

SCIENTIFIC REPORTS

OPEN

The mitochondrial genomes of the acoelomorph worms *Paratomella rubra*, *Isodiametra pulchra* and *Archaphanostoma ylvae*

Helen E. Robertson¹, François Lapraz^{1,2}, Bernhard Egger^{1,3}, Maximilian J. Telford¹ & Philipp H. Schiffer¹

Acoels are small, ubiquitous - but understudied - marine worms with a very simple body plan. Their internal phylogeny is still not fully resolved, and the position of their proposed phylum Xenacoelomorpha remains debated. Here we describe mitochondrial genome sequences from the acoels *Paratomella rubra* and *Isodiametra pulchra*, and the complete mitochondrial genome of the acoel *Archaphanostoma ylvae*. The *P. rubra* and *A. ylvae* sequences are typical for metazoans in size and gene content. The larger *I. pulchra* mitochondrial genome contains both ribosomal genes, 21 tRNAs, but only 11 protein-coding genes. We find evidence suggesting a duplicated sequence in the *I. pulchra* mitochondrial genome. The *P. rubra*, *I. pulchra* and *A. ylvae* mitochondria have a unique genome organisation in comparison to other metazoan mitochondrial genomes. We found a large degree of protein-coding gene and tRNA overlap with little non-coding sequence in the compact *P. rubra* genome. Conversely, the *A. ylvae* and *I. pulchra* genomes have many long non-coding sequences between genes, likely driving genome size expansion in the latter. Phylogenetic trees inferred from mitochondrial genes retrieve Xenacoelomorpha as an early branching taxon in the deuterostomes. Sequence divergence analysis between *P. rubra* sampled in England and Spain indicates cryptic diversity.

Acoel flatworms are small, soft-bodied, unsegmented, marine animals lacking a gut epithelium, coelomic cavity, and anus. Instead, they typically possess a ventral mouth opening, and a simple syncytial digestive system¹. Due primarily to the common attributes of acoelomate body and the absence of a through gut, Acoela were traditionally grouped as an order within the Platyhelminthes. The first molecular systematic studies on these animals using small subunit (SSU) ribosomal RNA gene sequences revealed that the Acoelomorpha are in fact a distinct lineage, quite separate from the main clade of the Platyhelminthes (Rhabditophora and Catenuilida)²⁻⁴. Instead, these initial molecular studies supported a position of the Acoelomorpha diverging prior to the protostome/deuterostome common ancestor. More recently, the Acoelomorpha have been linked to the similarly simple marine worm *Xenoturbella* in the new phylum Xenacoelomorpha, making sense of their shared simple body plan and other shared morphological characters, such as unusual ciliary ultrastructure⁵ and their simple basiepidermal nervous system⁶. Despite considerable efforts, the position of Xenacoelomorpha within the Metazoa remains unresolved, with alternative lines of evidence placing them either as the sister group to the remaining Bilateria (protostomes and deuterostomes)^{7,8}, or as a phylum within the deuterostomes⁹. A better understanding of acoel phylogeny and evolution is therefore integral to answering central questions concerning the evolution of Bilateria and its subtaxa. To this end more genomic data are needed.

Metazoan mitochondrial DNA (mtDNA) is a closed-circular molecule typically comprising 37 genes which are, for the most part, invariant across the Metazoa¹⁰. These include the two rRNAs of the mitochondrial ribosome, 22 tRNAs necessary for translation, and 13 protein-coding genes for the enzymes of oxidative phosphorylation. *atp8* is the only gene known to have been commonly lost from this complement, and this has been observed

¹Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, UK. ²Present address: CNRS/UMR 7277, institut de Biologie Valrose, iBV, Université de Nice Sophia Antipolis, Parc Valrose, Nice cedex 2, France. ³Present address: Institute of Zoology, University of Innsbruck, Technikerstr. 25, 6020, Innsbruck, Austria. Correspondence and requests for materials should be addressed to H.E.R. (email: helen.robertson.09@ucl.ac.uk) or P.H.S. (email: philipp.schiffer@gmail.com)

in a number of independent metazoan lineages, including the acoel *Symsagittifera roscoffensis*¹¹. In addition to primary sequence data, mtDNA has a number of other features which can be used for phylogenetic inference, including variations in mitochondrial genetic code¹²; a higher rate of sequence evolution than nuclear DNA¹³; changes in gene order; and changes in the secondary structure of rRNAs and tRNAs¹⁴.

Mitochondrial gene sequences have been used extensively for phylogenetic inference. In a recent paper, Rouse *et al.* used mitochondrial protein-coding sequence data from four newly discovered species of *Xenoturbella* (*X. hollandorum*, *X. churro*, *X. monstrosa*, and *X. profunda*) to infer the internal phylogeny of the Xenoturbellida¹⁵. Wider phylogenetic inference including mitochondrial proteins from these species placed Xenacoelomorpha with the deuterostomes¹⁵, corroborating previous mitochondrial phylogenetic analysis of this phylum^{9,16,17}.

Mitochondrial gene content is largely invariable across the Metazoa, with the order in which genes are arranged being fairly stable and conserved for up to hundreds of millions of years in some metazoan lineages. Rearrangement events, thought to occur via a model of 'duplication and deletion'^{14,18}, whereby a portion of the mitochondrial genome is duplicated, and the original copy of the duplicated gene subsequently deleted, are rare. The infrequency of such rearrangements, and the huge number of possible rearrangement scenarios, means that convergence on the same gene order in unrelated lineages is unlikely. Gene order is thus likely to retain evolutionary signals, with a common gene order being indicative of common ancestry and informative for the study of metazoan divergence¹⁹. Rearrangement of genes within the mitochondrial genome of different species can be a particularly powerful tool in the analysis of phylogenetic relationships¹⁴ and may also indicate accelerated evolution in a taxon.

In this study, we describe the mitochondrial genomes from three species of acoel: *Paratomella rubra*, *Isodiametra pulchra* and *Archaphanostoma ylvaе*. Adult specimens of all animals are approximately 1 mm in length, and, as is typical for small acoel species, they occupy the littoral and sub-littoral zones of marine ecosystems: *P. rubra* has been described across Europe and North America^{20,21}, and *I. pulchra* lives abundantly in the mud flats of Maine²². *A. ylvaе* has been described from the West coast of Sweden²³. All species move freely within the sediment by gliding on a multiciliated epidermis. First described by Rieger and Ott²¹, *P. rubra* is an elongate and flattened worm belonging to the family Paratomellidae^{24,25}. A 9.7 kb fragment of mitochondrial genome has previously been described from specimens of *P. rubra* collected on the Mediterranean coast of Spain²⁶. *I. pulchra* belongs to the family Isodiametridae; it can be maintained long-term in culture and has been used experimentally for *in situ* hybridisation, RNAi, and other molecular protocols^{22,27,28}. Its use as a 'model acoel' therefore makes this species particularly valuable for investigation. *A. ylvaе* also belongs to the Isodiametridae family of acoels. Originally described by K anneby *et al.* in 2014, its *cox1* gene has been sequenced and used for classification, but no further genes from its mitochondrial genome have been sequenced²³.

Results

Genomic composition. We assembled 14,954 base pairs of the *P. rubra* mitochondrial genome, starting from three genome assembly fragments and using Sanger sequencing of additional PCR fragments (Fig. 1a). We were unable to close the circular mitochondrial genome of *P. rubra*, but our 14.9 kb sequence contains all 13 protein-coding genes, both ribosomal genes and 22 putative tRNAs. Compared to the fragment of the genome previously published we have found four additional protein-coding genes and 12 additional tRNAs²⁶. All genes are found exclusively on one strand of the sequence. Allowing for overlap, protein-coding genes account for 74.79% of the genomic sequence; ribosomal genes 13.95%; tRNA genes 9.10%, and non-coding DNA just 2.04%. A 156 nucleotide-long stretch of non-coding sequence is found between *cytochrome c oxidase subunit 2 (cox2)* and *NADH dehydrogenase subunit 1 (nad1)*.

In *P. rubra*, *trnS2* is predicted entirely within the sequence coding for *nad1*, and also has clear deviation from the traditional 'cloverleaf' secondary structure of tRNA. In addition, three of the predicted tRNAs have minor overlaps with protein-coding genes: *trnA* with *nad3* (20 nucleotides); *trnK* with *nad4l* (18 nucleotides) and *trnS1* with *nad4* (six nucleotides); and all but five nucleotides of *trnL1* are predicted within the same sequence as *rrnL* (Fig. 1a, Table 1). With the exception of *trnT*, all predicted tRNAs have an amino-acyl acceptor stem composed of seven base pairs, and all predicted tRNAs apart from *trnT* and *trnS2* have a five base pair anticodon stem (Fig. 2). 11 tRNAs have one or two G-T mismatches in their acceptor or anticodon stems (A,C,G,I,K,L1,L2,P,Q,R,T). All tRNAs have a DHU arm of three or four nucleotides. The structure of the T ψ C arm shows greater variability, with a number of tRNAs having either a truncated stem, or the arm entirely lacking (Fig. 2).

For *I. pulchra*, we initially recovered a 13 kb contig, a 3.5 kb contig and a 19 kb contig of mitochondrial sequence from our transcriptomic data. The entire 13 kb contig and 2.4 kb of the 3.5 kb contigs were found to be perfectly matching subsets of the longer 19 kb sequence (Fig. 3). We designed several sets of PCR primers to verify the sequence between the 3' end of the 13 kb and 5' end of the 3.5 kb fragments found on the long 19 kb fragment (Fig. 3), however, no PCR amplification completely bridged the sequence between the 13 kb to 3.5 kb fragment. We found that the last (3') 300 bp of the 13 kb fragment was duplicated in the opposite orientation within the end (3') region of the 3.5 kb fragment. Although the long 19 kb fragment contained the repeated region between the 13 kb and 3.5 kb fragments, we were not able to connect sequences on both sides of the repeated region by PCR. The placement of Sanger sequencing fragments containing the repeat remained ambiguous. We are thus not confident in the assembly of the 19 kb transcriptomic sequence in this section, and therefore treat it as uncertain. Instead we focused on verifying the sequence of the two smaller fragments and on amplifying and sequencing the region lying between them. We reconfirmed the majority of the 13 kb fragment using PCR amplification and Sanger sequencing. We were also able to amplify and sequence fragments joining the 3' end of the 3.5 kb fragment with the 5' end of a 1.3 kb fragment containing the *rrnL* gene, which we had identified in transcriptome sequence data using blast. This contig included the duplicated region at the 3' end of the 3.5 kb fragment (Fig. 3).

In summary, we find the *I. pulchra* mitochondrial genome to have a span of at least 18,725 base pairs (Fig. 1b) based on our PCR validation of the transcriptomic data. This covers the region from the start of the 5' end of the

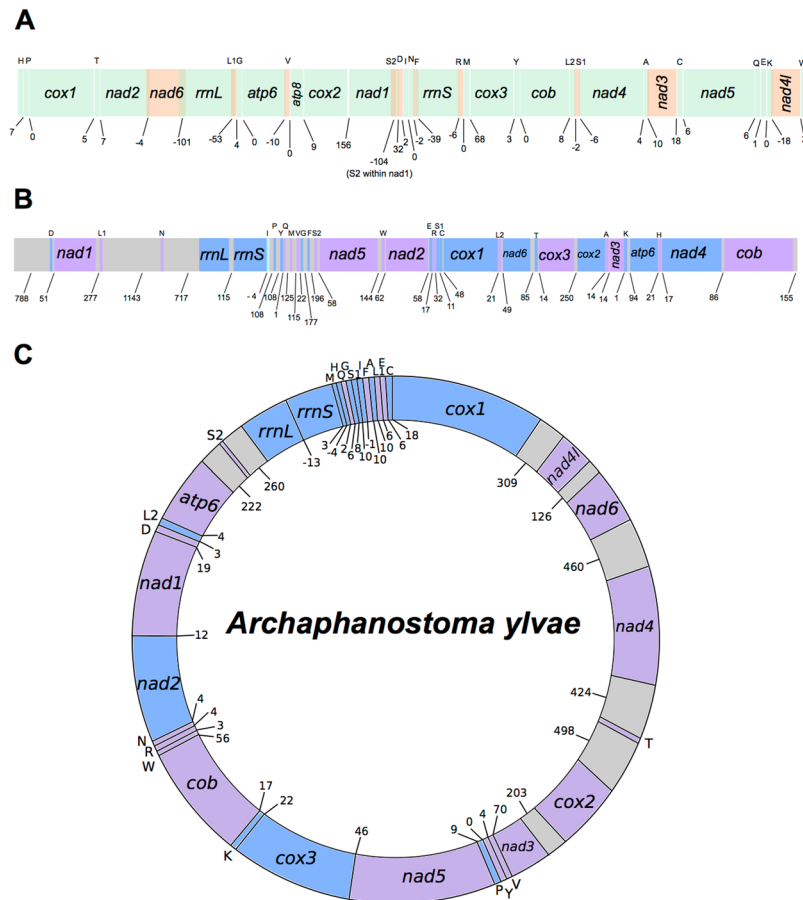


Figure 1. Overview of the mitochondrial genome sequences we resolve for *Paratomella rubra*, *Isodiametra pulchra* and *Archaphanostoma ylvae* (Xenacoelomorpha: Acoela). Genes not drawn to scale. Numbers beneath the sequences show intergenic spaces (positive values) or intergenic overlap (negative values). Protein-coding genes are denoted by three letter abbreviations; ribosomal genes by four letter abbreviations. tRNAs are shown by single uppercase letters. **(A)** *P. rubra* 14,957 base-pair long sequence. All genes found on the positive (forward) strand. Where genes, rRNAs or tRNAs are coloured orange, this is solely to demonstrate overlap with adjacent genes, rRNAs or tRNAs. **(B)** *I. pulchra* 18,725 base-pair long sequence. Genes found on the positive (forward) strand are coloured blue; genes on the negative (reverse) strand are coloured purple. Non-coding sequence shown in grey. **(C)** *A. ylvae* 16,619 nucleotide-long mitochondrial genome. Genes found on the positive (forward) strand are coloured blue; genes on the negative (reverse) strand are coloured purple. Non-coding regions greater than 100 nucleotides in length are shown in grey.

3.5 kb sequence which is linked through PCR amplicons to the 5' end of the 13 kb sequence (including the 1.3 kb *rrnL* contig), and up to the start of the duplicated sequence at the 3' end of the 13 kb sequence (Fig. 3). As we were not able to bridge the region between the 3' end of the 13 kb fragment and the 5' end of the 3.5 kb fragment with PCR, we could not confirm the validity of the duplicated sequence at this position nor fully close the circular mitochondrial genome. It is therefore likely that the entire mitochondrial genome is larger than 19 kb, and may include the duplicated sequence. The sequence we are confident on presenting contains both ribosomal genes, all tRNAs and 11 protein-coding genes. These protein-coding genes and RNAs are encoded on both the plus and minus strands. No sequences resembling either *atp8* or *nad4l* could be found in our sequence.

In the 18.7 kb sequence, protein-coding genes account for 56.66%; ribosomal genes contribute 8.15% and tRNA genes 7.77%. Compared to the *P. rubra* and *S. roscoffensis* mitochondrial genomes, intergenic space in the *I. pulchra* sequences is unusually high: non-coding DNA accounts for 22.72% of the sequences, including 14 intergenic regions of greater than 100 base pairs.

We identified all 22 expected tRNAs in the *I. pulchra* mitochondrial genome. Predicted sequences for *rrnS* and *trnI* overlap by four base pairs, but no other overlaps were found between any tRNAs or with any protein-coding genes (Fig. 1b, Table 2). All predicted tRNAs have an amino-acyl acceptor stem composed of seven base pairs and a five base pair anticodon stem, with the exception of *trnE*, *trnF* and *trnS2*, which have an anticodon stem composed of only four base pairs (Fig. 4). The structure of the DHU arms and T ψ C show greater variability, and are composed of either 3 or 4, or between 3 and 6, base pairs respectively, across the 22 tRNAs. Whilst the T ψ C arm is missing entirely in *trnQ*, and very truncated in *trnE*, *trnF*, *trnG* and *trnP*, more of the predicted tRNAs fit the

Feature	Strand	Start	Stop	Length (nucleotides)	Length (AA)	Start Codon	Stop Codon	Intergenic region
<i>trnH (gtg)</i>	+	368	426	59				7
<i>trnP (tgg)</i>	+	434	495	62				0
<i>cox1</i>	+	496	2058	1563	521	ATA	TAA	5
<i>trnT (tgt)</i>	+	2064	2123	60				7
<i>nad2</i>	+	2131	3105	975	325	ATT	TAG	-4
<i>nad6</i>	+	3102	3563	462	154	ATA	TAA	-101
<i>rrnL</i>	+	3463	4819	1357				-53
<i>trnL1 (tag)</i>	+	4767	4824	58				4
<i>trnG (tcc)</i>	+	4829	4887	59				0
<i>atp6</i>	+	4888	5496	609	203	ATA	TAG	-10
<i>trnV (tac)</i>	+	5487	5553	67				0
<i>atp8</i>	+	5554	5730	177	59	ATT	TA-	9
<i>cox2</i>	+	5740	6402	663	221	ATT	TAA	156
<i>nad1</i>	+	6559	7602	1044	348	ATT	TAA	-104
<i>trnS2 (tga)</i>	+	7499	7568	70				32
<i>trnD (gtc)</i>	+	7601	7662	62				2
<i>trnI (gat)</i>	+	7665	7728	64				0
<i>trnN (gtt)</i>	+	7729	7798	70				-2
<i>trnF (gaa)</i>	+	7797	7856	60				-39
<i>rrnS</i>	+	7818	8547	730				-6
<i>trnR (tcg)</i>	+	8542	8608	67				0
<i>trnM (cat)</i>	+	8609	8669	61				68
<i>cox3</i>	+	8738	9523	786	262	ATT	TAA	3
<i>trnY (gta)</i>	+	9527	9585	59				0
<i>cob</i>	+	9586	10668	1083	361	ATA	TAA	8
<i>trnL2 (taa)</i>	+	10677	10737	61				-2
<i>trnS1 (gct)</i>	+	10736	10799	64				-6
<i>nad4</i>	+	10794	12119	1326	442	ATC	TAA	4
<i>trnA (tgc)</i>	+	12124	12181	58				10
<i>nad3</i>	+	12192	12551	360	120	ATT	TAG	18
<i>trnC (gca)</i>	+	12570	12629	60				6
<i>nad5</i>	+	12636	14387	1752	584	ATA	TAG	6
<i>trnQ (ttg)</i>	+	14394	14449	56				1
<i>trnE (ttc)</i>	+	14451	14510	60				0
<i>trnK (ttt)</i>		14511	14576	66				-18
<i>nad4l</i>	+	14559	14867	309	103	ATA	TAA	2
<i>trnW (tca)</i>	+	14870	14933	64				

Table 1. Organisation of the *Paratomella rubra* 14.9 kb mitochondrial genome sequence. All genes found on the 'positive' strand.

stereotypical 'cloverleaf' secondary structure than has been found for other acoel species, including *S. roscoffensis* and *P. rubra* (Fig. 4).

The complete, closed circular mitochondrial genome of *A. ylvae* was recovered from genome sequencing data of *P. rubra* specimens collected from Yorkshire, UK. Contamination of the *P. rubra* samples was confirmed using NCBI Blast, which yielded a 99% identical sequence to the published *A. ylvae cox1*. The complete *A. ylvae* mitochondrial genome is 16,619 nucleotides in length, and contains 12 protein-coding genes, both rRNAs, and 22 predicted tRNAs (Fig. 1c, Table 3). With *cox1* at the start of the genome on the 'positive' strand, all other protein-coding genes apart from *cox3* and *nad2* are found on the 'negative' strand. Both rRNAs are found on the positive strand, and tRNAs are distributed between the two. Accounting for a small amount of overlap between genes - totalling 18 nucleotides of overlap across the whole genome - protein-coding genes make up 64.72% of the genome. tRNAs contribute 8.91%, and rRNAs 9.31%. As found for *I. pulchra*, non-coding DNA makes up a large amount of the genome, totalling 17.17%.

We identified possible sequences for all 22 mitochondrial tRNAs in the *A. ylvae* genome, although four of these (*trnE*, *trnI*, *trnK* and *trnS1*) have an e-value prediction of greater than 0.0001. All predicted secondary structures of the tRNAs in the *A. ylvae* mitochondrial genome have standard-length acceptor and anticodon stems, and the majority - with the exception of *trnK*, *L1*, *L2*, *N*, *S2* and *Y* - have a four nucleotide-long D-loop (Fig. 5). As in the other acoel mitochondrial genomes, the greatest variability is found in the T ψ C arm, which is truncated in *trnD*, *E*, *F*, *L1*, *P*, *V*, *W* and *Y* and missing in *trnQ* (Fig. 5).

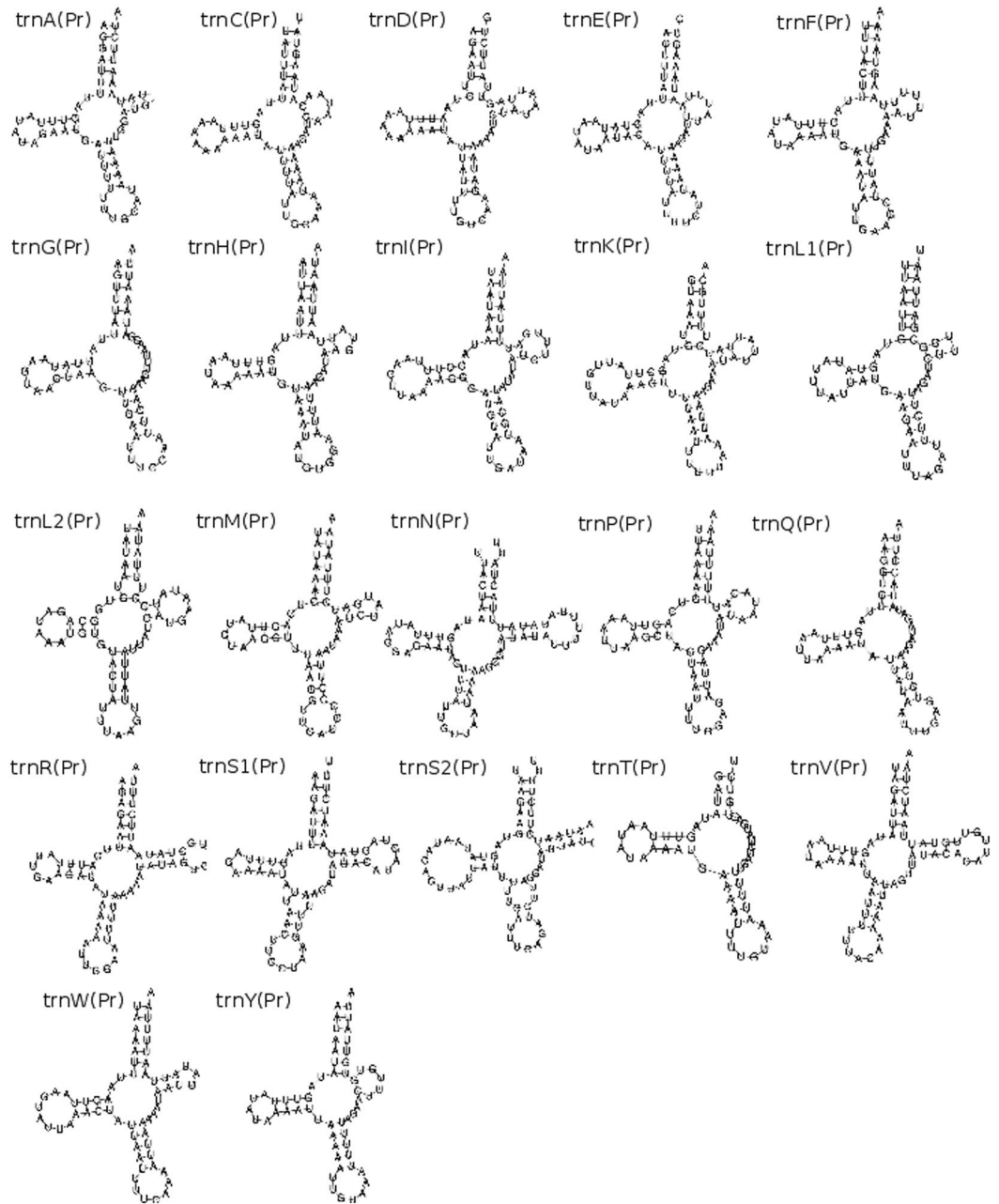


Figure 2. Predicted secondary structures of tRNAs from the mitochondrial genome sequence of *Paratomella rubra* as predicted by MiTFi in Mitos.

The *P. rubra* genome is 78.15% A + T rich, which is higher than the A + T content calculated for the *I. pulchra* genomic sequence at 67.28%, and the *A. ylvae* complete genome at 74.70%. Overall nucleotide usage on the plus strand of *P. rubra* is: A = 29.29%; T = 48.86%; C = 6.77% and G = 15.10%; GC-skew = 0.38 and absolute AT-skew = 0.25. Overall nucleotide usage for *I. pulchra* is: A = 34.04%; T = 33.24%; C = 16.45% and G = 16.27%; GC-skew = 0.006 and AT-skew = 0.012. For *A. ylvae*, A = 40.41%; T = 34.29%; C = 12.82% and G = 12.47%; GC-skew = 0.014 and AT-skew = 0.082. GC-skew and AT-skew absolute values for *P. rubra* are much higher than that of *S. roscoffensis*, although the absolute values for *I. pulchra* and *A. ylvae* are comparatively low¹¹. AT-skew value for the *P. rubra* sequence is just 0.01 different from that of the published partial *P. rubra* genome, and GC-skew is slightly higher (published *P. rubra* GC-skew = 0.32)²⁶.

Gene order and gene arrangement. All thirteen protein-coding genes in *P. rubra* have complete initiation codons: ATA (x5) and ATT (x8). Five of the protein-coding genes previously published differ in the nucleotide sequence of their start codons: *nad2*, *atp8*, *cox2*, and *cox3* all have ATA as an initiation codon in our analysis, compared to ATT found in previous analysis²⁶. Twelve of the genes have full stop codons: TAA (x9) or TAG (x3). *atp8* was found to have a truncated stop codon (TA-), which is thought to be completed during

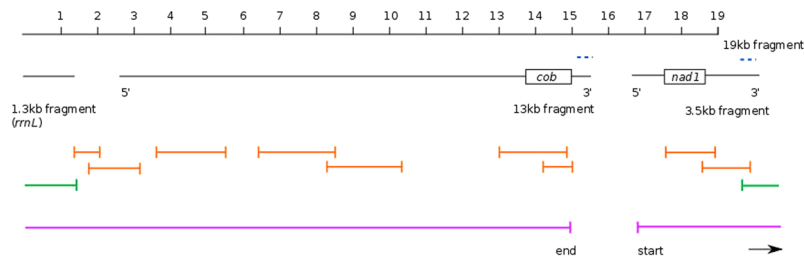


Figure 3. Overview of the initial transcriptome assembly fragments and PCR strategy for scaffolding the *Isodiametra pulchra* mitochondrial genome. 1.3 kb, 13 kb and 3.5 kb fragments aligned to a continuous 19 kb fragment, with the location of the duplicated sequence in the 13 kb and 3.5 kb fragments shown by blue dashed lines. The ‘start’ and ‘end’ regions of the 13 kb and 3.5 kb fragments are annotated by 5’ (start) and 3’ (end). The approximate location of *cob* and *nad1* protein-coding sequence are shown for reference. Reliable PCR-amplicons are shown in orange; the green PCR fragment indicates successful joining of the 3’ end of the 3.5 kb fragment to the *rrnL* fragment, including the duplicated section. The 18,725 base-pair long sequence we resolve is indicated by the pink lines, from ‘start’ to ‘end’.

Feature	Strand	Start	Stop	Length (nucleotides)	Length (AA)	Start Codon	Stop Codon	Intergenic region
<i>trnD (gtc)</i>	+	789	848	60				51
<i>nad1</i>	–	900	1784	885	295	ATG	TAA	277
<i>trnL1 (tag)</i>	–	2062	2130	69				1143
<i>trnN (gtt)</i>	–	3274	3339	66				717
<i>rrnL</i>	+	4057	4657	601				115
<i>rrnS</i>	+	4773	5698	926				–4
<i>trnI (gat)</i>	+	5695	5768	74				108
<i>trnP (tgg)</i>	+	5877	5939	63				108
<i>trnY (gta)</i>	+	6048	6111	64				1
<i>trnQ (ttg)</i>	–	6113	6173	61				125
<i>trnM (cat)</i>	–	6299	6360	62				115
<i>trnV (tac)</i>	–	6476	6543	68				22
<i>trnG (tcc)</i>	+	6566	6627	62				177
<i>trnF (gaa)</i>	+	6805	6873	69				196
<i>trnS2 (tga)</i>	–	7070	7139	70				58
<i>nad5</i>	–	7198	8907	1710	570	ATG	TAA	144
<i>trnW (tca)</i>	–	9052	9118	67				62
<i>nad2</i>	–	9181	10233	1053	351	ATG	TAA	58
<i>trnE (ttc)</i>	+	10292	10355	64				17
<i>trnR (tcg)</i>	–	10373	10439	67				32
<i>trnS1 (tct)</i>	+	10472	10539	68				11
<i>trnC (gca)</i>	+	10551	10613	63				48
<i>cox1</i>	+	10662	12197	1536	512	ATA	TAG	21
<i>trnL2 (taa)</i>	–	12219	12286	68				49
<i>nad6</i>	+	12336	12811	476	159	ATG	T–	85
<i>trnT (tgt)</i>	+	12897	12963	67				14
<i>cox3</i>	–	12978	13775	798	266	ATG	TAA	250
<i>cox2</i>	+	14026	14640	615	205	ATA	TAA	14
<i>trnA (tgc)</i>	–	14655	14718	64				14
<i>nad3</i>	–	14733	15110	378	126	ATG	TAA	1
<i>trnK (ttt)</i>	+	15112	15178	67				94
<i>atp6</i>	+	15273	15955	683	228	ATA	TA–	21
<i>trnH (gtg)</i>	–	15977	16042	66				17
<i>nad4</i>	+	16060	17403	1344	448	ATG	TAA	86
<i>cob</i>	–	17490	18570	1081	361	ATA	T–	155

Table 2. Organisation of the *Isodiametra pulchra* 18.7 kb mitochondrial genome fragment sequence.

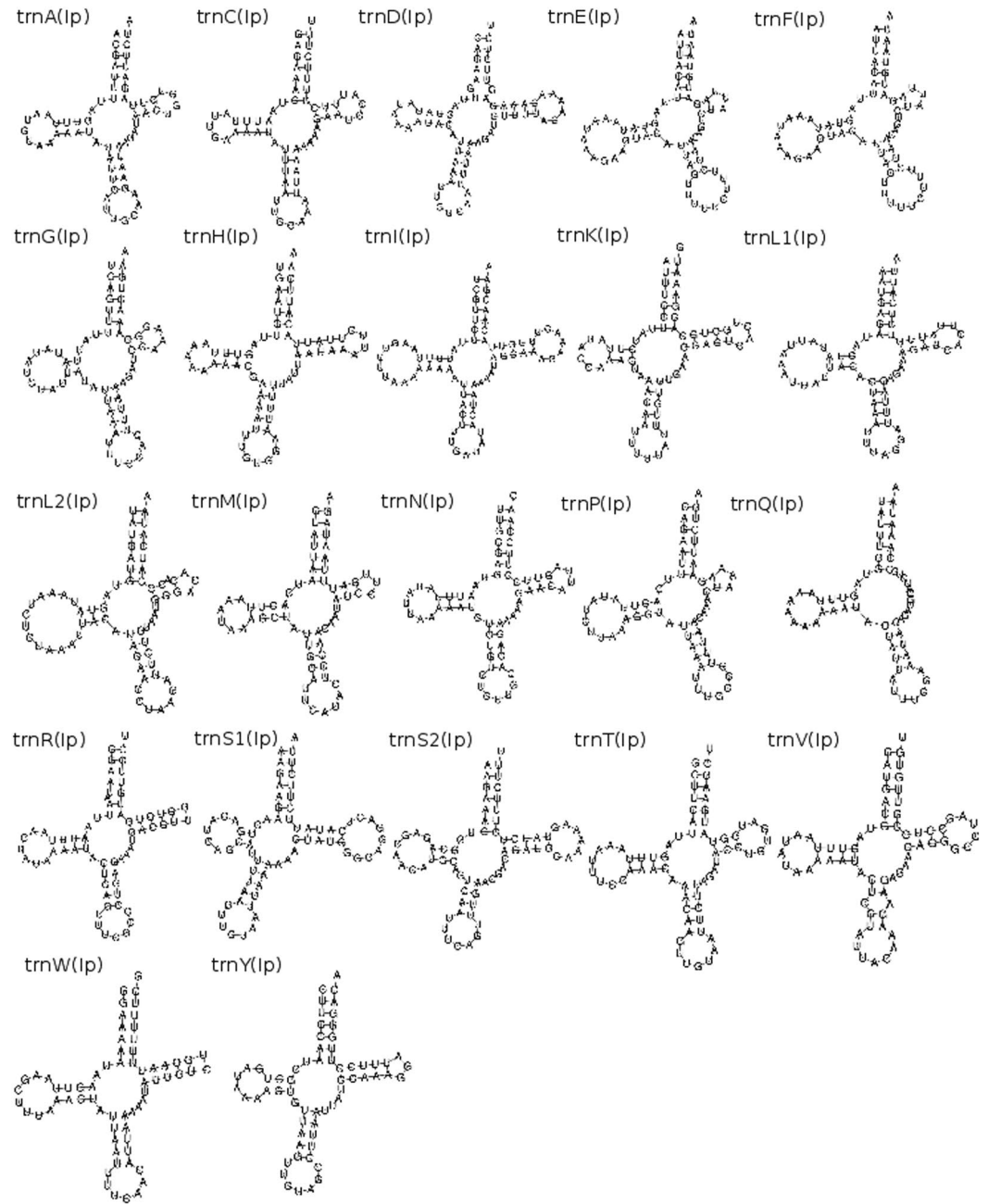


Figure 4. Predicted secondary structure of tRNAs from the mitochondrial genome sequences of and *Isodiametra pulchra* as predicted by MITFi in Mitos.

post-transcriptional modification (Table 1). The eleven protein-coding genes found for *I. pulchra* also have full initiation codons: ATA (x4) and ATG (x7). Eight of the genes for this species have full stop codons: TAA (x7) and TAG (x1); *nad6*, *atp6* and *cob* are inferred to have truncated stop codons (Table 2). Initiation codes in *A. ylvae* are: ATA x4, ATG x7 and ATT x1; all genes have TAA as stop codons, with the exception of *nad6*, which has TAG (Table 3). As in other invertebrate mitochondrial genomes, our data indicates a deviation from the ‘standard’ genetic code, with ATA encoding the start codon methionine, M, instead of isoleucine, I.

We found all genes in *P. rubra* on the ‘plus’ strand. In *I. pulchra*, genes are distributed over the plus and minus strands, with just two ‘blocks’ of genes with the same transcriptional polarity clustered together (*rrnL-rrnS-trnI-trnP-trnY*; *trnS2-nad5-trnW-nad2*). Similarly, in *A. ylvae* genes are distributed across the two strands, with two clustered ‘blocks’ of genes and tRNAs (*nad4l-nad6-nad4-trnT-cox2-nad3-trnV-trnY*; *trnM-trnH-trnQ-trnG-trnS1-trnI-trnF-trnA-trnL1-trnE-trnC*). Whilst the *P. rubra* mitochondrial sequence has a large degree of overlap between adjacent genes, the opposite is true for *I. pulchra* and *A. ylvae*. Unlike other metazoan mitochondrial genomes, where genes are adjacent or overlapping and one or two larger non-coding regions are commonly found, *I. pulchra* non-coding sequence is found consistently between protein-coding genes and

Feature	Strand	Start	Stop	Length (nucleotides)	Length (AA)	Start Codon	Stop Codon	Intergenic
<i>cox1</i>	+	1	1539	1539	513	ATA	TAA	309
<i>nad4l</i>	–	1849	2133	285	95	ATG	TAA	126
<i>nad6</i>	–	2260	2727	468	156	ATG	TAG	460
<i>nad4</i>	–	3188	4531	1344	448	ATA	TAA	424
<i>trnT (tgt)</i>	–	4956	5025	70				498
<i>cox2</i>	–	5524	6171	648	216	ATA	TAA	203
<i>nad3</i>	–	6375	6743	369	123	ATG	TAA	70
<i>trnV (tac)</i>	–	6814	6882	69				4
<i>trnY (gta)</i>	–	6887	6953	67				0
<i>trnP (tgg)</i>	+	6954	7022	69				9
<i>nad5</i>	–	7032	8687	1656	552	ATG	TAA	46
<i>cox3</i>	+	8734	9519	786	262	ATG	TAA	22
<i>trnK (ttt)</i>	+	9542	9611	70				17
<i>cob</i>	–	9629	10714	1086	362	ATT	TAA	56
<i>trnW (tca)</i>	–	10771	10837	67				3
<i>trnR (tcg)</i>	–	10841	10908	68				4
<i>trnN (gtt)</i>	–	10913	10984	72				4
<i>nad2</i>	+	10989	11990	1002	334	ATG	TAA	12
<i>nad1</i>	–	12003	12872	870	290	ATG	TAA	19
<i>trnD (gtc)</i>	–	12892	12954	63				3
<i>trnL2 (taa)</i>	+	12958	13025	68				4
<i>atp6</i>	–	13030	13731	702	234	ATA	TAA	222
<i>trnS2 (tga)</i>	–	13954	14020	67				260
<i>rrnL</i>	+	14281	15089	809				–13
<i>rrnS</i>	+	15077	15814	738				3
<i>trnM (cat)</i>	+	15818	15879	62				–4
<i>trnH (gtg)</i>	+	15876	15945	70				2
<i>trnQ (ttg)</i>	–	15948	16007	60				6
<i>trnG (tcc)</i>	+	16014	16079	66				8
<i>trnS1 (tct)</i>	+	16088	16151	64				10
<i>trnI (gat)</i>	+	16162	16231	70				–1
<i>trnF (gaa)</i>	–	16231	16296	66				10
<i>trnA (tgc)</i>	+	16307	16373	67				10
<i>trnL1 (tag)</i>	–	16384	16450	67				6
<i>trnE (ttc)</i>	–	16457	16522	66				6
<i>trnC (gca)</i>	+	16529	16601	73				18

Table 3. Organisation of the *Archaphanostoma ylvae* 16.6 kb mitochondrial genome.

between tRNAs, ranging in length from eleven to 277 base pairs. In addition, three long non-coding regions of 788, 1143 and 717 base pairs are found at the start of our sequence; between *trnL1* and *trnN*; and *trnN* and *rrnL*. The A + T content of these three sections are 68.78%, 65.79% and 76.15% respectively. The compositional difference between the 717 base pair non-coding region and the rest of the genome is statistically different ($\chi^2 = 25.629$, $p < 0.0001$), with a higher A + T content indicating that it could function as a transcriptional control region. The A + T content of the 788 base pair non-coding region is not significantly higher than the rest of the sequence ($\chi^2 = 0.85$). Similarly, there is a large portion of intergenic, non-coding sequence in the *A. ylvae* genome. Eight regions of non-coding sequence greater than 100 base pairs are found throughout the genome, with 24 additional smaller intergenic regions, ranging in size from 3 to 70 base pairs. Of the larger non-coding sequences, three have an A + T content that is statistically higher than the entire sequence: 309 nucleotides between *cox1* and *nad4l* ($\chi^2 = 3.944$, $p < 0.1$); 126 nucleotides between *nad4l* and *nad6* ($\chi^2 = 6.964$, $p < 0.01$) and 260 nucleotides between *trnS2* and *rrnL* ($\chi^2 = 9.654$, $p < 0.01$). The *P. rubra* sequence has just one longer non-coding intergenic sequence, of 196 base pairs.

As was already indicated by the 9.7 kb published partial genome²⁶, the gene arrangement we found in *P. rubra* is unique amongst published metazoan mitochondrial genomes. Similarly, neither *I. pulchra* nor *A. ylvae* show any similarity to other published metazoan mitochondrial genomes (Fig. 6). The species analysed in this study share only the small 'block' of *nad3-atp6-nad4-cob* (*I. pulchra*) and *cob-nad4-nad3* (*P. rubra*). However, the order is reversed between the two, and the genes are distributed across both strands in *I. pulchra*, so it is unlikely that this represents a feature inherited from a common ancestor. To quantify the number of common gene arrangements between the species in this study and other mitochondrial genomes, protein-coding gene and ribosomal

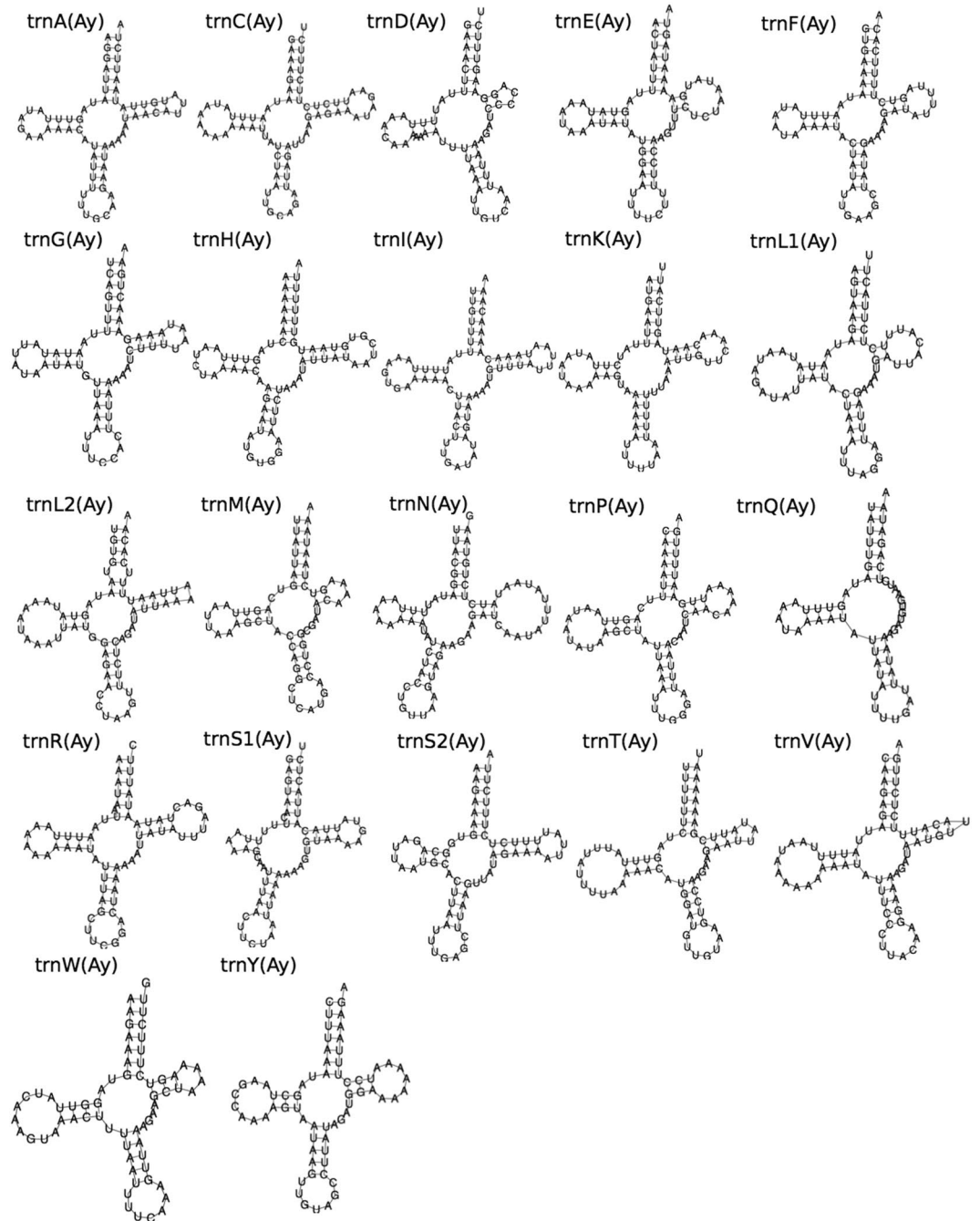


Figure 5. Predicted secondary structure of tRNAs from the mitochondrial genome sequence of *Archaphanostoma ylvae* as predicted by MiTFi in Mitos.

RNA gene order was analysed using CREx²⁹ (compared to the acoel *S. roscoffensis*, the xenoturbellid *Xenoturbella bocki*, and the metazoan mitochondrial 'ground plan', represented by *L. polyphemus*). Conserved mitochondrial gene 'blocks' (that is, a series of genes, regardless of their order within the grouping) were very infrequent between the species. Of the genomes compared, the highest number of common gene blocks was found between *X. bocki* and *P. rubra*. This result was not significant, finding only 16 common intervals out of a possible 150, and confirming the visual observation that gene order between these species is highly variable.

Phylogenetic analysis, and population differentiation. We used our new mitochondrial data from *P. rubra*, *I. pulchra*, and *A. ylvae* to investigate the internal phylogeny of the acoels and to test support for an Acoela-Xenoturbellida affinity. We observed that including the fast-evolving tunicates into our phylogeny leads to a clustering of these species and the acoels in an artificial long-branched clade (Supplementary Figure S1). Removing the tunicates, Bayesian phylogenetic inference was carried out using the protein-coding genes of *P. rubra*, *I. pulchra* and *A. ylvae* on a trimmed concatenated amino acid alignment, including the additional species

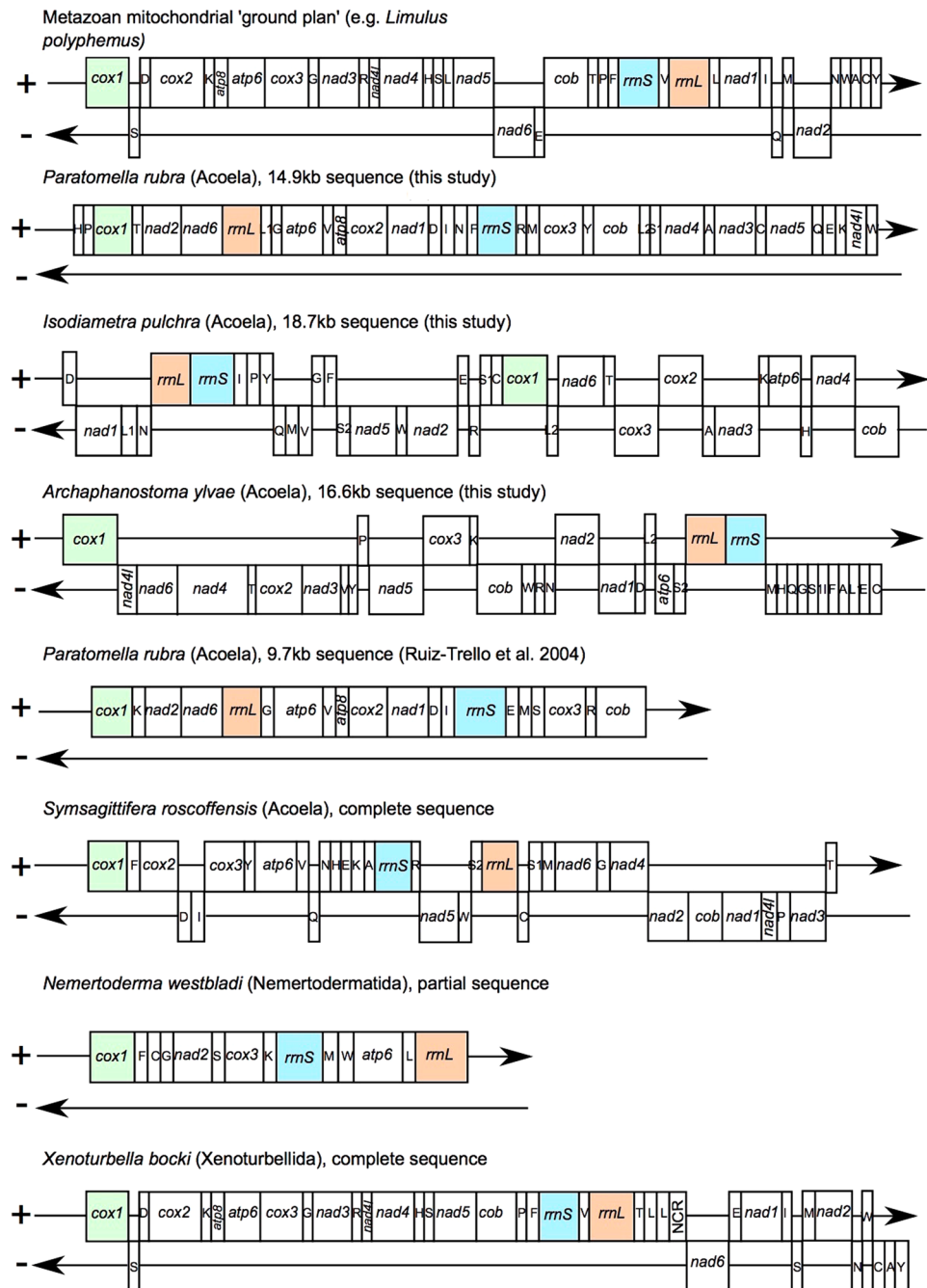


Figure 6. Comparisons of gene orders in the mitochondrial genome sequences resolved for *Paratomella rubra*, *Isodiametra pulchra* and *Archaphanostoma ylvae* compared to a published *P. rubra* fragment; the acoel *Symsagittifera roscoffensis*; the xenoturbellid *Xenoturbella bocki*; the nemertodermatid *Nemertoderma westbladi* and the metazoan mitochondrial 'ground plan' gene order, represented by *Limulus polyphemus*. Genes are not drawn to scale. Coloured genes chosen to show 'anchors' and divergence from the ground plan gene order in other species.

listed in Supplementary Table S2. The data set reached a MaxDiff of 0.17 after 39,525 trees were sampled across 10 chains discarding the first 400 trees (per chain) as burnin and sampling every 10th tree. In both, this and the maximum likelihood approaches, the protostome/deuterostome split was correctly inferred and Xenacoelomorpha were found splitting off inside Deuterostomia. *P. rubra*, *I. pulchra*, and *A. ylvae* were grouped inside Acoela, as expected (Fig. 7).

We detected the Nemertodermatida species *Nemertoderma westbladi* inside Mollusca and we therefore conclude that the limited data available on Genbank for this species is a contamination.

Having access to the published partial *P. rubra* mitochondrial sequence from a population sampled near Barcelona (Spain) and our own samples from Yorkshire, UK, we could estimate total sequence divergence and

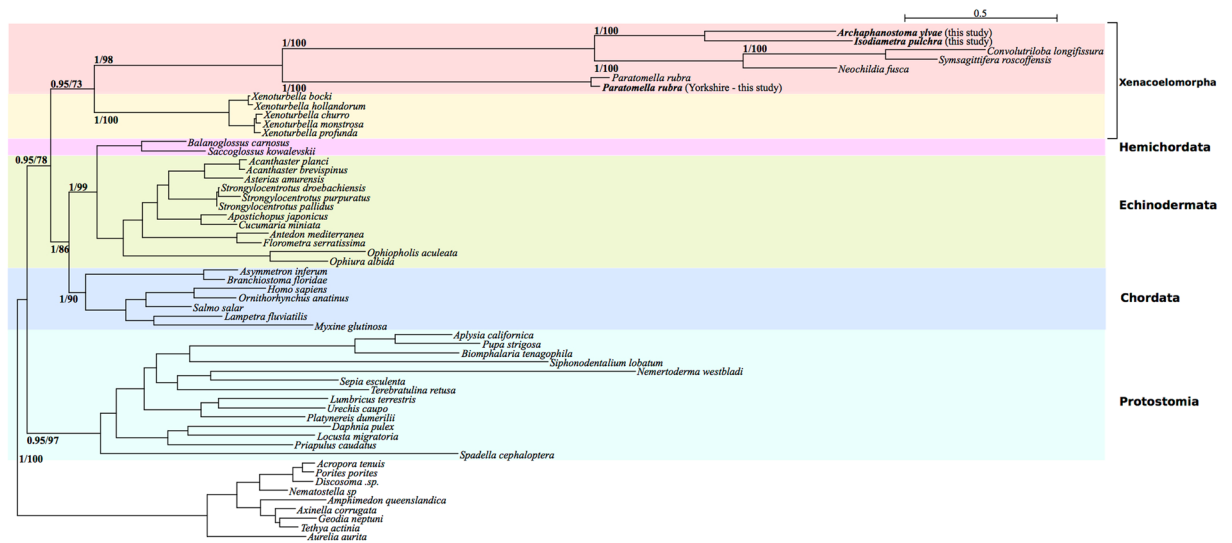


Figure 7. Bayesian (using PhyloBayes⁵³) and Maximum Likelihood (using RAxML⁵¹) phylogenetic analysis of mitochondrial protein-coding genes from the Metazoa, including *P. rubra*, *I. pulchra* and *A. ylvae* with posterior probability and bootstrap support values, respectively, at relevant nodes. Analysis carried out on trimmed alignment. Topology of both trees is identical.

compare non-synonymous to synonymous substitutions in eight protein-coding genes found on the Spanish fragment to the same genes from the Yorkshire mitochondrial genome. We found the 9.7 kb sequences to be 82.62% similar at the nucleotide level. The number of substitutions varied between, for example, 23 in the shortest gene alignment (*atp8*; 177 bp), to 161 in *nad2* (972 bp), and 116 in the 1401 bp long *cox1* alignment (Supplementary Table S3). Notably, non-synonymous substitutions appear to be frequent with, for example, 13 in *atp8*, 104 in *nad2*, and 25 in *cox1* (see Supplementary Table S3). Similarity of the *cox1* sequences on the nucleotide level is only 91% over 666 bp, thus higher when compared to species pairs (Supplementary Table S4), but lower than the 95–98% percent threshold used to distinguish species in *cox1* based barcoding³⁰.

Discussion

The 14.9 kb sequence of the *P. rubra* mitochondrial genome determined in our analysis contains the full complement of 37 genes typical of metazoan mitochondrial DNA. Numerous lab-based and computational efforts to close the circular mitochondrial genome were unsuccessful, but the complete gene complement and length of our final sequence indicates that this fragment covers the majority of the *P. rubra* complete mitochondrial genome. The difficulty we encountered in attempting to close the circular mitochondrial sequence may be attributed to the AT-rich, repetitive sequence found at both ends of the fragment, which could have prevented PCR amplification. Similar regions have been shown as problematic in studies of other mitochondrial genomes³¹. As no long stretch of non-coding sequence was found for this species in our study, the missing sequence might represent its mitochondrial transcription control region. Nonetheless, the overall AT content of the *P. rubra* mitochondrial sequence (78.15%) is high even for mtDNA, and greater than the A + T content of the mitochondrial genome of the acoel *S. roscoffensis* (75.3%)¹¹ and the published partial *P. rubra* genome (76.4%)²⁶.

The validity of the duplicated sequence found in our analysis of the *I. pulchra* mitochondrial genome could not be confirmed by PCR or computational efforts to map short reads to resolve it. Duplications within mitochondrial genomes are not uncommon, and changes to mitochondrial gene order are widely thought to arise as a result of a sequence ‘duplication and deletion’ mechanism^{14, 32, 33}. A number of mitochondrial genomes with duplicated sequences have been reported in species with a divergent mitochondrial gene order^{33–35}. Given the highly unusual gene order of the *I. pulchra* mitochondrial genome, a genomic duplication such as this could provide evidence for a genomic ‘duplication and deletion’ rearrangement of genes. The rearrangement and separation of protein-coding genes in other mitochondrial genomes has been attributed to long, non-tandem, inverted repeats³⁴, and this could be true for *I. pulchra*. Furthermore, very long nematode mitochondrial genomes with variable duplicated regions have been found with a conserved region containing the majority of the protein-coding genes³⁶: in *I. pulchra* the protein-coding genes and tRNAs, with the exception of *nad1* and *trnD* and *L1* are found in one long region, outside of the duplicated section. However, long non-coding duplications are frequently adjacent to tRNAs or other sequences capable of forming stem-and-loop structures³⁷. This is not true for the potential duplicate in *I. pulchra*. Most puzzling, both occurrences of the duplicate are identical, nucleotide-by-nucleotide, and unless the duplication occurred exceptionally recently, it is likely that spontaneous mutations would result in differences between the two copies of the sequence, especially given the elevated mutation rate of mitochondrial genomes. While it is true that the duplicated sequences appear at the start and end point of transcriptome assembly contigs, meaning it is possible that the duplication observed occurred only as a result of a sequencing and assembly error, their existence is nevertheless supported by PCR products which show an identical sequence being adjacent to both *rrnL* and to *cob*.

	<i>Isodiametra pulchra</i>	<i>Paratomella rubra</i>	<i>Symsagittifera roscoffensis</i>	<i>Archaphanostoma ylvae</i>
<i>cox1</i>	1536	1563	1551	1539
<i>cox2</i>	615	663	741	648
<i>cox3</i>	798	786	792	786
<i>nad1</i>	881	1053	870	870
<i>nad2</i>	1053	1014	990	1002
<i>nad3</i>	378	390	393	369
<i>nad4</i>	1344	1326	1350	1344
<i>nad4l</i>	absent	309	270	285
<i>nad5</i>	1710	1752	1776	1656
<i>nad6</i>	476	462	480	468
<i>cob</i>	1134	1083	1161	1086
<i>atp6</i>	681	609	702	702
<i>atp8</i>	absent	177	absent	absent

Table 4. Length of protein-coding genes in acoel mitochondrial genomes. All gene lengths in base pairs.

The 14.9 kb mitochondrial genome of *P. rubra*, the 18.7 kb sequence from *I. pulchra* and the complete 16.6 kb *A. ylvae* mitochondrial genome show no significant organisational similarity to any other published metazoan mitochondrial genome (Fig. 6). Comparison of the 14.9 kb *P. rubra* sequence with the published 9.7 kb *P. rubra* fragment shows an identical protein-coding and ribosomal gene order, but with variation in tRNA order (Fig. 6). tRNAs are reported to show much more frequent gene translocation compared to larger genes³⁸, which could account for these discrepancies. This, and the differences on the nucleotide level, including the relatively low level of similarity in the *cox1* barcoding gene might indicate that *P. rubra* collected from Barcelona (Spain)²⁶ and our animals from Yorkshire (England) should be regarded as cryptic species and not just divergent populations. Given the large and mostly unresolved diversity in benthic communities³⁹, and the marine environment in general⁴⁰, a differentiation into (cryptic) species cannot be seen as surprising.

All species analysed in this study are unique in the orientations and orders of their genes: *P. rubra* has genes exclusively on one strand; *I. pulchra* has an almost-equal distribution of genes across both the plus (18 genes) and minus (17 genes) strand; with *cox1* in a 'forward' orientation at the start of the genome, the majority of the protein-coding genes for *A. ylvae* are found on the minus strand. Furthermore, genes in *I. pulchra* and *A. ylvae* are not clustered into groups of 'gene blocks' on the same strand, but are found frequently as one or two genes on each strand. The finding of a unique gene order for these species seems to be typical for the acoels: analysis of the complete *S. roscoffensis* mitochondrial genome found no gene order similarity to any other species published to date, suggesting great variability in mitochondrial gene order amongst the acoels¹¹.

In addition to an unusual gene order, the mitochondrial genome of *P. rubra* shows frequent overlaps between protein-coding genes and tRNAs. tRNAs have been reported within protein-coding genes in other metazoan mitochondrial genomes^{41,42}, and given that no other location could be predicted for these sequences, this overlap could represent the simultaneous coding for both tRNAs and protein-coding genes. Overlap in coding sequence could be the result of selection to reduce genome size, accompanied by a reduction in non-coding sequence⁴², and truncated tRNAs with incomplete secondary structure⁴³, both of which are also found for the *P. rubra* sequence. The opposite is true for the *I. pulchra* and *A. ylvae* sequences. For *I. pulchra*, the sequence we could confidently verify makes the minimal length of the *I. pulchra* mitochondrial genome 18,725 base pairs, and it is likely to be longer in the complete closed circular genome. As has been found for other 'long' mitochondrial genomes, this is largely due to a large portion of the genome being non-coding⁴⁴. The lengths of protein-coding genes inferred for *I. pulchra* are similar to those of other acoel species (Table 4), and in addition, two protein-coding genes (*atp8* and *nad4l*) appear to have been lost from the genome, contributing to a reduced proportion of protein-coding gene sequence within the genome. The loss of *atp8* is not unusual, and has been reported in a number of unrelated taxa, as well as *S. roscoffensis* and in our *A. ylvae* mitochondrial genome¹¹. The absence of *nad4l* is more unusual, and could be a result of its existence in a portion of the genome that we have been unable to sequence. Although non-coding sequence contributes a relatively large proportion of the *A. ylvae* mitochondrial genome (17.17% compared to 22.72% in the *I. pulchra* sequence), the total genome is not exceptionally long.

The internal phylogeny we resolve for Acoela is in line with that proposed by Jondelius *et al.*²⁵. *I. pulchra* and *A. ylvae* group together in the family Isodiametridae. Isodiametridae groups with *S. roscoffensis*, *Neochildia fusca* and *Convolutriloba longifissura*, which are all members of the Convolutidae. *P. rubra* forms a separate branch outside the Convolutidae, representing the Paratomellidae. We interpreted the initial grouping of the acoels and tunicates as a classical example of long branch attraction (LBA) (Supplementary Figure S1). The accelerated substitution rates in mitochondrial DNA are also evidenced by the cryptic divergence we find in *P. rubra*, and may well lead to LBA in phylogenies derived from mitochondrial protein-coding genes, owing to the clustering of rapidly evolving lineages. This is of particular relevance for acoel species, which already demonstrate a very rapid rate of nucleotide substitution compared to other metazoans, leaving them vulnerable to LBA. Excluding the urochordates, we do retrieve Xenacoelomorpha as a branch of the deuterostomes, as expected (Fig. 7).

The mitochondrial genome sequences we analysed for the acoel species *P. rubra*, *I. pulchra* and *A. ylvae* have very divergent gene orders compared to other metazoan species. Furthermore, these species have very different

mitochondrial features: a large amount of genomic overlap in *P. rubra*, and a lot of non-coding sequence in *I. pulchra* and *A. ylva*. It is also possible that the mitochondrial genome of *I. pulchra* has a non-tandem inverted duplication - which could provide a mechanism for gene order variation - but this could not be confirmed by lab or computational based methods. Although limited to four species, the uniqueness of acoel mitochondrial genomes analysed so far^{11, 26} means that gene order and other mitochondrial genome features may not be phylogenetically informative for this order, although further mitochondrial genomes from other members of the Acoela would no doubt aid in this comparative analysis. Similarly, the cryptic divergence found between *P. rubra* samples from Yorkshire and Barcelona illustrate the usefulness of studying mitochondrial genomes to understand hidden species diversity. Our data clearly emphasise the still problematic placing of Xenacoelomorpha, with the clade firmly placed inside deuterostomes, but LBA drawing the acoels towards the outgroup when the tunicates were included. In summary, more data from genomes of early branching taxa are needed to resolve phylogenetic and biological questions.

Methods

Specimen collection, DNA extraction and PCR. Live *Paratomella rubra* specimens were isolated from sand samples collected from Filey, North Yorkshire and were immediately frozen and stored at -70°C following identification. Specimens of *Isodiametra pulchra* were cultured in petri dishes with nutrient-enriched f/2 sea water and fed *ad libitum* on *Nitzschia curvilineata* diatoms. DNA was extracted from live specimens of *I. pulchra* and frozen specimens of *P. rubra* using the QIAamp DNA Micro Kit (Qiagen: Product No. 56304) with the manufacturer recommended protocol.

All PCRs were done using the GeneAmp PCR System 2700 (Applied Bioscience). PCRs were carried out using the Expand Long-Range PCR Kit (Roche Applied Sciences: Product No. 11681834001), following manufacturer recommendations for 50 μl reaction set-up. General cycling protocol was: 92°C for 2 min; 15 cycles of: 92°C for 10 sec, 57°C for 15 sec, 68°C at initial elongation time (approximated as 1 min per 1000 base pairs to be amplified); 2 cycles each of: 92°C for 10 sec, 57°C for 15 sec, 68°C at 40 sec longer than initial elongation time, repeated at increasing 40 sec intervals for a further 14 cycles; a final elongation stage at 68°C for 7 min and a 4°C 'hold' stage. Where PCRs were not successful using this protocol, they were repeated using the Q5 High-Fidelity PCR Kit (New England Biolabs: Product No. E0555L), following manufacturer recommendations for a 25 μl reaction. Cycling protocol was: 92°C for 2 min; 40 cycles of: 92°C for 10 sec, 58°C for 15 sec, 68°C at initial elongation time (approximated as 1 min per 1000 base pairs to be amplified); a final elongation stage at 68°C for 7 min. Amplified products were visualised on ethidium-bromide stained TAE 0.8% gels. Bands of expected size were purified using the High Pure PCR Product Purification Kit (Roche Applied Sciences: Product No. 11732668001) and sent for sequencing by Source BioScience Life Sciences. Only amplifications that resulted in one clear band on the TAE 0.8% agarose gel were sequenced.

Three fragments of sequence from the mitochondrial genome of *P. rubra*, of size ~ 5.8 kb, ~ 4 kb, and ~ 1.2 kb, were generated from our gDNA assembly. Fragments were verified using a translated nucleotide query blast with invertebrate codon usage (blastx NCBI), and their orientation determined by gene annotation in comparison to the published 9.7 kb section of the *P. rubra* mitochondrial genome²⁶. Primers were designed in conserved gene regions to:

(1) Amplify across the 'N-stretches' present in the 5.8 kb and 4 kb fragments (8 and 9 N-stretches respectively, all of arbitrary length 50 base pairs).

(2) Cover the whole 1.2 kb fragment, with the aim of resolving the two frameshift mutations within the assembled sequence.

(3) Close the circular mitochondrial genome, by joining the 5.8 kb fragment to the 1.2 kb fragment; the 1.2 kb fragment to the 4 kb fragment; and the 4 kb fragment to the 5.8 kb fragment (see Supplementary Figure S5). Amplification of the fragments joining the 1.2 kb fragment to the 4 kb fragment and to close the mitochondrial genome using standard PCR cycling were unsuccessful. These were repeated using a touchdown protocol with Expand Long-Range polymerase. Annealing temperature was set at 65°C with decreasing 2°C intervals every 2 cycles down to 49°C . Initial elongation time was calculated as before, increasing 30 sec every two cycles of the touchdown, with a final 6 cycles at 49°C . This successfully amplified the region joining the 1.2 kb fragment to the 4 kb fragment, but we could not close the circular genome. Design of three new forward and reverse primers, tried in all combinations and using variable PCR parameters were unsuccessful in closing the mitochondrial genome. Additional RNA-Seq and DNA genomic sequencing data corroborated the stretches of sequence at either end of the mitochondrial genome but did not aid in closing the circle.

Three mitochondrial contigs of size ~ 13 kb, ~ 3.5 kb and ~ 1.3 kb were identified from *I. pulchra* Trinity transcriptome assembly from total RNA sequencing. A further contig of ~ 19 kb was also identified, covering the entire ~ 1.3 and 13 kb regions, and ~ 2.4 kb of the 3.5 kb sequence. Fragments were verified using blastx, NCBI, as outlined for *P. rubra*, and approximations for the location of protein-coding genes and tRNAs determined using the MITOS mitochondrial genome annotation server⁴⁵ (<http://mitos.bioinf.uni-leipzig.de/help.py>). Primers were designed to span the 13 kb contig in two ~ 5 kb sections, and to join the 13 kb contig to the 3.5 kb contig in both directions, to close the mitochondrial genome and check the validity of the duplicated region (Fig. 3). RNA-Seq data for *I. pulchra* were mapped to the long transcriptome assembly contigs and PCR sequencing results using NextGenMap⁴⁶, and visualised using Tablet⁴⁷.

We accidentally co-sequenced *A. ylva* at very low coverage as part of a *P. rubra* genome sequencing experiment. From an initial paired end assembly of Illumina HiSeq data with the CLC assembly cell software (v.5.0) we extracted the full mitochondrial circle of *A. ylva* in a single contig identified with BLAST. This was then annotated using MITOS and manual refinement as described above.

Genome annotation. For *P. rubra*, all sequenced fragments were aligned against the initial scaffold of the 9.7 kb published sequence²⁶; the 5.8 kb, 4 kb and 1.2 kb genome assembly sequences; and an additional long genome assembly fragment of length 14,954 (see Supplementary Figure S5). All contigs and PCR sequencing results were similarly aligned for *I. pulchra*, but without a reference sequence (see Fig. 3). Alignments were visualised using Mesquite (<http://mesquiteproject.org>) with invertebrate mitochondrial translated amino acid state colour coding. Where ambiguity remained between PCR sequencing results and genome or transcriptome assembly fragments, the genome or transcriptome assembly nucleotide sequence was used to establish a final ‘consensus’ sequence for each mitochondrial genome. This was of particular relevance for repetitive AT regions – for example, within *P. rubra nad1*, for which PCR sequencing results were inconclusive. In the case of *I. pulchra*, where the validity of the duplicated sections could not be confidently determined, we resolved a consensus sequence of 18,725 base pairs (Fig. 3).

The region for each protein-coding mitochondrial gene (*nad1–6*, *nad4l*, *cox1–3*, *cob*, *atp6* and *atp8*) in the *P. rubra*, *I. pulchra* and *A. ylvae* sequences were compared against published mitochondrial genomes using a translated nucleotide query (blastx, NCBI) with NCBI translation table number 5 ‘invertebrate mitochondrial’. Published genes from the mitochondrial genomes of the acoels *Symsagittifera roscoffensis* and *P. rubra* were downloaded from the NCBI Nucleotide database and aligned to the new consensus gene sequences of both *P. rubra*, *I. pulchra* and *A. ylvae* to verify the location of protein-coding and ribosomal RNA-encoding genes. The 5′ end of protein-coding genes were inferred to start from the first in-frame start codon (ATN, GTG, TTG, or GTT), even if this appeared to overlap with the preceding gene. Similarly, the terminal of protein-coding genes was inferred to be the first in-frame stop codon (TAA, TAG, or TGA). If no stop codon was present, a truncated stop-codon (T- or TA-) prior to the beginning of the next gene was assumed to be the termination codon, completed by post-transcription polyadenylation. tRNA sequences and putative secondary structures were identified using the Mitf program within MITOS.

Sequence alignment, phylogenetic analysis, and evolutionary rates. Phylogenetic analysis was performed using a concatenated amino acid alignment of all thirteen protein-coding genes for *P. rubra*, and all eleven protein-coding genes present in *I. pulchra*. Nucleotide sequences from all three acoel taxa and an additional set of species comprising 54 metazoans, taken from a range of published metazoan mitochondrial genomes representing deuterostomes, protostomes, cnidarians, and two species of poriferans as an outgroup (Supplementary Table S2) were aligned using TranslatorX (<http://www.translatorx.co.uk/>) independently for all genes with the appropriate mitochondrial genetic code set for each taxon included, using ClustalOmega⁴⁸ for amino acid alignment (Supplementary Table S2). Protein alignments were reduced to the most informative residues using trimAl v.1.4.rev15⁴⁹ with standard settings. Regions showing ambiguity in alignment were excluded, so that only blocks of well-aligned sequence were included for analysis.

We initially re-constructed neighbour nets in SplitsTrees v.4⁵⁰ to screen our dataset for potentially non-tree-like patterns, which could impede our phylogenetic analysis. Subsequently, we used RAXML⁵¹ v. 8.2.9 to infer maximum likelihood phylogenies from the original and the reduced alignments under the MTZOA model⁵². Bootstrapping was conducted employing the ‘autoMRE’ option in RAXML and the trees visualised with figtree v.1.4 (<http://tree.bio.ed.ac.uk/software/figtree>). We carried out Bayesian inference on the same trimmed alignment with PhyloBayes v.4.1⁵³ under the MTZOA model. We ran 10 chains in parallel and stopped the tree search at ~3950 trees per chain, with a maximum difference of 0.17, when discarding 400 trees as burnin and sampling every 10th tree per chain.

We used the Geneious software (v.8) to calculate sequence differences between a 9.7 kb section of the *P. rubra* mitochondrial genome originating from worms sampled in Filey, Yorkshire (UK) and Barcelona (Spain)²⁶. For eight protein-coding genes found on this section we used ParaAT (v.2.0)⁵⁴ to calculate translation alignments and the KaKs calculator (v1.2)⁵⁵ to access substitution rates. Also using Geneious we calculated a difference matrix for the *cox1* barcoding gene of the two *P. rubra* populations in comparison to acoel *coI* sequences retrieved from Genbank.

References

- Bourlat, S. J. & Hejnol, A. Acoels. *Curr Biol* **19**, R279–280, doi:10.1016/j.cub.2009.02.045 (2009).
- Ruiz-Trillo, I., Riutort, M., Littlewood, D. T., Herniou, E. A. & Bagaña, J. Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. *Science* **283**, 1919–1923, doi:10.1126/science.283.5409.1919 (1999).
- Katayama, T., Yamamoto, M., Wada, H. & Satoh, N. Phylogenetic position of Acoel turbellarians inferred from partial 18S rDNA sequences. *Zoolog Sci* **10**, 529–536 (1993).
- Philippe, H., Brinkmann, H., Martinez, P., Riutort, M. & Bagaña, J. Acoel flatworms are not platyhelminthes: evidence from phylogenomics. *PLoS One* **2**, e717, doi:10.1371/journal.pone.0000717 (2007).
- Lundin, K. The epidermal ciliary rootlets of *Xenoturbella bocki* (Xenoturbellida) revisited: new support for a possible kinship with the Acoelomorpha (Platyhelminthes). *Zoologica Scripta* **27**, 8–270, doi:10.1111/zsc.1998.27.issue-3 (1998).
- Raikova, O. I., Reuter, M., Jondelius, U. & Gustafsson, M. K. The brain of the Nemertodermatida (Platyhelminthes) as revealed by anti-5HT and anti-FMRamide immunostainings. *Tissue Cell* **32**, 358–365, doi:10.1054/tice.2000.0121 (2000).
- Hejnol, A. *et al.* Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Biol Sci* **276**, 4261–4270, doi:10.1098/rspb.2009.0896 (2009).
- Cannon, J. T. *et al.* Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89–93, doi:10.1038/nature16520 (2016).
- Philippe, H. *et al.* Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature* **470**, 255–258, doi:10.1038/nature09676 (2011).
- Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res* **27**, 1767–1780, doi:10.1093/nar/27.8.1767 (1999).
- Mwinyi, A. *et al.* The phylogenetic position of Acoela as revealed by the complete mitochondrial genome of *Symsagittifera roscoffensis*. *BMC Evol Biol* **10**, 309, doi:10.1186/1471-2148-10-309 (2010).
- Telford, M. J., Herniou, E. A., Russell, R. B. & Littlewood, D. T. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc Natl Acad Sci USA* **97**, 11359–11364, doi:10.1073/pnas.97.21.11359 (2000).

13. Saccone, C., De Giorgi, C., Gissi, C., Pesole, G. & Reyes, A. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* **238**, 195–209, doi:[10.1016/S0378-1119\(99\)00270-X](https://doi.org/10.1016/S0378-1119(99)00270-X) (1999).
14. Boore, J. L. & Brown, W. M. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev* **8**, 668–674, doi:[10.1016/S0959-437X\(98\)80035-X](https://doi.org/10.1016/S0959-437X(98)80035-X) (1998).
15. Rouse, G. W., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. New deep-sea species of Xenoturbella and the position of Xenacoelomorpha. *Nature* **530**, 94–97, doi:[10.1038/nature16545](https://doi.org/10.1038/nature16545) (2016).
16. Perseke, M. *et al.* The mitochondrial DNA of Xenoturbella bocki: genomic architecture and phylogenetic analysis. *Theory in Biosciences* **126**, [10.1007/s12064-007-0007-7](https://doi.org/10.1007/s12064-007-0007-7) (2007).
17. Bourlat, S. J., Rota-Stabelli, O., Lanfear, R. & Telford, M. J. The mitochondrial genome structure of Xenoturbella bocki (phylum Xenoturbellida) is ancestral within the deuterostomes. *BMC Evolutionary Biology* **9**, 107, doi:[10.1186/1471-2148-9-107](https://doi.org/10.1186/1471-2148-9-107) (2009).
18. Moritz, C. & Brown, W. M. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proc Natl Acad Sci USA* **84**, 7183–7187, doi:[10.1073/pnas.84.20.7183](https://doi.org/10.1073/pnas.84.20.7183) (1987).
19. Boore, J. L., Lavrov, D. V. & Brown, W. M. Gene translocation links insects and crustaceans. *Nature* **392**, 667–668, doi:[10.1038/33577](https://doi.org/10.1038/33577) (1998).
20. Crezee, M. Paratomella rubra, Rieger and Ott, an amphiatlantic acael turbellarian. *Cahiers De Biologie Marine* **19**, 1–9 (1978).
21. Rieger, R. & Ott, J. Gezeitenbedingte Wanderungen von Turbellarien und Nematoden eines nordadriatischen Sandstrandes. *Vie Milieu (Suppl.)* **22**, 425–447 (1971).
22. Achatz, J. G. & Martinez, P. The nervous system of Isodiametra pulchra (Acoela) with a discussion on the neuroanatomy of the Xenacoelomorpha and its evolutionary implications. *Front Zool* **9**, 27, doi:[10.1186/1742-9994-9-27](https://doi.org/10.1186/1742-9994-9-27) (2012).
23. Känneby, T., Bernvi, D. C. & Jondelius, U. Distribution, delimitation and description of species of Archaphanostoma (Acoela). *Zoologica Scripta* **44**, 218–231, doi:[10.1111/zsc.12092](https://doi.org/10.1111/zsc.12092) (2015).
24. Achatz, J. G., Chiodin, M., Salvenmoser, W., Tyler, S. & Martinez, P. The Acoela: on their kind and kinships, especially with nemertodermatids and xenoturbellids (Bilateria incertae sedis). *Organisms, Diversity & Evolution* **13**, 267–286, doi:[10.1007/s13127-012-0112-4](https://doi.org/10.1007/s13127-012-0112-4) (2013).
25. Jondelius, U., Wallberg, A., Hooge, M. & Raikova, O. I. How the Worm Got its Pharynx: Phylogeny, Classification and Bayesian Assessment of Character Evolution in Acoela. *Systematic Biology* **60**, 845–871, doi:[10.1093/sysbio/syr073](https://doi.org/10.1093/sysbio/syr073) (2011).
26. Ruiz-Trillo, I., Riutort, M., Fourcade, H. M., Baguna, J. & Boore, J. L. Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyly of the Platyhelminthes. *Mol Phylogenet Evol* **33**, [10.1016/j.ympev.2004.06.002](https://doi.org/10.1016/j.ympev.2004.06.002) (2004).
27. De Mulder, K. *et al.* Characterization of the stem cell system of the acael Isodiametra pulchra. *BMC Developmental Biology* **9**, 1–17, doi:[10.1186/1471-213x-9-69](https://doi.org/10.1186/1471-213x-9-69) (2009).
28. Egger, B. *et al.* To be or not to be a flatworm: the acael controversy. *Plos One* **4** (2009).
29. Bernt, M. *et al.* CREx: inferring genomic rearrangements based on common intervals. *Bioinformatics* **23**, 2957–2958, doi:[10.1093/bioinformatics/btm468](https://doi.org/10.1093/bioinformatics/btm468) (2007).
30. Hebert, P. D. N., Ratnasingham, S. & de Waard, J. R. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**, S96–S99, doi:[10.1098/rsbl.2003.0025](https://doi.org/10.1098/rsbl.2003.0025) (2003).
31. Sakai, M. & Sakaizumi, M. The complete mitochondrial genome of Dugesia japonica (Platyhelminthes; order Tricladida). *Zoolog Sci* **29**, 672–680, doi:[10.2108/zsj.29.672](https://doi.org/10.2108/zsj.29.672) (2012).
32. Moritz, C., Dowling, T. E. & Brown, W. M. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* **18**, [10.1146/annurev.es.18.110187.001413](https://doi.org/10.1146/annurev.es.18.110187.001413) (1987).
33. Braband, A., Podsiadlowski, L., Cameron, S. L., Daniels, S. & Mayer, G. Extensive duplication events account for multiple control regions and pseudo-genes in the mitochondrial genome of the velvet worm Metaperipatus inae (Onychophora, Peripatopsidae). *Molecular Phylogenetics and Evolution* **57**, 293–300, doi:[10.1016/j.ympev.2010.05.012](https://doi.org/10.1016/j.ympev.2010.05.012) (2010).
34. Hyman, B. C., Beck, J. L. & Weiss, K. C. Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. *Genetics* **120**, 707–712 (1988).
35. Zhou, X., Lin, Q., Fang, W. & Chen, X. The complete mitochondrial genomes of sixteen ardeid birds revealing the evolutionary process of the gene rearrangements. *BMC Genomics* **15**, 573, doi:[10.1186/1471-2164-15-573](https://doi.org/10.1186/1471-2164-15-573) (2014).
36. Hyman, B. C., Lewis, S. C., Tang, S. & Wu, Z. Rampant gene rearrangement and haplotype hypervariation among nematode mitochondrial genomes. *Genetica* **139**, 611–615, doi:[10.1007/s10709-010-9531-3](https://doi.org/10.1007/s10709-010-9531-3) (2011).
37. Stanton, D. J., Daehler, L. L., Moritz, C. C. & Brown, W. M. Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. *Genetics* **137**, 233–241 (1994).
38. Bermudez-Santana, C. *et al.* Genomic organization of eukaryotic tRNAs. *BMC Genomics* **11**, 1–14, doi:[10.1186/1471-2164-11-270](https://doi.org/10.1186/1471-2164-11-270) (2010).
39. Fonseca, V. G. *et al.* Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications* **1**, 98, doi:[10.1038/ncomms1095](https://doi.org/10.1038/ncomms1095) (2010).
40. Schiffer, P. H. & Herbig, H.-G. Endorsing Darwin: global biogeography of the epipelagic goose barnacles Lepas spp. (Cirripedia, Lepadomorpha) proves cryptic speciation. *Zoological Journal of the Linnean Society* **177**, 507–525, doi:[10.1111/zoj.12373](https://doi.org/10.1111/zoj.12373) (2016).
41. Morrison, D. A. How and where to look for tRNAs in Metazoan mitochondrial genomes, and what you might find when you get there. *arXiv.org* (2010).
42. He, Y., Jones, J., Armstrong, M., Lamberti, F. & Moens, M. The Mitochondrial Genome of Xiphinema americanum sensu stricto (Nematoda: Enoplea): Considerable Economization in the Length and Structural Features of Encoded Genes. *Journal of Molecular Evolution* **61**, 819–833, doi:[10.1007/s00239-005-0102-7](https://doi.org/10.1007/s00239-005-0102-7) (2005).
43. Doublet, V. *et al.* Large gene overlaps and tRNA processing in the compact mitochondrial genome of the crustacean Armadillidium vulgare. *RNA Biology* **12**, 1159–1168, doi:[10.1080/15476286.2015.1090078](https://doi.org/10.1080/15476286.2015.1090078) (2015).
44. Minxiao, W., Song, S., Chaolun, L. & Xin, S. Distinctive mitochondrial genome of Calanoid copepod Calanus sinicus with multiple large non-coding regions and reshuffled gene order: Useful molecular markers for phylogenetic and population studies. *BMC Genomics* **12**, 73, doi:[10.1186/1471-2164-12-73](https://doi.org/10.1186/1471-2164-12-73) (2011).
45. Bernt, M. *et al.* MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* **69**, 313–319, doi:[10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023) (2013).
46. Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791, doi:[10.1093/bioinformatics/btt468](https://doi.org/10.1093/bioinformatics/btt468) (2013).
47. Milne, I. *et al.* Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics* **14**, 193–202, doi:[10.1093/bib/bbs012](https://doi.org/10.1093/bib/bbs012) (2012).
48. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539–539, doi:[10.1038/msb.2011.75](https://doi.org/10.1038/msb.2011.75) (2011).
49. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973, doi:[10.1093/bioinformatics/btp348](https://doi.org/10.1093/bioinformatics/btp348) (2009).
50. Huson, D. H. & Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* **23**, 254–267, doi:[10.1093/molbev/msj030](https://doi.org/10.1093/molbev/msj030) (2006).
51. Stamatakis, A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, doi:[10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) (2014).

52. Rota-Stabelli, O., Yang, Z. & Telford, M. J. MtZoa: A general mitochondrial amino acid substitutions model for animal evolutionary studies. *Molecular Phylogenetics and Evolution* (2009).
53. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288, doi:[10.1093/bioinformatics/btp368](https://doi.org/10.1093/bioinformatics/btp368) (2009).
54. Zhang, Z. *et al.* ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochemical and Biophysical Research Communications* **419**, 779–781, doi:[10.1016/j.bbrc.2012.02.101](https://doi.org/10.1016/j.bbrc.2012.02.101) (2012).
55. Zhang, Z. *et al.* KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging. *Genomics, Proteomics & Bioinformatics* **4**, 259–263, doi:[10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2) (2006).

Acknowledgements

This work was supported by the European Research Council (ERC-2012-AdG 322790), the Leverhulme Trust (grant F/07 134/DA) and the Biotechnology and Biological Sciences Research Council (grant BBS/H006966/1). MJT was supported by a Royal Society Wolfson Research Merit Award. We would like to thank Richard Copley (The Wellcome Trust Centre for Human Genetics, University of Oxford) and the Oxford Wellcome Trust for their assistance with *P. rubra* sequencing, We are indebted to Anne Zakrzewski for her assistance with animal collection trips, associated animal identification, and valuable comments on the manuscript.

Author Contributions

Conceived and designed the experiments: H.E.R., F.L., B.E. Performed the experiments: H.E.R., B.E., F.L., P.H.S. Analysed the data: H.E.R., P.H.S., M.J.T. Wrote the paper: H.E.R., M.J.T., P.H.S.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-01608-4](https://doi.org/10.1038/s41598-017-01608-4)

Competing Interests: The authors declare that they have no competing interests.

Accession codes: The mitochondrial sequences for *P. rubra*, *A. ylvae* and *I. pulchra* have been submitted to NCBI GenBank under Accession Numbers KY825222, KY825223 and KY825224, respectively.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017