid	Codon	*Rhopalura litoralis*	Amino acid	Codon	*Rhopalura litoralis*
A	GCA	25	F	TTC	62
	GCC	11		TTT	343
	GCG	–	L	TTA	362
	GCT	27		TTG	29
R	CGA	21		CTA	66
	CGC	1		CTC	30
	CGG	1		CTG	1
	CGT	7		CTT	71
Y	TAT	137	I	ATC	32
	TAC	39		ATT	348
N	AAT	168	V	GTA	65
	AAC	39		GTC	5
D	GAT	42		GTG	13
	GAC	5		GTT	54
C	TGT	23	S	TCA	52
	TGC	5		TCC	63
E	GAA	76		TCG	2
	GAG	12		TCT	120
P	CCA	40		AGT	29
	CCC	24		AGC	11
	CCG	–		AGA	63
	CCT	30		AGG	8
T	ACA	51	H	CAT	45
	ACC	25		CAC	13
	ACG	3	K	AAA	138
	ACT	88		AAG	13
Q	CAA	34	M	ATG	61
	CAG	2		ATA	261
G	GGA	47	W	TGA	38
	GGC	7		TGG	6
	GGG	14	Stop codons	TAA	10
	GGT	19		TAG	2

one reversal event and two gene deletions (*atp8* and *trnQ*) (Fig. 6).

Starting from any Orthonectida mt genome arrangement, the number of changes is approximately the same (Table 7). This is explained by the fact that it is not possible to calculate correctly all rearrangements, because programs do not consider duplications and loss of genes. In these cases, all rearrangements for all genomes are represented only by reversals or reverse translocations. All known species of Orthonectida have a unique gene order which was not defined previously (Fig. 7). If we compare Orthonectida to the putative ground pattern of Lophotrochozoa we observe three common gene blocks (*nad4* and *nad5, rrnL* and *rrnS, nad1, nad6* and *cob*) along with complicated rearrangement landscape within Orthonectida (Fig. 7a, c). If we compare Orthonectida with Pleistoannelida we can see only two common gene blocks (*rrnL* and *rrnS, nad6* and *cob*) (Fig. 7a, c). The exception is *I. variabili*, which has three common gene blocks with Lophotrochozoa and Pleistoannelida (Fig. 7b). Interestingly, ancestral genome has four common gene blocks with Lophotrochozoa and three with Pleistoannelida (Fig. 7d).

Phylogenetic analysis

According to previous studies, position of the Orthonectida on the phylogenetic tree has been changed (Mikhailov

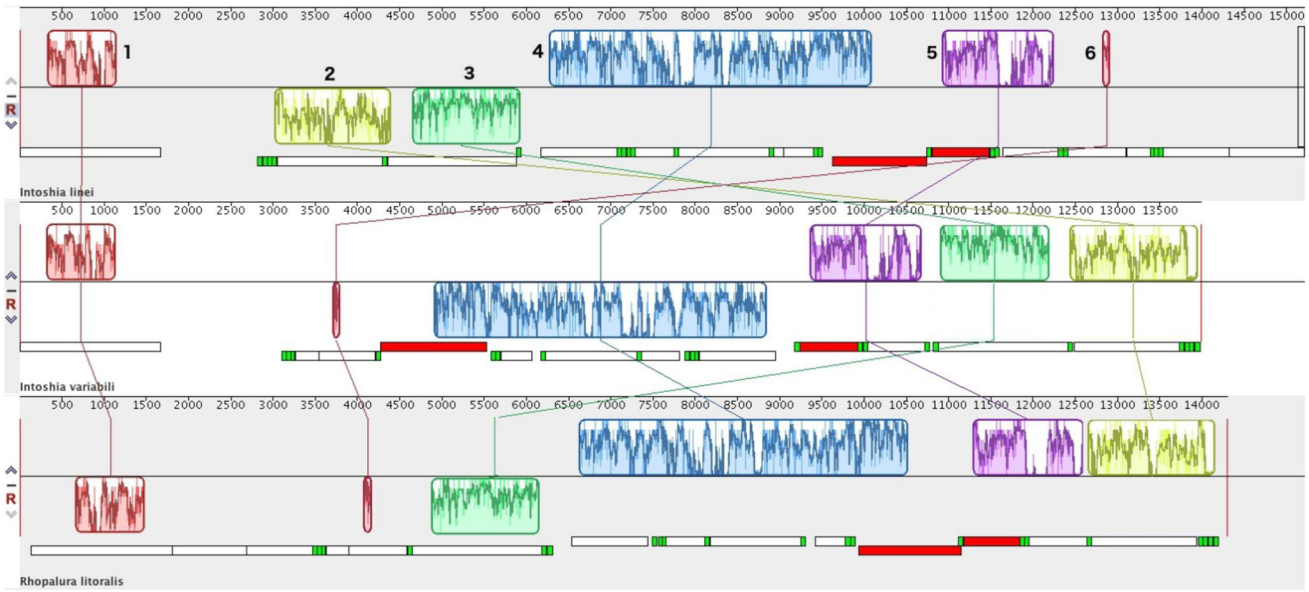

Fig. 4 A Mauve alignment of *Intoshia linei, Intoshia variabili*, and *Rhopalura litoralis* mitochondrial genomes. The colored blocks represent regions of homology between mt genomes as determined by Mauve alignment on default settings. Inside each block, Mauve draws a similarity profile of the genome sequence. Lines indicate which regions in each genome are homologous

Table 6 Genes located in the regions, determined by Mauve program

Number of regions	Genes
1	*nad5*
2	*trnK, trnV, trnT, trnY, nad4, trnS*
3	*cox1, trnF*
4	*nad1, trnQ, trnE, trnP, trnR, nad6, trnL, cob, trnW, atp8, nad3, trnI, trnC, rrnL*
5	*rrnS, trnL, trnD, cox2*
6	*atp6*

et al. 2016). Based on genomic data of *I. linei*, Mikhailov et al. identified orthonectids as highly simplified spiralians located close to Annelida in the phylogenetic tree. The most recent studies based on single mt genome of *I. linei* defined orthonectids as a highly degenerate annelid worms but could not precisely place them within Annelida (Schiffer et al. 2018). Based on only one Orthonectida species the authors found a short block of genes (*nad1, nad6, cob*) which are found in the same order as in the lophotrochozoan ancestor but did not find a similar block in the Pleistoannelida (Schiffer et al. 2018).

Our gene order analysis and ancestral gene order reconstruction showed that orthonectids and their ancestral genome have up to three common gene blocks with Pleistoannelida (Fig. 7) while in basal branch of Annelida we observe a variety of gene orders and, therefore, it is more difficult to find common gene blocks between all basal branch species and orthonectids (Weigert et al. 2016). According to this, we proposed that Orthonectida are closer to Pleistoannelida, than to the Basal Branch Annelida. To check our hypothesis, we inferred phylogenetic

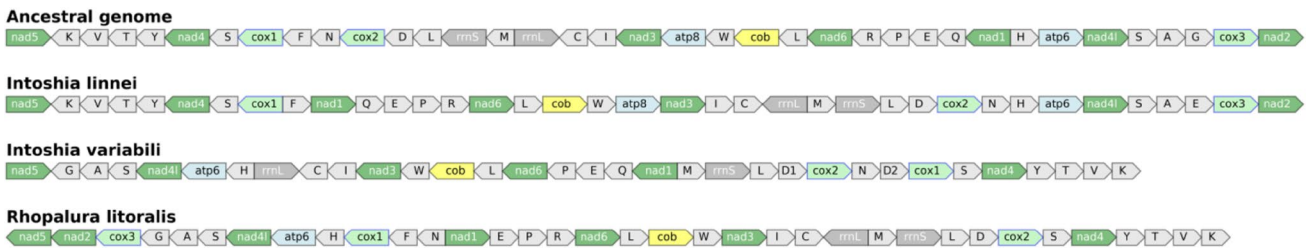

Fig. 5 Comparison of ancestral gene order with Orthonectida species. Protein-coding and ribosomal genes are denoted by their names and one capital letter indicates the amino acid for the tRNAs

Table 7 Pairwise distance matrix of the Orthonectida mt genomes

	Intoshia linei	*Intoshia variabili*	*Rhopalura litoralis*	*Ancestral genome*
Intoshia linei	0	3	4	2
Intoshia variabili	3	0	4	8
Rhopalura litoralis	4	4	0	6

relationships between Orthonectida and Annelida species where most of them belong to Pleistoannelida and several of them to basal branch as an outgroup.

Our results suggest that all orthonectid species are grouped together within Annelida. Moreover, they are forming a long branch, which is close to Clitellata (Fig. 8). As seen from the analysis, Orthonectida might have separated from Clitellata a long time ago, and their mt genomes evolved faster than Annelida mt genomes, hence orthonectids are located on a longer branch. Noteworthy, a clitellum made of secretory cells in the midbody has been registered in some orthonectids (Metschnikoff 1881; Shtein 1953).

To summarize, mt genomes of orthonectids have undergone reduction and gene loss, they have complicated organization, which is manifested in strand asymmetry in nucleotide composition, as well as in some features of intergenic non-coding regions, tRNA duplication and folding. Moreover, all species of Orthonectida have a unique gene order with complicated rearrangement landscape. Significant differences in mitochondrial genomes in the three orthonectid species could be explained by the fact that their host species belong to different taxa (flat worms, nemertines and gastropods). Among analyzed mt genomes of Orthonectida, *I. linei* possesses the closest gene order to the ancestral genome. All Orthonectida species are monophyletic, and in the phylogenetic tree are close to Pleistoannelida, and specifically, to Clitellata.

Fig. 6 A minimum number of events that rearranges the mitochondrial gene order of ancestral genome (top) via the intermediate gene order (middle) into the gene order of Orthonectida (bottom)

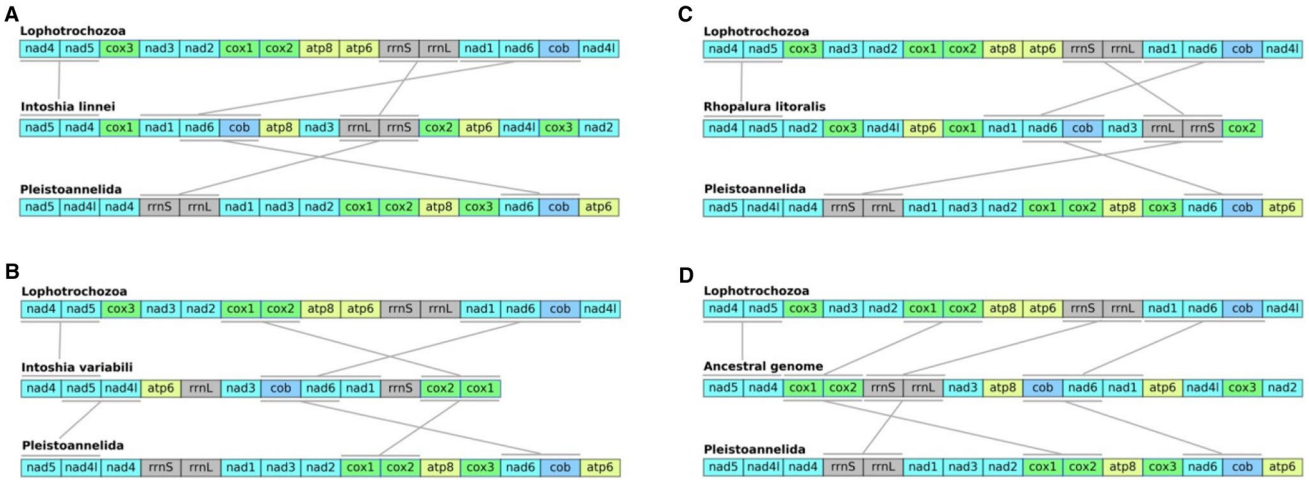

Fig. 7 Comparison of Orthonectida and ancestral mitochondrial genes with the putative ground pattern of Pleistoannelida and Lophotrochozoa. Only PCGs are shown. The direction of the transcription is not considered

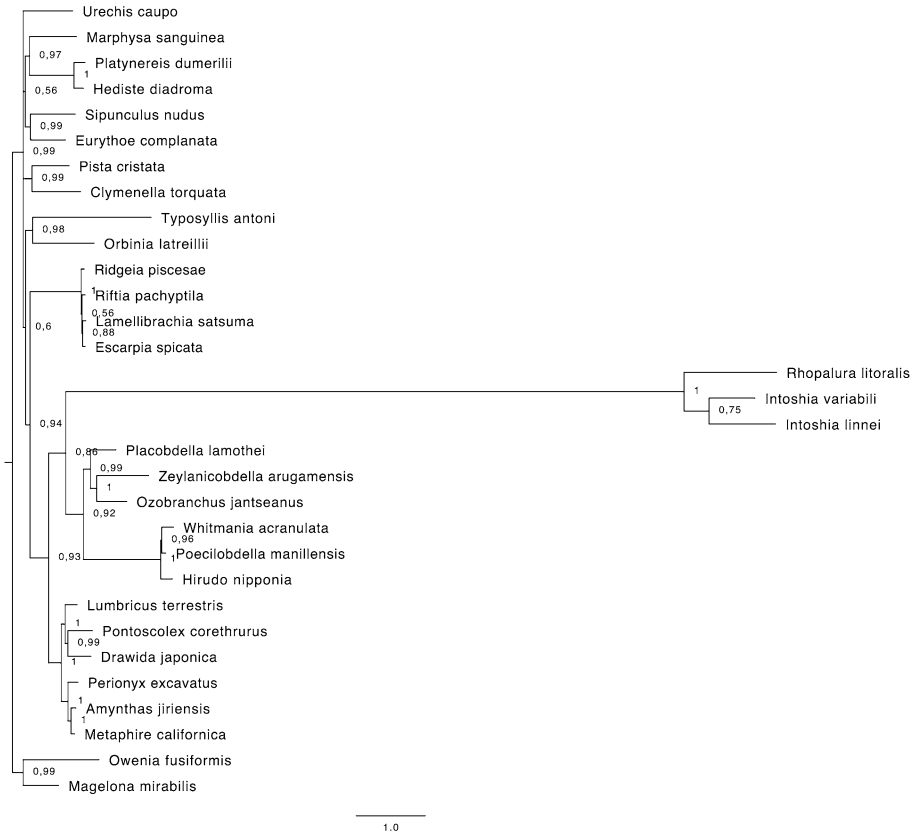

Fig. 8 Phylogenetic relationships of Orthonectida based on mitochondrial genome data (31 taxa, 2759 amino acid positions). Bootstrap support values (BS) are shown for each node. Node numbers show poster probability values. Scale bar represents substitutions/site

Acknowledgements We are thankful to Dr. M. Makarov for providing accommodation and facilities for field work at the Barentz Sea, and also to the staff of the Marine Biological Station of the Saint Petersburg State University for assistance in collecting material. The work was supported by Grant of Russian Foundation for Basic Research (RFBR) no. 19-04-0021 and Saint Petersburg State University research Grant 1.40.496.2017. Bioinformatics was supported with Russian Science Foundation 17-74-10103 Grant to Natalya Bondarenko. This study utilized equipment of the "Centre Biobank" and "Computing Centre SPbU" of Saint Petersburg State University.

References

Abascal F, Zardoya R, Posada D (2006) GenDecoder: genetic code prediction for metazoan mitochondria. Nucleic Acids Res 34:389–393

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021

Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M (2007) CREx: Inferring genomic rearrangements based on common intervals. Bioinformatics 23:2957–2958

Bernt M, Braband A, Schierwater B, Stadler PF (2013) Genetic aspects of mitochondrial genome evolution. Mol Phylogenet Evol 69:328–338

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

Curole JP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol Evol 14:394–398

Darling AE, Mau B, Blatter FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403

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

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Kozloff EN (1992) The genera of the phylum Orthonectida. Cah Biol Mar 33:377–406

Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21:1095–1109

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

Metschnikoff E (1881) Untersuchungen über Orthonectiden. Zeitschrift fuer Wissenschaftliche Zoologie 35:282–303

Mikhailov KV, Slyusarev GS, Nikitin M, Logacheva MD, Penin AA, Aleoshin VV, Panchin YV (2016) The genome of *Intoshia linei* affirms orthonectids as highly simplified spiralians. Curr Biol 26:1768–1774

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358

Plazzi F, Puccio G, Passamonti M (2016) Comparative large-scale mitogenomics evidences clade-specific evolutionary trends in mitochondrial DNAs of Bivalvia. Genome Biol Evol 8:2544–2564. https://doi.org/10.1093/gbe/evw187

Rambaut A, Drummond AJ (2009) Tracer version 1.5. http://tree.bio.ed.ac.uk/software/tracer/. Accessed 30 Nov 2009

Rota-Stabelli O, Yang Z, Telford M (2009) MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. Mol Phylogenet Evol 52:268–272

Rota-Stabelli O, Kayal E, Gleeson D, Daub J, Boore JL, Telford MJ, Pisani D, Blaxter M, Lavrov DV (2010) Ecdysozoan mitogenomics: evidence for a common origin of the legged invertebrates, the Panarthropoda. Genome Biol Evol 2:425–440. https://doi.org/10.1093/gbe/evq030

Schiffer PH, Robertson HE, Telford MJ (2018) Orthonectids are highly degenerate annelid worms. Curr Biol 28:1970–1974. https://doi.org/10.1016/j.cub.2018.04.088

Shtein GA (1953) Ortonektidy roda Rhopalura Giard nekotorykh molliusko v Barentsova moria. Uchenye Zapiski Karelo-Finskogo Universiteta. Biol Nauki 5:171–206

Slusarev GS (2018) Ortonectida. Miscellaneous Invertebrates. Handbook of Zoology. https://www.degruyter.com/view/Zoology/bp_048833-3_2. Accessed 26 Oct 2018

Slyusarev GS (1994) The fine structure of the female *Intoshia variabili* (Alexandrov & Sljusarev) (Mesozoa: Orthonectida). Acta Zool 75:311–321

Slyusarev GS, Ferraguti M (2002) Sperm structure of *Rhopalura litoralis* (Orthonectida). Invertebr Biol 121:91–94

Tesler G (2002) GRIMM: genome rearrangements web server. Bioinformatics 18:492–493

Weigert A, Golombek A, Gerth M, Schwarz F, Struck TH, Bleidorn C (2016) Evolution of mitochondrial gene order in Annelida. Mol Phylogenet Evol 94:196–206

Wey-Fabrizius AR, Podsiadlowski L, Herlyn H, Hankeln T (2013) Platyzoan mitochondrial genomes. Mol Phylogenet Evol 69:365–375. https://doi.org/10.1016/j.ympev.2012.12.015