Genomics 111 (2019) 343–355

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Probing recalcitrant problems in polyclad evolution and systematics with novel mitochondrial genome resources

Nathan J. Kenny[a], Carolina Noreña[b], Cristina Damborenea[c], Cristina Grande[d,*]

[a] Department of Life Sciences, The Natural History Museum of London, Cromwell Road, London SW7 5BD, UK
[b] Museo Nacional de Ciencias Naturales (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain
[c] División Zoologia Invertebrados, Museo de La Plata, CONICET, Argentina
[d] Departamento de Biologia, Facultad de Ciencias, Universidad Autonoma de Madrid, Cantoblanco, 28049 Madrid, Spain

ARTICLE INFO

Keywords:
Mitochondria
Phylogenetics
Platyhelminthes
Polycladida
Tricladida
Cotylea
Acotylea

ABSTRACT

For their apparent morphological simplicity, the Platyhelminthes or "flatworms" are a diverse clade found in a broad range of habitats. Their body plans have however made them difficult to robustly classify. Molecular evidence is only beginning to uncover the true evolutionary history of this clade. Here we present nine novel mitochondrial genomes from the still undersampled orders Polycladida and Rhabdocoela, assembled from short Illumina reads. In particular we present for the first time in the literature the mitochondrial sequence of a Rhabdocoel, *Bothromesostoma personatum* (Typhloplanidae, Mesostominae). The novel mitochondrial genomes examined generally contained the 36 genes expected in the Platyhelminthes, with all possessing 12 of the 13 protein-coding genes normally found in metazoan mitochondrial genomes (*ATP8* being absent from all Platyhelminth mtDNA sequenced to date), along with two ribosomal RNA genes. The majority presented possess 22 transfer RNA genes, and a single tRNA gene was absent from two of the nine assembled genomes. By comparison of mitochondrial gene order and phylogenetic analysis of the protein coding and ribosomal RNA genes contained within these sequences with those of previously sequenced species we are able to gain a firm molecular phylogeny for the inter-relationships within this clade.

Our phylogenetic reconstructions, using both nucleotide and amino acid sequences under several models and both Bayesian and Maximum Likelihood methods, strongly support the monophyly of Polycladida, and the monophyly of Acotylea and Cotylea within that clade. They also allow us to speculate on the early emergence of Macrostomida, the monophyly of a "Turbellarian-like" clade, the placement of Rhabditophora, and that of Platyhelminthes relative to the Lophotrochozoa (=Spiralia). The data presented here therefore represent a significant advance in our understanding of platyhelminth phylogeny, and will form the basis of a range of future research in the still-disputed classifications within this taxon.

1. Introduction

The Polycladida includes some of the most beautiful and fascinating invertebrates found on earth. This order of the Platyhelminthes consists of approximately 800 known species found in diverse marine environments, from coral reefs to deep-sea vents, with some freshwater species also known (e.g. *Limnostylochus borneensis* (Stummer-Traunfels 1902)). Some of the Polycladida have become well known for their vibrant colours, and almost all are predatory, generally preying on sessile invertebrates. This contrasts strongly with the lifestyle of the obligately parasitic Platyhelminthes that are the best-studied members of this Phylum.

How polyclads are related to other Platyhelminthes, and the internal phylogeny of this clade, has been subject to some recent debate, although some consensus is beginning to be reached regarding basic inter-relationships with the aid of molecular phylogenetic methods. Three classes of parasitic flatworms, Cestoda, Trematoda and Monogenea, are now known to form a monophyletic group, the Neodermata [8,16], which almost certainly evolved from a free-living ancestor [36]. Free-living flatworms, which were formerly grouped into

Abbreviations: atp6 and 8, ATPase subunits 6 and 8; bp, base pair(s); cob, cytochrome *b*; cox1–3, cytochrome *c* oxidase subunits I–III; kb, kilobase; LBA, long branch attraction; ML, maximum likelihood; mt, mitochondria(l); mtDNA, mitochondrial DNA; nad, NADH dehydrogenase subunit; NCBI, National Centre for Biotechnology Information; rRNA, ribosomal RNA; rrnS and rrnL, small and large subunit ribosomal RNA; tdrl, tandem-duplication-random-loss; tRNA, transfer RNA

* Corresponding author.
E-mail addresses: nathanjameskenny@gmail.com (N.J. Kenny), mcnnj92@mncn.csic.es (C. Noreña), cdambor@fcnym.unlp.edu.ar (C. Damborenea), cristina.grande@uam.es (C. Grande).

https://doi.org/10.1016/j.ygeno.2018.02.009
Received 27 October 2017; Received in revised form 13 February 2018; Accepted 16 February 2018
Available online 24 February 2018
0888-7543/

the paraphyletic "Turbellaria" (see [22]), have proven to be more problematic to classify [30].

Of the clades formerly incorporated into the Turbellaria (free-living flatworms), acoel flatworms and the Xenoturbellida are now generally positioned as one or several separate Phyla (e.g [11].). Within the Platyhelminthes, the Catenulida is reasonably well established as the earliest branching clade [15,31], sister to all other Platyhelminthes, which are grouped together in the Rhabditophora. Within the Rhabditophora relationships are complex. Two excellent recent analyses of these inter-relationships have been performed [15,31] allowing some broad conclusions can be made, although some contentious nodes remain to be resolved before platyhelminth phylogeny is fully understood. Macrostomorpha is highly likely to represent the sister taxon to all other Rhabditophoran species. Polycladids, along with either the Lecithoepitheliata or Prorhynchida, are then sister taxa to one another, and this clade is sister to the other Rhabditophora. Further rhabditophoran inter-relationships are still subject to some debate, and we refer the interested reader to Egger et al. [15] and Laumer et al. [31] for further comparisons.

Polyclad species were traditionally placed into one of two sub-orders, the Cotylea or Acotylea, based on the presence or absence of a sucker located posterior to the female genital pore, and the morphology of the reproductive and digestive systems was then used for formulating further taxonomic hypotheses (e.g [4,27].). However, a range of recent investigations has led to doubt concerning the monophyly of the Cotylean and Acotylean clades. While the presence of a sucker can be considered an apomorphy for the Cotylea, the lack of it cannot be considered an apomorphy for the Acotylea. This provokes a controversial situation where a deficiency of apomorphies results in the lack of demonstrable monophyly for Cotylea and Acotylea [4,17,18,39,43]. More recently, molecular analysis has been able to infer acotylean monophyly, but cotyleans are not always robustly supported as a monophyletic clade (e.g [5,42].). Other investigations have shown cotylean monophyly [1], but sampling has remained sparse compared to true Polyclad diversity, limiting the power of previous analyses.

Within the cotylean and acotylean clades, specific investigations have been performed to more discretely catalogue the internal phylogeny of these groups. Historically these hypotheses were based on morphological data, and this is still the primary tool for understanding the systematics of the Polycladida (e.g [41].). However, molecular data has been used on a limited basis to corroborate and extend these frameworks. Rawlinson et al. [42] performed comparative work across this clade, an analysis of the nuclear *28S rRNA* gene sequence of eight cotylean and six acotylean species, building on earlier work sampling more discretely within polyclad diversity (e.g [37]. (sampling the Pseudocerotidae), [53] (sampling the Cotylea)). In Rawlinson et al. [42] the Acotylea was shown to be monophyletic, but cotylean traits appeared to be paraphyletic, with the cotylean *Pericelis cata* indicated as the sister taxon to the Acotylea with good posterior probability but very poor bootstrap support under ML analysis. This made it difficult to discern whether the presence of a posterior sucker was a plesiomorphy of the Polycladida as a whole, secondarily lost in the Acotylea. However, that analysis did support some extant generic assignments within the Cotylea, and suggest that further sampling and a wider variety of molecular data was necessary to discern deeper level interrelationships in the Polycladid clade. Furthermore, most recently, analysis of *28S* rRNA data by Bahia et al. [5] has strongly suggested that the acotyleans *Cestoplana* and *Theama* were nested within the Cotylea. That work provides a robust structure for assessing polycladid relationships, and suggests that traditionally used morphological characters may be misleading. However, work based on a single locus can also be misleading, and it is useful to consider additional evidence from a number of molecular loci.

The advent of wider molecular data from a broader array of species has already been of some aid in establishing platyhelminth systematic relationships [36]. Mitochondrial genome sequences in particular have begun to be used to attempt to estimate polycladid and platyhelminth inter-relationships (see, for example, [1,47]). Mitochondrial genomes can be useful for phylogenetic inference, as they contain both faster and slower evolving sequence areas, are well conserved in arrangement and content between species, and as intact mitochondrial genomes can be assembled from next-generation sequencing reads, allowing for rapid generation of datasets for analysis. Presently, four complete polyclad mitochondrial genomes are available in the published record, with just under 20 other complete platyhelminth mtDNA sequences (excluding those described here for the first time).

Platyhelminth mitochondrial genomes are reasonably small (approximately 14 kb in length) in comparison to other metazoan mitochondria, and all sequenced to date seem to lack the *atp8* gene. With the exception of the catenulids [28], which use the standard invertebrate mitochondrial transcriptional code, all Platyhelminthes share a derived translational code, with several changes, including "TGA" representing tryptophan instead of "stop". Start/stop codons also seem relaxed, with alternate codons and possible frame shifting often substituting other codons in place of the more usual "ATG", "TAG" and "TAA" codes [1]. These changes seem universal in non-catenulid flatworms, although the still limited number of mitochondrial sequences available in this Phylum means that this is yet to be fully confirmed.

To extend our understanding of platyhelminth mitochondrial diversity, phylogeny and evolution, we present the mitochondrial sequences of 9 novel species of free-living flatworm, including for the first time analyses of the sequence of a Rhabdocoel, *Bothromesostoma personatum*. These have been drawn from samples from a range of geographic locations and a number of important polycladid clades (Fig. 1). We have used these to investigate the inter-relationships of the Polycladida in particular and the Platyhelminthes in general, and this information will be of utility for ongoing efforts in untangling the true phylogeny of this fascinating clade.

2. Material and methods

2.1. Acquisition of samples and sequencing of DNA

Specimens were collected from several localities (Fig. 1, Table 1). DNA was extracted from samples following a standard phenol-chloroform extraction. Samples were sent to AllGenetics (Coruña, Spain) for sequencing. There they were quantified using fluometric methods (Qubit, Life Technologies) and the Illumina Nextera DNA Sample Prep Kit was used to prepare DNA samples for sequencing according to the standard protocol, with nominal library size of 300 bp. A Qubit HS DNA Assay was then used to check sample integrity before paired end sequencing was performed on the Illumina HiSeq 2000 platform at 101 bp read length, alongside other samples, with a nominal 1/6th or 1/7th of a total lane devoted to each sample. Reads were then assigned to species on the basis of their respective library indices (Table 2).

2.2. Read cleaning, assembly and identification of mitochondrial sequences

Initial assays of sequencing quality were performed using FastQC [3]. To remove library indices and adapter sequence, and as poor read quality was observed in some samples, Trimmomatic [9] was run using the following settings: ILLUMINACLIP: ./adapter.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 (where adapter.fa was a file containing the sequence of the specific adapter used in sequencing). This resulted in smaller libraries of good sequencing quality for assembly. Assembly was performed using Velvet [52] with an initial *k* mer size of 61, −min_contig_lgth 100 and -cov_cutoff 3. After initial assembly, TBLASTN (−db_gencode 9) was used with mitochondrial protein sequences of known homology to assay approximate mitochondrial genome coverage within individual species, by examining the coverage of the best blast hits, which is reported in Velvet fasta

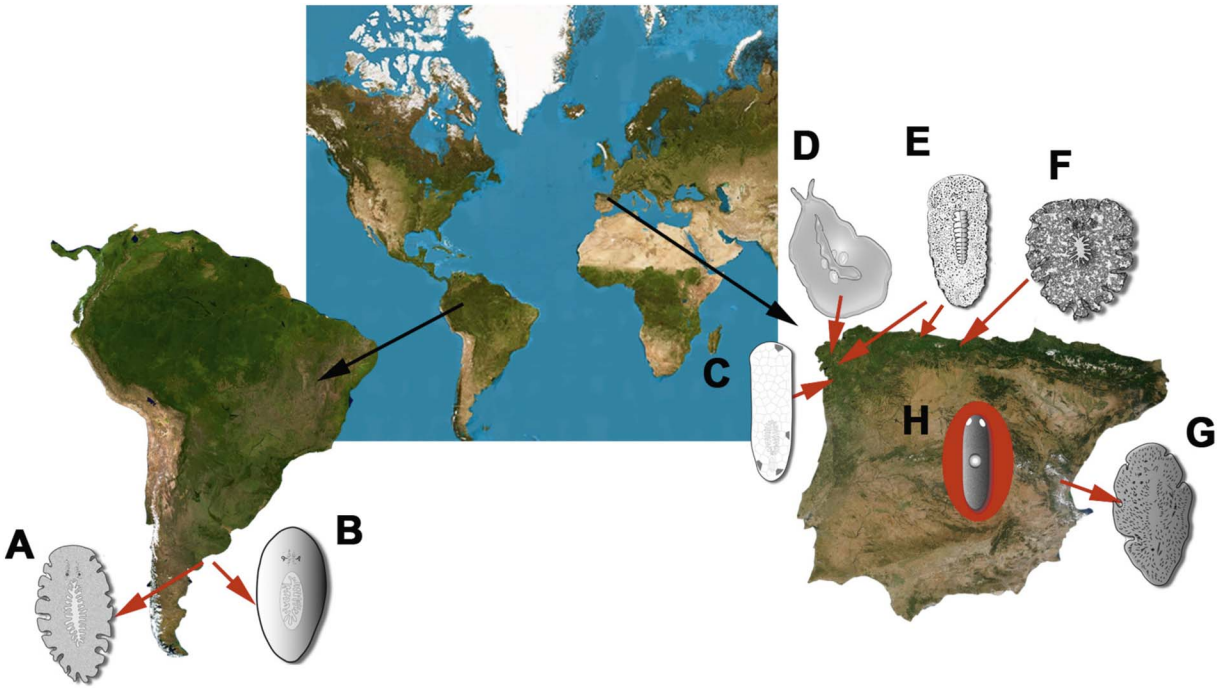

Fig. 1. Images of specimens sequenced for the first time in this study, and a map indicating the source of these samples (South America, left, and the Iberian Peninsula, right, enlarged). A: *Crassiplana albatrossi*; B: *Notocomplana palta*; C: *Cryptocelis alba*; D: *Imogine fafai*; E: *Discocelis tigrina* (Galicia at left, Asturias at right); F: *Bothromesostoma personatum*; G: *Imogine stellae*; H: *Eurylepta cornuta*.

Table 1
Identity and Accession Numbers of Samples Used in Analysis.

	Novel	Species	Classification:	Accession #:
Platyhelminthes:		*Taenia saginata*	Platyhelminthes; Cestoda; Eucestoda; Cyclophyllidea; Taeniidae	NC_009938
		Diphyllobothrium latum	Platyhelminthes; Cestoda; Eucestoda; Diphyllobothriidea; Diphyllobothriidae	AB269325
		Benedenia hoshinai	Platyhelminthes; Monogenea; Monopisthocotylea; Capsalidae	NC_014591
		Gyrodactylus derjavinoides	Platyhelminthes; Monogenea; Monopisthocotylea; Gyrodactylidae	NC_010976
		Microcotyle sebastis	Platyhelminthes; Monogenea; Polyopisthocotylea; Microcotylidae	NC_009055
		Clonorchis sinensis	Platyhelminthes; Trematoda; Digenea; Opisthorchiida; Opisthorchiata; Opisthorchiidae	NC_012147
		Schistosoma japonicum	Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae	HM120848
		Schistosoma mansoni	Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae	NC_002545
		Dugesia japonica	Platyhelminthes; Rhabditophora; Seriata; Tricladida; Continenticola; Geoplanoidea; Dugesiidae	NC_016439
		Schmidtea mediterranea	Platyhelminthes; Rhabditophora; Neophora; Tricladida; Continenticola; Geoplanoidea; Dugesiidae	JX398125
		Girardia sp.	Platyhelminthes; Rhabditophora; Neophora; Tricladida; Continenticola; Geoplanoidea; Dugesiidae	KP090061
	*	*Bothromesostoma personatum*	Platyhelminthes; Rhabditophora; Rhabdocoela; Typhloplanidae	MF993329
		Microstomum lineare	Platyhelminthes; Rhabditophora; Macrostomorpha; Macrostomida; Microstomidae	AY228756
	*	*Crassiplana albatrosi*	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Stylochoidea; Callioplanidae	MF993330
	*	*Imogine stellae*	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Stylochoidea; Stylochidae	MF993336
	*	*Imogine fafai*	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Stylochoidea; Stylochidae	MF993335
	*	*Discocelis tigrina* (Galicia)	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Ilyplanoidea; Discocelidae	MF993333
	*	*Discocelis tigrina* (Asturias)	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Ilyplanoidea; Discocelidae	MF993332
	*	*Cryptocelis alba*	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Leptoplanoidea; Cryptocelidae	MF993331
	*	*Notocomplana palta*	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Leptoplanoidea; Notoplanidae	MF993337
		Stylochoplana maculata	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Leptoplanoidea; Stylochoplanidae	KP965863
		Hoploplana elisabelloi	Platyhelminthes; Rhabditophora; Polycladida; Acotylea; Leptoplanoidea; Leptoplanidae	KT363735
		Enchiridium sp.	Platyhelminthes; Rhabditophora; Polycladida; Cotylea; Euryleptoidea; Prosthiostomidae	KT363734
		Prosthiostomum siphunculus	Platyhelminthes; Rhabditophora; Polycladida; Cotylea; Euryleptoidea; Prosthiostomidae	KT363736
	*	*Eurylepta cornuta*	Platyhelminthes; Rhabditophora; Polycladida; Cotylea; Euryleptoidea; Eurylepidae	MF993334
Outgroups:		*Roboastra europaea*	Mollusca; Gastropoda; Heterobranchia; Euthyneura; Nudipleura; Nudibranchia; Doridina; Anadoridoidea; Polyceridae	NC_004321
		Katharina tunicata	Mollusca; Polyplacophora; Neoloricata; Chitonida; Acanthochitonina; Mopaliidae	NC_001636
		Lumbricus terrestris	Annelida; Clitellata; Oligochaeta; Haplotaxida; Lumbricina; Lumbricidae; Lumbricinae	NC_001673
		Platynereis dumerilii	Annelida; Polychaeta; Palpata; Aciculata; Phyllodocida; Nereididae	AF178678
		Rotaria rotatoria	Rotifera; Bdelloidea; Philodinida; Philodinidae	NC_013568
		Philodina citrina	Rotifera; Bdelloidea; Philodinida; Philodinidae	NC_019806
		Pallisentis celatus	Acanthocephala; Eoacanthocephala; Gyracanthocephala; Quadrigyridae	NC_022921
		Macracanthorhynchus hirudinaceus	Acanthocephala; Archiacanthocephala; Oligacanthorhynchida; Oligacanthorhynchidae	NC_019808
		Gnathostomula armata	Gnathostomulida; Bursovaginoidea; Gnathostomulidae	KP965860
		Gnathostomula paradoxa	Gnathostomulida; Bursovaginoidea; Gnathostomulidae	KP965861
		Lepidodermella squamata	Gastrotricha; Chaetonida; Paucitubulatina; Chaetonidae	KP965862

Table 2
Sequencing statistics.

Species	Fraction, HiSeq lane	Library index	Number of read pairs (raw)	Number of Read Pairs (after filtering)	GC% (after cleaning)
Bothromesostoma personatum	1/7	CGTACTAG/TATCCTCT	35,027,807	17,793,260	38
Crassiplana albatrosi	1/6	TAAGGCGA/AGAGTAGA	24,830,194	15,926,205	43
Imogine stellae	1/6	TAAGGCGA/TATCCTCT	14,875,505	6,360,346	36
Imogine fafai	1/7	TCCTGAGC/AGAGTAGA	19,570,261	12,749,160	42
Discocelis tigrina (Galicia)	1/6	TAAGGCGA/AGAGTAGA	43,146,443	27,235,585	39
Discocelis tigrina (Asturias)	1/6	CGTACTAG/AGAGTAGA	25,719,396	17,854,794	40
Cryptocelis alba	1/7	TAAGGCGA/TATCCTCT	24,429,868	14,930,278	41
Notocomplana palta	1/6	CGTACTAG/TATCCTCT	43,464,112	22,936,663	41
Eurylepta cornuta	1/6	CGTACTAG/AGAGTAGA	47,500,054	32,640,518	39

headers. The -cov_cutoff setting was then revised upwards to exclude low coverage *k* mers from resulting assemblies, which, empirically, greatly increased contiguity of mitochondrial sequence. TBLASTN, again with known mitochondrial protein sequences and -db_gencode 9 setting, was used to identify mitochondrial sequences, which in almost all cases were fully assembled mitochondrial genomes. Overlapping sequences at either end of the contig were identified by homology and removed manually. In the cases where assemblies did not contain the entire mitochondrial sequence, the 2–3 contigs containing mitochondrial sequence were examined manually, and these were found to overlap by fewer than 60 bp at their ends. These overlapping areas were used for manual assembly of complete sequence, with gene order then compared with fully assembled sequence to confirm correct assembly, and overlapping sequence at ends removed when necessary.

2.3. Sequence and annotation

The MITOS webserver [6] was used for annotation, using the "09 - Echinoderm/Flatworm" code for translation. Manual curation was necessary for these annotations, as start and stop codons were not always correctly identified. When necessary, homology to known genes was used to inform start/stop location when canonical codons were not present. *OrganellarGenomeDRAW* [38] was used to generate the mitochondrial figure for *Bothromesostoma personatum* with the relative GC content at a given location represented on the inner circle.

2.4. Phylogenetic analysis

rrnL, *rrnS* and 12 mitochondrial protein nucleotide sequences, and amino acid sequences from the latter genes, were used for phylogenetic reconstruction of polyclad interrelationships under both maximum likelihood and Bayesian methods. Novel sequences were analyzed along with mitochondrial sequences from previously sequenced species (Table 1). Where the partial *Microstomum lineare* mitochondrial sequence contained a gene, two versions of alignments were made, one with and one without the *M. lineare* gene. Gene sequences were aligned, on an individual gene by gene basis, in MAFFT [23,24] using the FFT-NS-i model. Alignments were then individually assayed with Gblocks [12] under the three "relaxed" parameters, excluding excessively variable regions from further analysis. FASconCAT [26] was then used to combine the alignments, with one set of "with *Microstomum lineare*" alignments, and one "without *Microstomum lineare*" sequences. These had final lengths of 1388 residues (protein) and 6442 bp (nucleotide) for the "with *Microstomum lineare*" set (5/7 genes) and 2290 residues (protein) and 10,890 bp (nucleotide) for the "without *Microstomum lineare*" set (12/14 genes). Sequences and alignments can be found in Supplementary File 1.

Nucleotide and amino acid alignments were fed into jModelTest2 [14] and ProtTest 3.2 [13] respectively, to infer the best fit models of substitution for these datasets (GTR + I + G and MTArt + I + G + F) respectively, although MTzoa was substituted when it was found to outperform MTart empirically (ln likelihood MTArt: −90,141.25, MTzoa: −88,439.80, BIC, for the "without *M. lineare*" dataset). RAxML v. 8.2.3 [48] was used to perform Maximum Likelihood (ML) analyses, with 1000 bootstrap replicates under the rapid bootstrapping mode. For Bayesian Inference (BI), Phylobayes 4.1 [29] was run, with the CAT-GTR model used, and for BI both amino acid and nucleotide data four gamma categories, maximum discrepancy 0.1 and minimum effective size 100 were used. Following the run, 20% of sampled points were discarded by readpb as "burn-in", before the remaining samples were used to generate the figures shown here. Phylogenies were displayed in Figtree v1.4.3 [40], with the Bayesian topology displayed in figures used here. Where differences in topology were observed between BI and ML (in the nucleotide tree, with *M. lineare*, Fig. 6A) these nodes are shown with an asterisk, indicating a polytomy in the ML tree.

3. Results

3.1. Sequencing

Sequencing results for all samples are shown in Table 2. The number of reads recovered per sample was highly variable, with the most highly recovered sample, *Eurylepta cornuta*, possessing more than three times the number of paired end reads of the least-well sequenced sample (*Imogine stellae*). We speculate that it is a result of variation introduced in the course of loading the flow cell. The GC% of the samples was relatively stable in all cases, between 36 and 43%. While read quality as assessed using FastQC was reasonable in the forward read direction, read quality declined markedly in the reverse direction, with lower quartile read quality often dipping below a Phred score of 30 as early as the 60th base of the read. In some cases residual library index adapter contamination was also detected. We therefore filtered stringently for read quality and content to exclude poor quality and adapter sequence, and the number of reads left after filtering can be seen in Table 2. The read count remained highly variable between samples (with the best covered sample now 5-fold more well-covered than the least well covered sample), but read quality was markedly improved; with lower quartile Phred scores above 30 through to the final base in the retained reads.

3.2. Assembly and annotation

Assembly was performed using the filtered read pairs (unpaired reads remaining after filtering were not used, but could be utilized in future analyses). Mitochondrial genomes were assembled and identified as described in the methods, and basic statistics relating to these can be seen in Table 3. The disparity observed in read number did not affect the assembly of complete mtDNA sequence, with all assemblies resulting in easily identifiable complete, circular mtDNA sequences. Sizes are relatively consistent, varying from 14,783 bp to 15,909 bp.

Table 3
Assembly statistics for novel mitochondrial genomes, indicating GC% and non-coding portion, along with overlapping sequence.

	Size (bp)	GC%	Non-coding bp	Overlapping bases
Bothromesostoma personatum	15,017	25.9	2007	Total: 58, 7 cox1/trnF, 4 rrnS/trnT, 1 trnI/trnW, 11 nad2/trnR, 35 nad1/trnN
Crassiplana albatrosi	15,798	30.9	2364	Total: 46, 37 nad1/trnL2, 8 rrnS/trnN, 1 trnM/cox2
Imogine stellae	14,902	30	772	Total: 10, 1 trnG/rrnL, 8 rrnS/trnN, 1 trnM/cox2
Imogine fafai	15,064	30.7	1470	Total: 6, 5 rrnS/trnN, 1 trnM/cox2
Discocelis tigrina (Galicia)	14,783	31.3	897	Total: 15, 1 trnQ/trnK, 13 trnC/rrnS, 1 rrnS/trnN
Discocelis tigrina (Asturias)	14,872	31.3	893	Total: 15, 1 trnQ/trnK, 13 trnC/rrnS, 1 rrnS/trnN
Cryptocelis alba	15,268	32.9	1242	Total: 10, 8 rrnS/trnN, 1 trnP/nad3, 1 trnM/cox2
Notocomplana palta	14,934	29.9	1053	Total: 35, 1 trnL2/trnQ, 26 trnH/cob, 8 rrnS/trnN
Eurylepta cornuta	15,909	34.2	2036	Total: 100, 52 rrnL/trnL1, 6 trnL1/trnE, 5 atp6/nad1, 4 nad1/trnT, 5 cob/nad4L, 5 trnF/trnY, 23 trnC/rrnS

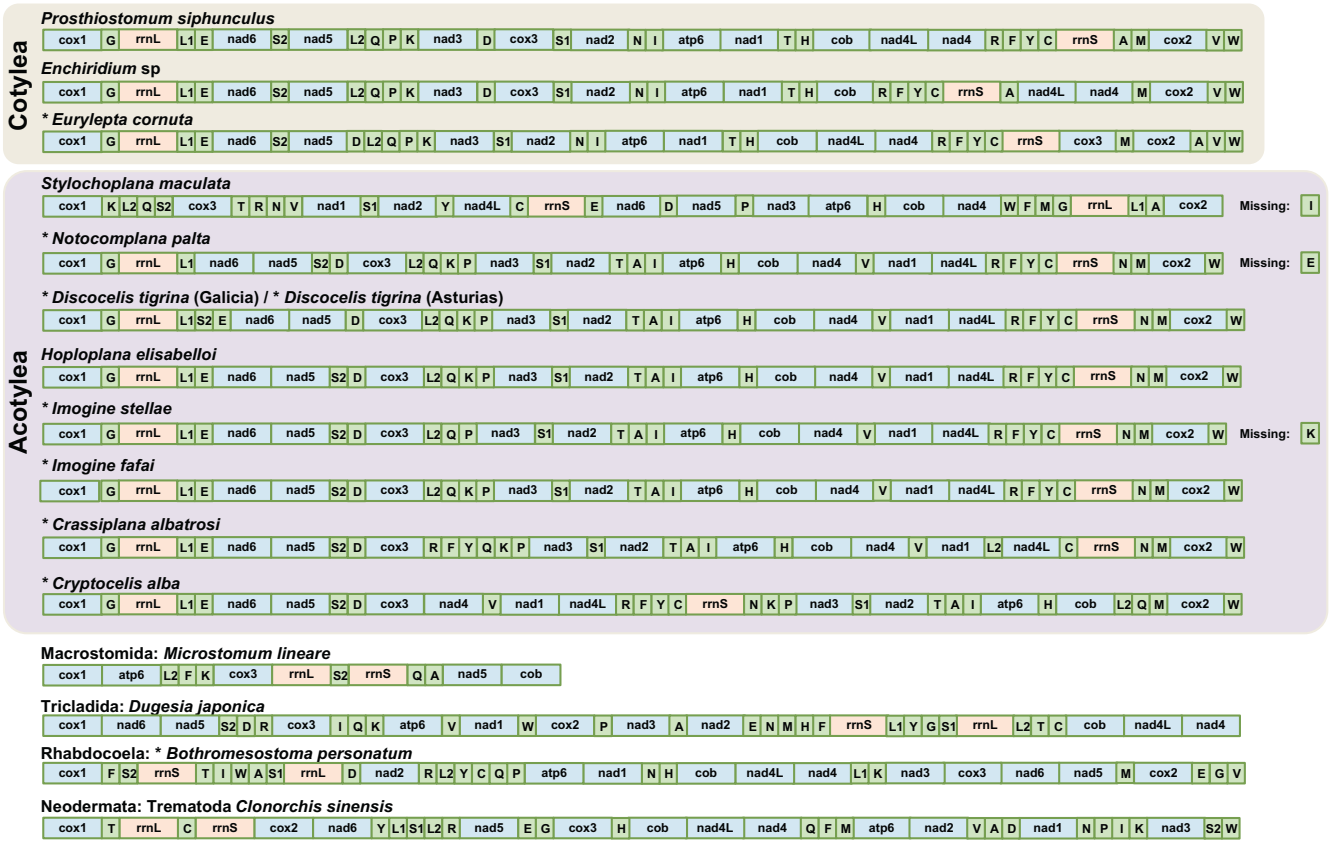

Fig. 2. Relative arrangement of mitochondrial gene order within the Platyhelminthes. Genes are colour coded according to their families – tRNA genes in green, rRNA in orange, cytochrome oxidase and reductase, NADH and ATP synthase genes in blue. Acotylean and Cotylean clades indicated by shaded boxes. Asterisks (*) indicate novel data presented here. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Annotation revealed that all mitochondrial genomes contained the expected complement of twelve protein coding genes, along with two ribosomal RNA genes and around 22 transfer RNA genes. A single tRNA gene was absent from two of the nine novel assembled genomes, *Notocomplana palta* and *Imogine stellae* as noted on Fig. 2 (please note, a tRNA in *Stylochoplana maculata* is also shown as absent in our results, but this sequence was recovered in [21]). Fig. 2 also shows the relative order of these genes within these mtDNA molecules, oriented with *cox1* as the first gene in all cases. The order of genes within the Cotylean and Acotylean clades is relatively consistent. In the Cotylea the relative location of *cox3*, *nad4L* and *nad4* varies from species to species. This will provide further evidence for clades within this group (for example, *Prosthiostomum siphunculus* and *Enchiridium* sp. sharing the placement of *cox3* relative to other genes, and therefore being more likely to be sister taxa). Across the Cotylea, some tRNA arrangements remain constant, even when protein coding genes move. *trnN* and *trnI*, *trnT* and *trnH*, and *trnV* and *trnW* appear in a constant arrangement relative to one another, not observed in the Acotylea.

In the Acotylea, a great variety of gene orders is observed, although the larger number of species sequenced here and in previous works mean than more data points are available in the Acotylea than the Cotylea. *S. maculata* differs from other acotylean species, with a unique arrangement of genes (*cox3*, *nad1*, *nad2*, *nad4L* and *rrnS*) following *cox1*, while all other Acotylea possess a stereotypical *cox1*, *rrnL*, *nad6*, *nad5* and *cox3* arrangement, generally followed by *nad3* (with the exception of *Cryptocelis alba*). Indeed, the sequences of the species of the genera *Discocelis*, *Hoploplana* and *Imogine* differ only in the arrangement of tRNA genes, particularly *trnS2*.

The direction of transcription was consistent in all sampled mtDNA sequences, with all genes (including tRNA genes) transcribed in the same "+" frame relative to one other. The GC% of the assembled genomes (Table 3) is much lower than that of the reads (Table 2).

The distribution of non-coding regions throughout the genome is reasonably consistent between all sampled mtDNA molecules, with the total number of non-coding bp varying between as much as 2364 in *Crassiplana albatrossi* and as little as 772 in *I. stellae* (Table 3). A small

Table 4
Start/stop codon usage in novel species examined here.

	B. personatum		*C. albatrosi*		*C. alba*		*D. tigrina Astur.*		*D. tigrina Gal.*		*E. cornuta*		*I. fafai*		*I. stellae*		*N. palta*	
	Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop	Start	Stop
atp6	ATG	TAA	TTG	TAA	GAT	TAA	ATG	TAA	ATG	TAA	ATG	TAA	TTG	TAA	TTG	TAA	TTG	TAA
cob	ATG	TAA	TCT	TAA	TCA	TAA	ATG	TTT	ATG	TTT	CGT	TAG	TCT	TAG	TCT	TAA	ATG	TAA
cox1	AAC	TAG	GCA	TAA	ATG*	TAA	ATG*	TAA	ATG*	TAA	ATG	TAA	ATG	TAG	GTG	TAA	ATG*	TAA
cox2	ATG	TAG	ATG	TAA	ATG	TAG	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAA	ATG	TAA	ATG	TAA
cox3	ATG	TAA	ATG	TTT	TTC	TAG	TTT	TAA	TTT	TAA	ATG	TAA	TTT	TAA	TTT	TAA	TTT	TAA
nad1	ATG	TAA	ATT	TAG	ATG	TAA	TTG	TAG	TTG	TAG	ATG	TAG	ATG	TAG	ATG	TAG	TTG	TAA
nad2	ATG	TAA	TTG	TCA	AGA	AGT	ATG	TAA	ATG	TCT	ATG	TAG	ATT	TAC	ATG	TAG	TTG	TCC
nad3	ATG	TAA	ATG	TGA	ATG	TAG	ATG	TAA	ATG	TAA	GTG	TAG	ATG	TAG	ATG	TAA	TTG	TAG
nad4	TGG	TAA	ATC	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATC	TAA	ATG	TAA	ATG	TAA	GTA	TAA
nad4l	ATG	TAA	TTT	TAA	GTG	TAG	ATG	TAA	ATG	TAA	ATG	TAG	TTG	TAA	ATG	TAA	TTG	TAA
nad5	ATA	TAA	CAA	TAA	ATG	TAG	AAA	TAA	AAA	TAA	ATG	TAA	AAC	TAA	AAT	TAA	TTA	TTA
nad6	GTG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAA	ATG	TAA	TTG	TTA

* Frameshift follows.

amount of overlapping sequence, from 6 to 100 bp in different species, was observed between some adjacent genes, with some genes showing evolutionarily conserved overlap (e.g. *rrnS/trnN*, overlapping by one bp in a variety of species seen here). Two specimens of *Discocelis tigrina* were sequenced in this work, and 112 differences at the nucleotide level were observed between them. These correlated to occasional differences at the protein coding level.

The start and stop codons used by platyhelminth species within mtDNA genomes are known to vary from the canonical ATG start and TAA/TGA/TAG stop sequences. In Table 4 the start/stop codons observed in the species sequenced here are presented. ATG and TAA are the most common codons for all novel species, and the "echinoderm and flatworm mitochondrial code" [50] seems robustly supported as appropriate for use with these species, with TGA used both to indicate "STOP" and tryptophan.

Alternative start codons were more numerous and diverse than alternative stops, with 45 predicted alternative start codons seen in the nine species examined here, and only 9 alternative stop codons. With only one exception (*C. alba nad2*, AGT) all alternate stop codons used began with "T", a trait seen in other polycladid species [1]. Of the 45 alternate start codons, TTG (13 occurrences) was the most commonly used. *N. palta* is the species that most frequently uses non-canonical codons, with 9/12 start codons differing from the standard ATG, and 3/12 alternate stop codons. As these were assigned on the basis of homology of sequence to known genes, definitive confirmation would require a proteomic study, but we are confident in our likely assignments, which are often shared in several species.

3.3. Bothromesostoma personatum mtDNA sequence

Particular attention was paid to the arrangement and contents of the *Bothromesostoma personatum* mitochondrial genome, as the first sequenced representative of the Rhabdocoela, to our knowledge, at the time of publication. The size and coding quantity (c.f. non-coding fraction) of this genome did not differ markedly from that of all other mitochondrial sequences examined in this manuscript. The arrangement of genes in the mitochondrial genome of *B. personatum* can be seen in Fig. 3. However, the GC% is particularly low (25.9% - the lowest observed in the novel genomes examined in this manuscript), and the arrangement is markedly different from that observed in previously investigated platyhelminth species and shown here. Small syntenic elements, such as a *cob*, *nad4L* and *nad4* group, are conserved between this species and other orders of Platyhelminthes, but the order of rhabdocoel mitochondrial genomes seems to distinctly differ.

The analysis of codon usage bias in this species (using [49]) found evidence of bias towards certain codons for any given amino acid. In all cases, codons ending in "T" were favoured over any other combination when a codon ending in 'T' existed, with the two exceptions being arginine, where CGA (22 uses) was favoured over CGT (16 uses), and leucine, where TTA was favoured (out of 6 options). For example, in proline, CCT is found in 47% of cases (of four options); in alanine, GCT is used by 63% of all codons coding for that amino acid; for cysteine, TGT is used 89% in of cases; and for phenylalanine, TTT is used 90% of the time. Further evidence, along with exact counts, can be found in Supplementary File 2.

No differences between *B. personatum* start codon usage and that of other species were observable (Table 4). However, no non-canonical stop codons are seen in the *B. personatum* mitochondrial genome, unlike polycladid species described in this manuscript and other Platyhelminthes described previously.

3.4. Phylogenetic analysis

Well-supported trees were reliably recovered by our analyses, although they slightly changed with and without the presence of additional data. When *M. lineare* was not included in analyses (due to its incomplete mitochondrial sequencing, which precludes analysis with a fully populated matrix for all genes) we gain highly congruent results. Figs. 4 and 5 show the results of phylogenetic analysis of nucleotide and amino acid alignments respectively. Both exhibit the same basic topology for platyhelminth clade relationships, the only difference being the relative position of some outgroup taxa, where the chiton *Katharina tunicata* is placed in a monophyletic molluscan clade in the nucleotide dataset tree, but in a monophyletic annelid + mollusc clade in the amino acid dataset.

Polycladida, Acotylea and Cotylea are recovered as monophyletic clades with the maximum possible support from bootstrapping and posterior probability. This is also observed for Tricladida and Neodermata. Within the Polycladida, clade support tends to be more robust under BI than ML analysis. The lowest posterior probability for BI in this clade is 0.96, whereas ML-derived bootstrap support values are sometimes much lower. This is particularly the case for our amino acid dataset, analyzed under the Mtzoa model. The GTR site-heterogenous model implemented for nucleotide analysis has higher support values for some nodes under ML analysis than those in the amino acid tree shown in Fig. 5. The Mtzoa matrix will not necessarily be representative of polyclad molecular evolution, and thus the GTR model used for nucleotides may have resulted in the higher ML support values for some nodes, as can be seen in Fig. 4. It should be noted, however, that the additional data provided by rRNA genes could also have had contributory effects to these support values.

Platyhelminthes is supported as a monophyletic clade distinct from the remainder of the lophotrochozoan (=spiralian) outgroups analyzed here. The existence of a clade of free living flatworm species, similar to

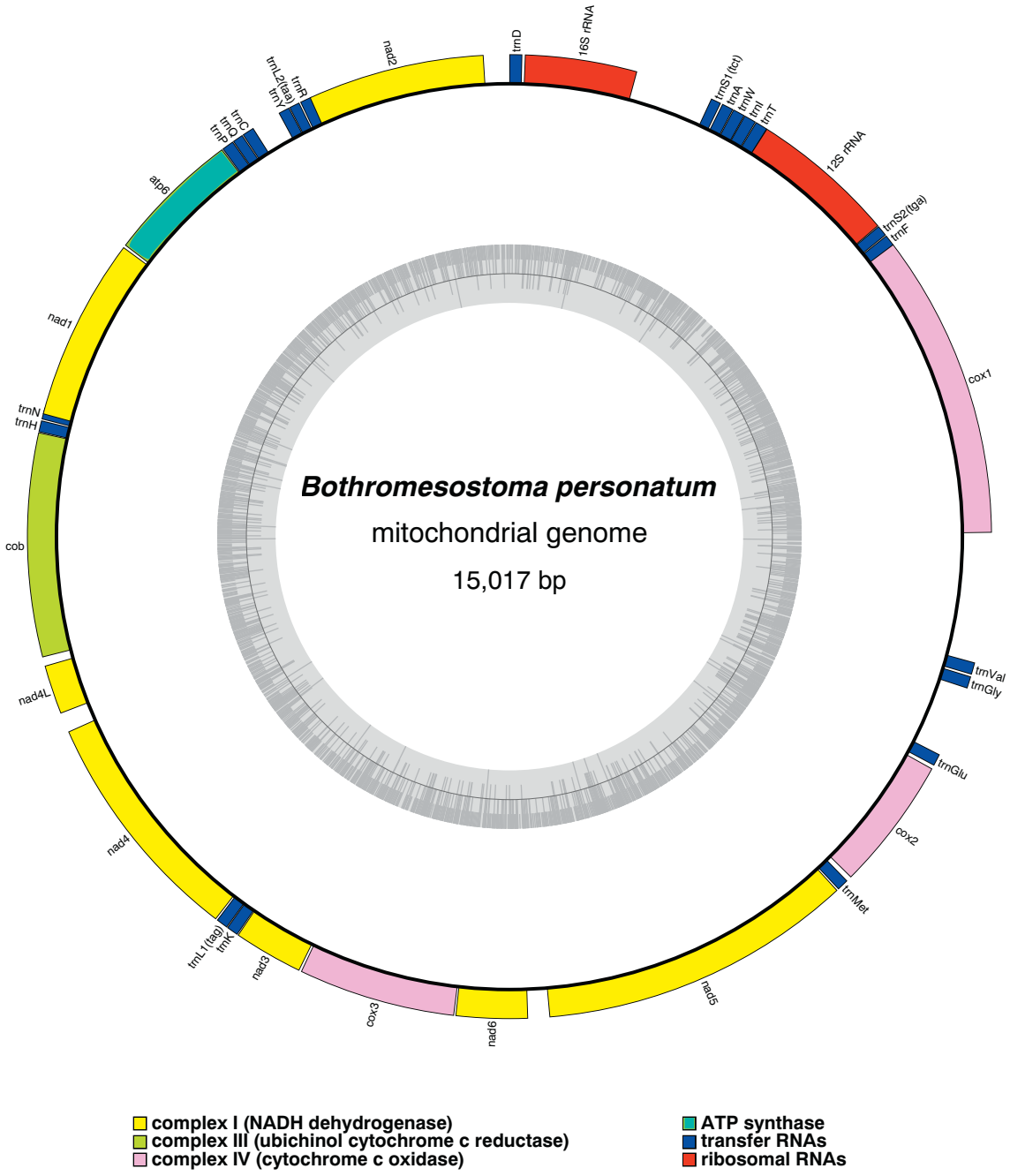

Fig. 3. The circular mitochondrial genome of *Bothromesostoma personatum* as represented by OrganellarGenomeDRAW [38]. Orientation of gene transcription represented by the outside circle - all genes transcribed in same frame. Local GC content (GC dark grey, AT light grey) represented on the inner ring. Colour key at base of figure indicates families of genes within the mitochondrial genome.

those previously gathered as the class "Turbellaria", which includes all flatworms except the Neodermata, is supported with almost maximal values under BI and ML analysis of nucleotide data (0.91/100 respectively) and with reasonable support for amino acid analyses (1/60). However, this is contentious, and discussed further in Sections 3.5 and 4.3.

Tricladida is well supported as a monophyletic clade, and Rhabdocoels are recovered as the sister taxa to the Tricladida by both nucleotide and amino acid based analyses, albeit with lower support values under some models. Only BI of amino acid datasets gives compelling support (posterior probability = 0.96), CAT-GTR, with ML bootstrap support less strong under both the Mtzoa model (35) and GTR model of nucleotide substitution (41). We find no evidence of a "Neophoora" clade (Rhabdocoela + Tricladida + Neodermata).

All Cotylea studied here belong to superfamily Euryleptoidea. Both *P. siphunculus* and *Enchiridium* sp. (Prosthiostomidae) form a monophyletic group in the analysis, with high levels of support in both nucleotide and amino acid analyses, and *E. cornuta* (Euryleptidae) is recovered as the sister taxa of them.

In the Acotylea, *C. alba*, the only member of the superfamily Cryptoceloidea Bahia 2017 presented here, appears to the sister group of the remaining acotyleans. Two further clades, representing the Stylochoidea and Leptoplanoidea, are well recovered in our analyses by multiple forms of evidence, confirming recent analysis [2,5]. The Leptoplanoidea includes members of the former Superfamily Ilyplanoidea (*D. tigrina*) with strong, almost maximal, support. The Stylochoidea are maximally well supported as a clade by BI, although ML analysis gives more ambivalent results in our amino acid tree.

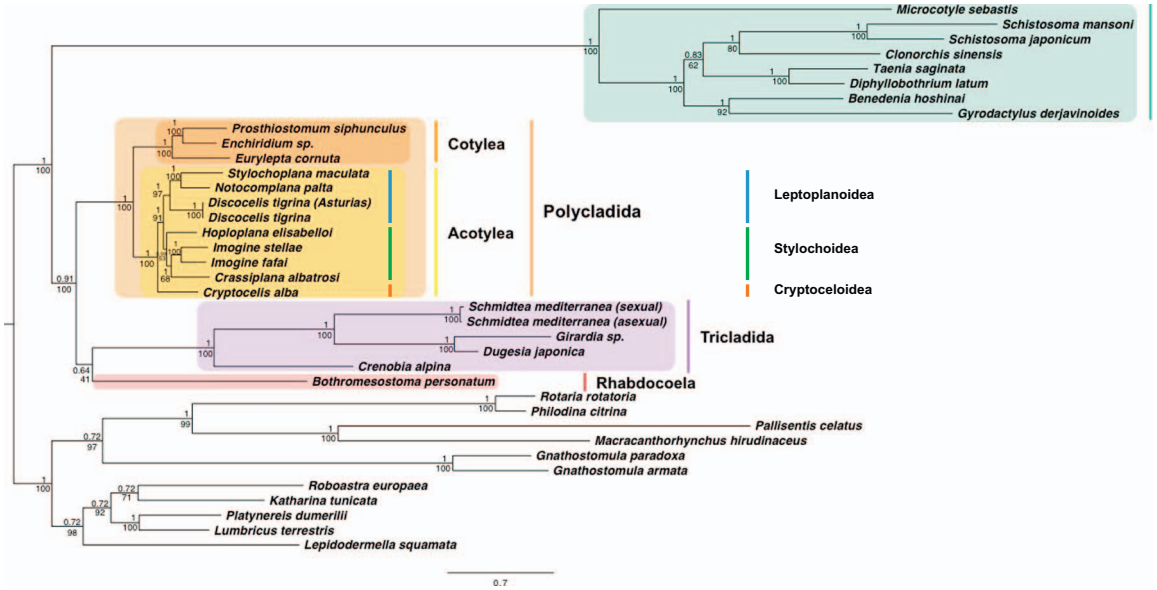

Fig. 4. Inter-relationships of Platyhelminthes and related taxa, based on Bayesian Inference (BI) and Maximum Likelihood analysis of nucleotide sequences for protein coding and rRNA genes within the mitochondrial genome of these species, with BI-derived tree displayed here. Orders of Platyhelminth are indicated with coloured boxes as marked on the Figure, along with the sub-orders Acotylea and Cotylea. Numbers at bases of nodes indicate posterior probability/bootstrap replicates (percentage, from 1000 replicates).

3.5. Phylogenetic analysis, incorporating the Macrostomorpha

The Macrostomorpha has only been sparsely sequenced to date, with the incomplete mitochondrial sequence of *M. lineare*, to our knowledge, the only known mtDNA resource available. However, the Macrostomorpha could be key for understanding the wider inter-relationships of the Platyhelminthes as a whole. We therefore repeated our analysis, including the partial sequence of *M. lineare* (5 protein coding genes, 2 rRNA genes) with the other sequences shown here. The results of this analysis can be seen in Fig. 6.

These trees are generally similar to those recovered by our "complete" analyses. In both amino acid and nucleotide-based analyses, Platyhelminthes is recovered as a monophyletic clade with good support from posterior probability/bootstrap results - not always found in previous analyses (e.g. Fig. 4, [1]).

The most interesting and important differences to note, however, are in the relative placement of the Rhabdocoela, the monophyly (or lack thereof) of free-living flatworms, and in the placement of *M. lineare* itself. Rhabdocoels are not recovered as the sister taxa of the Tricladida by amino acid analysis when *M. lineare* is present, nor is it strongly supported based on nucleotide evidence (Bayesian posterior probability for (Tricladida + *B. personatum*) = 0.6, ML support < 10). Instead, Rhabdocoela is recovered as the sister taxa to the Polycladida with some bootstrap support (0.96/56).

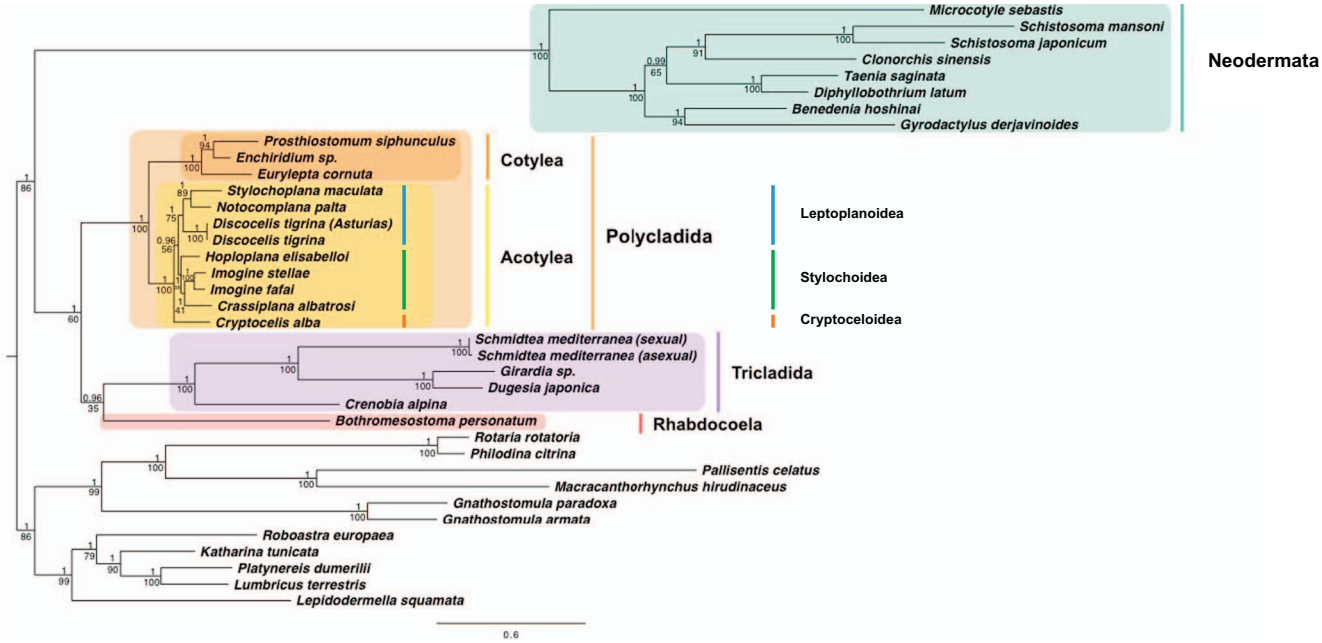

Fig. 5. Phylogeny indicating relationships between Platyhelminthes and related Phyla, based on amino acid sequences of genes within the mitochondria of these species. Bayesian Inference (BI) and Maximum Likelihood (ML) analysis resulted in this arrangement, with BI-derived tree displayed here. Numbers at base of nodes indicates BI posterior probability (top)/ML bootstrap support (% of 1000 replicate), and substitutions per unit length given by scale bar at base of figure. Coloured boxing/lines indicate orders of Platyhelminth, along with the sub-orders Cotylea and Acotylea.

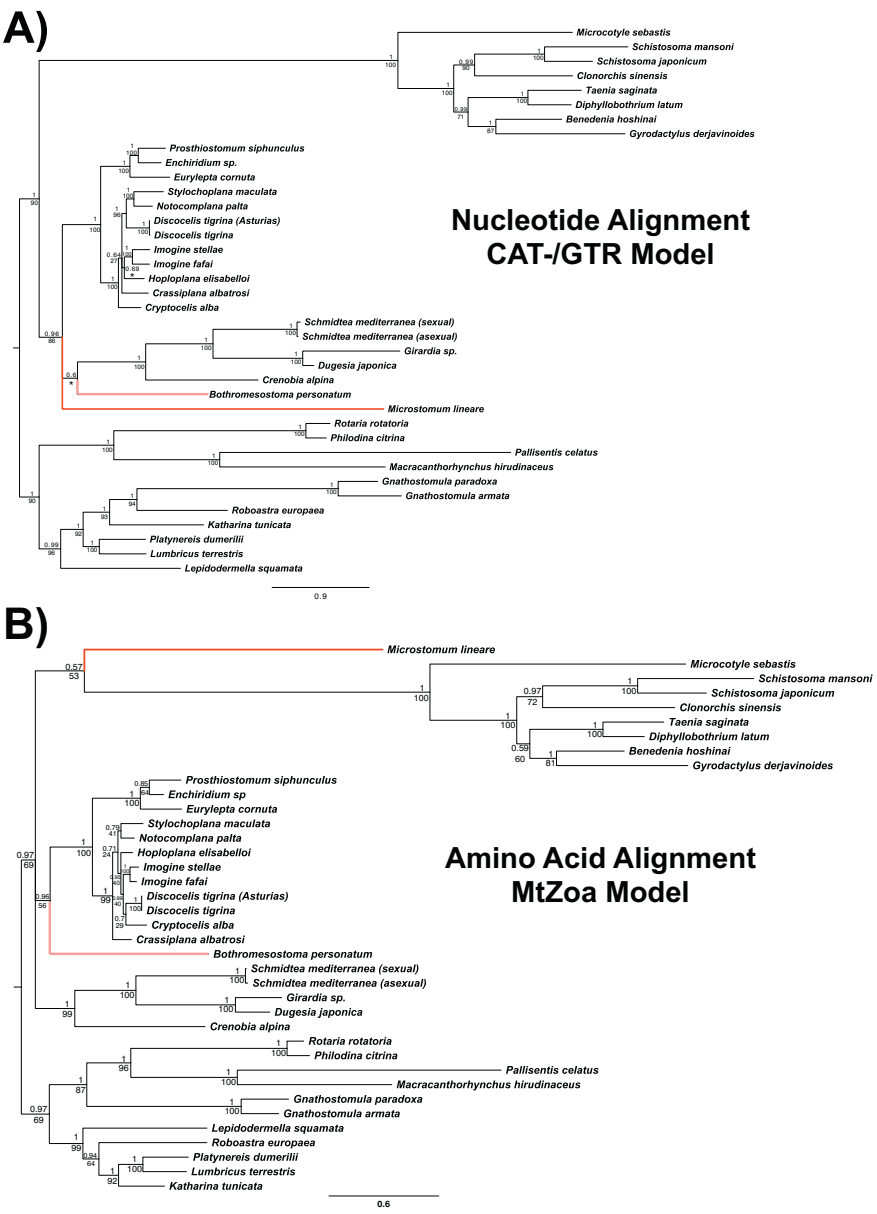

Fig. 6. Phylogenies reproduced with the addition of the partial *Microstomum lineare* mitochondrial genome sequence information to previously considered datasets, under Bayesian Inference (BI)/Maximum Likelihood (ML) analysis. A) above, shows phylogeny based on nucleotide alignment of five protein coding genes, along with *rrnL* and *rrnS*, under the GTR + I + G (ML)/ CAT-GTR + I + G (BI) models, while B) below shows relationships based on amino acid alignments of five protein coding genes within the mitochondria of the species displayed (Mtzoa +I + G + F model). Upper numbers at base of nodes indicates BI posterior probability support, lower numbers ML bootstrap support (%age, of 1000 replicates). Bayesian analysis-derived trees are shown, and where ML trees did not provide support for nodes, this has been indicated with a * and the nodes have been collapsed to a polytomy. Scale bar below trees represents substitutions per given unit length. Position of *M. lineare* indicated by colouring of line leading to this taxa in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Free living flatworms (turbellarians, to use a previously posited term) are still supported as the sister taxa of the Neodermata by nucleotide based evidence, and with reasonable support values (0.96/86). However, under amino acid-based analyses, *M. lineare* appears as the sister clade of the Neodermata, albeit with only slight posterior probability (0.57) and bootstrap (53) support. Again, this contrasting arrangement suggests that deeper sampling of flatworm diversity will be necessary for disentangling inter-relationships within the Platyhelminthes, and the short (5 protein coding gene) sequence of *M. lineare* currently available will constrain analysis somewhat. This result is also potentially influenced by the long branch leading to the Neodermata, coupled to *M. lineare*'s position near the root of the tree (which itself couples well with Macrostomida's position at the base of the Platyhelminth radiation). Completing the sequencing of *M. lineare*, and adding more Macrostomida to the sequenced record, would both aid in the resolution of this tree.

Even with the addition of Macrostomid data to our analysis, no evidence for the existence of the Neoophora clade can be found, and Rhabdocoela, Tricladida and Neodermata (along with other taxa) are not supported as sister clades to the exclusion of the Polycladida. While future sequencing efforts will certainly test this further, the Neoophora hypothesis is not supported by any of the data presented here.

4. Discussion

4.1. The Whole Genome Shotgun (WGS) approach

This paper has used the WGS approach to mitochondrial genome assembly, with excellent results. Whole mitochondrial assemblies were gained using simple DNA extraction techniques, followed by the increasingly commonplace mechanisms of NGS on the Illumina platform followed by assembly using freely available short read assembly tools. While sequencing and library construction has some associated upfront cost, the ability to multiplex multiple samples into a single lane of sequencing renders these reasonable, when compared to the cost of primer design, sanger sequencing and personnel costs associated with more traditional methods. This approach is recommended, building upon 454 sequencing approaches assayed earlier (e.g [47].), and will be easily applicable to other platyhelminth species in the future.

4.2. Mitochondrial genome sequences and their utility

The provision of the mitochondrial DNA sequences presented here has added markedly to the sampling of the Polycladida as a whole, particularly for the acotylean clade. The addition of a rhabdocoel sequence to the public record will aid in understanding the evolution and diversification of the Platyhelminthes. The genome of *B. personatum* has interesting characteristics, with a low GC% and highly constrained canonical stop codons observed. Mitochondrial sequences have proved their usefulness in a variety of contexts, and these novel sequences are therefore likely to be used in a range of future analyses.

The size of the mtDNA sequences studied here were relatively consistent, varying from 14,783 bp to 15,909 bp, in agreement with previous results on this clade (e.g [1].). The GC% of the assembled genomes are entirely consistent with those seen in other platyhelminth species sequenced to date [1,47,54], and are not unusual for mitochondrial genomes as a whole, which tend to be AT-rich [46]. The overlapping sequences detected between genes may have contributed to the maintenance of the syntenic relationships of these genes across evolutionary time.

The start and stop codons of platyhelminth species within mtDNA genomes are known to vary from the canonical ATG start and TAA/TGA/TAG stop sequences. Previous investigations in platyhelminths [1, 32,33,45, 51, 54] have noted the use of “TA” and ‘T’ stop signals, as is common in many metazoans, along with a diversity of other inferred start and stop sequences, although ATG and TAA are the most commonly observed. The most common start/stop codons for all novel species sequenced here agree with the “echinoderm and flatworm mitochondrial code” [50], with ATG and TAA the most common, and TGA used to code for both “STOP” and tryptophan. This provides further evidence that the “alternative flatworm mitochondrial code” (where TAA codes for tyrosine rather than “STOP”) [7] is not generally used across the Platyhelminthes, and could instead be artifactual or limited due to sample size [50]. Given previous knowledge, and with the additional evidence from the species presented here, it seems that the use of alternate codons is prevalent throughout the Platyhelminthes.

The two specimens of *D. tigrina* sequenced in this work showed occasional differences at the protein coding level between them. This may reflect differences at the population level in this sequence, as the samples were drawn from two different sites on the Iberian Peninsula, Asturias and Galicia. Further investigation will be necessary to discern whether this reflects population structure, normal variation, or cryptic speciation.

Phylogenetic inference based on these sequences has provided a coherent and consistent set of results regarding polyclad relationships, the results of which are discussed in more detail in Section 4.3. This understanding can be further corroborated by mitochondrial gene order results, which may provide convenient rare genomic change data that can support clades suggested by phylogenetic analysis. Well conserved syntenic areas can be seen in the arrangement of mitochondrial genes across the polyclad species examined, and some potential cotylean synapomorphies in tRNA and *nad3* arrangement can be observed, as noted above. The detailed syntenic arrangements and resulting evolutionary arguments that can be made as a result of their patterns of change (i.e. tandem duplication and random loss (tdrl), [10]) are themselves a good argument in support of the continued sequencing of mitochondrial genomes in the Platyhelminthes.

4.3. Phylogenetic inference and platyhelminthes relationships

Well-supported trees were reliably recovered by our analyses. As phylogenetic analysis was performed using both ML and BI, both amino acid and nucleotide data, and site-heterogenous and empirical matrix-based models, all of which consistently supported essentially the same tree, we are confident that our results are robust and well supported. This work has begun to stringently assess established views of platyhelminth relationships from a mitochondrial standpoint.

Platyhelminthes is supported as a monophyletic clade. While this is evident based on morphological evidence, previous molecular systematic studies have struggled to recover a monophyletic Platyhelminthes, given the long branch leading to the Neodermata (see, for example, Fig. 4, [1]). The Rhabditophora, which is generally considered to include the Neodermata and all free-living flatworms with the exception of the Catenulids, was not tested by our analysis, and the sequencing of one or more catenulid mitochondrial genomes would be useful for examining this in future work. The existence of a clade of free-living flatworm species, similar to those previously gathered as the class “Turbellaria”, is supported. However, we have not sampled the entirety of free-living flatworm diversity here, and the existence of this grouping as a true monophyletic clade is unlikely, with the Neodermata likely diverging from a free-living ancestor (see [15,16]). Sequencing of a member of the Proseriata, in particular, would aid in testing this monophyly (e.g [44].). Our results including *M. lineare*, below in this section, also aid in this testing process. While rhabdocoels are noted as the possible sister taxon to the Tricladida by our results, additional data from other rhabdocoel species, alongside that of other clades of flatworm, particularly the Prolecithophora, which are often noted as the sister taxa to Tricladida (e.g [44].) will allow to examine these relationships in more detail in the future.

Our analyses recover Polycladida, Cotylea and Acotylea as monophyletic. Recent results have brought into question our understanding of the Superfamily relationships within the Acotylea, with the Leptoplanoidea recovering part of the previously accepted superfamily Ilyplanoidea, *D. tigrina*, while other species of the former Ilyplanoidea, such as *Adenoplana*, become part of the superfamily Cryptocelidae (see [5]).

In the Acotylea, the arrangement of the superfamilies Stylochoidea, Leptoplanoidea and Cryptoceloidea as proposed in [5] is confirmed by the data presented here. In particular, the Leptoplanoidea appears to include members of the former superfamily Ilyplanoidea (*D. tigrina*). The Stylochoidea are maximally well supported by BI, although ML analysis gives more ambivalent results with poorer bootstrap results in our amino acid tree. Nevertheless, the superfamily Stylochoidea (including, in our analysis, *I. fafai*, *I. stellae*, *C. albatrossi* and *H. elisabelloi)* is nonetheless supported, a result that coincides with those obtained by Bahia et al. [5]. As in that analysis, *C. alba* is separated from Leptoplanoida, its previous location, and is placed in another superfamily, the Cryptoceloidea Bahia 2017.

The topology of the Cotylea, with only three species sequenced to date, is uncontroversial in our analysis, but has not been tested robustly. According to recent research [5] the suborder Cotylea is constituted by 5 superfamilies: Cestoplanoidea, Periceloidea, Chromoplanoidea, Prosthiostomoidea and Pseudocerotoidea. The species analyzed in our study are part of the superfamily Prosthiostomoidea (*P. siphunculus* and *Enchiridium* sp.) and Pseudocerotoidea (*E. cornuta*). A close relationship between two members of the Prosthiostomidae, *Enchiridium* sp. and *P. siphunculus*, has been confirmed here. Future work will more stringently test cotylean interrelationships, particularly in light of the findings of Bahia et al. [5], where supposed acotyleans, like *Theama* and *Cestoplana*, were found to be nested within the Cotylea sensu stricto.

The *B. personatum* mitochondrial genome is, to our knowledge, the first sequenced representative of the Rhabdocoela at the time of publication. The size and coding quantity (c.f. non-coding fraction) of this genome did not differ markedly from that of all other mitochondrial sequences examined in this manuscript. However, no non-canonical stop codons were seen in the *B. personatum* mitochondrial genome, unlike polycladids, where these are commonly encountered. Stop codon usage may therefore be more constrained in the rhabdocoels than in other clades.

The addition of sequences from *B. personatum* also helps to increase the sampling of flatworm diversity. The placement of the Rhabdocoela

as the sister clade of the Tricladida is of interest, but is poorly supported by ML analysis in particular. Only under Bayesian analysis of amino acid data is this hypothesis particularly well supported (posterior probability = 0.96). Comparison of gene order between the Rhabdocoela and Tricladida (Fig. 2) is unconclusive due to widespread rearrangement between the species examined. As noted in this work, we heavily recommend further sequencing, which could confirm the Rhabdocoela as sister to the Tricladida, relationship also mentioned by Egger et al. [15]. Given the placement of the *B. personatum* here under a variety of models with both amino acid and nucleotide data, it is perhaps the most likely placement given current evidence, but the sequencing of mitogenomes from the Proseriata, Prolecithophora, Haplopharyngida, Lecithoepitheliata and PNUK (*Piscinquilinus*, *Notentera*, *Urastoma* and *Kronborgia*) [34,35] groups would provide a conclusive test for the true phylogenetic position of the Rhabdocoela as a whole.

The short sequence of *M. lineare*'s mitochondrial genome currently available constrains the utility of phylogenetic analysis, leading to uncertainty and poor support values for its placement in the Platyhelminthes tree (Fig. 6.). This is probably not aided by its likely true placement towards the base of the Platyhelminthes radiation, which will make it susceptible to changes in position near the node of the Platyhelminthes under phylogenetic analysis [15,30,31]. The provision of further Macrostomorpha sequences would therefore be useful for confirming the true position of that clade, and in the meantime analyses with the truncated sequence should continue to be interpreted with caution.

In comparison with previous trees using polyclad mitochondrial sequence data (e.g [1].) our trees are more robust, especially when *M. lineare* is not incorporated, and are not prey to the same odd, likely artifactual errors (for example, the recovery of the Neodermata as Gnathiferans). Even when *M. lineare* is included we recover a monophyletic Platyhelminthes and our topology changes little. This is almost certainly due to our increased sampling, and confirms the advantages of deeper sampling in the future.

4.4. Is Neoophora monophyletic?

The Neoophoran clade (uniting a range of clades with heterocellular female gonads and ectolecithal embryos) is not supported by any of the phylogenetic data we present. The Neodermata, Tricladida and Rhabdocoela would be gathered into a single clade to the exclusion of the Polycladida if the Neoophoran hypothesis was correct. This could mean that the ectolecithal condition is ancestrally shared within the Platyhelminthes, and subsequently lost in the Polycladida, or alternately it could imply that ectolecithal embryos have been evolved independently on two occasions - once in the Neodermata and once in the (Tricladida + Rhabdocoela) clade.

In contrast, the results shown here would argue for a sister-clade relationship of free living flatworm species sampled, to the exclusion of the Neodermata. The Turbellarian hypothesis, uniting all free-living flatworms into a monophyletic clade, is generally discounted by both molecular and morphological evidence. It does, however, gain some support in our data, but we recognize that not all orders of flatworm are present in our analysis, and that the unique characteristics and rapid molecular evolution of the Neodermata may separate them artificially from the free living flatworm species examined in this study, leading to an artifactual placement. The revisiting of this question later, with more mitochondrial sequences in hand, would likely answer this question more conclusively.

4.5. Platyhelminthes and its place in the wider Lophotrochozoa

Platyhelminthes is recovered as a monophyletic group in all our analyses, unlike in previous investigations (including mitochondrial ones) (e.g. Fig. 4, [1]). The monophyly of the Platyhelminthes is however well supported by previous evidence. Outgroups chosen in this study were drawn from a range of other lophotrochozoan species. Both our nucleotide-based tree (Fig. 4) and our amino acid based tree (Fig. 5) strongly support the sister taxa relationship of Rotifera and Acanthocephala, although further rotiferan sequences would probably see Acanthocephala nested within the Rotifera (e.g [20].). Both of these trees also support the Gnathostomulida as the sister taxa to this Phyla, supporting the Gnathiferan hypothesis (e.g [19,25].). This clade is sister to a clade formed by gastrotrichs, molluscs and annelids (with slightly varying topology and support values).

All of the sister taxa are more closely related to one another than they are to the Platyhelminthes, and our trees offer no support to a "Rouphozoa" hypothesis [55] linking Gastrotricha and Platyhelminthes. Instead, our trees suggest gastrotrichs are more closely related to Trochozoa. This exclusion of Platyhelminthes from a group consisting of other lophotrochozoans such as gastrotrichs is potentially interesting, as its inclusion with them in the "Platyzoa" in phylogenies and due to characteristics such as a shared positive GC skew ([54], sometimes light-heartedly called the "Longbranchozoa") in some analyses may well have been due to long branch attraction rather than an underlying biological reality. An early-branching difference between Platyhelminthes and other lophotrochozoans could make some sense given previously published results (e.g [19].) but will only be rigorously tested by the addition of mitochondrial sequences from other Phyla and testing with a variety of outgroups.

4.6. Future sampling efforts

Representative sequencing of the Platyhelminthes remains a work in progress. This fact precludes stronger conclusions in several aspects. Catenulid species sequences would allow analysis of the base of the platyhelminth radiation, and sequences of species of Proseriata and Prolecithophora (and to a lesser degree, those of Haplopharyngida, Lecithoepitheliata and "PNUK" [34,35] would allow a better understanding of higher level Platyhelminthes relationships. To test cotylean monophyly and superfamilial relationships, it is the Acotylea that should be targeted for further sampling. Sequencing the mitochondrial genomes of representatives of the Theamatidae and Cestoplanidae would allow conclusive understanding of their position within the Polycladida, and would provide a robust test of the phylogeny presented by Bahia et al. [5] and Aguado et al. [2].

The Cotylea and Acotylea, which are both well supported as monophyletic clades in the present analysis, could have this further tested by targeted sequencing of the mitochondrial genomes of species such as *Pericelis* (Cotylea; Pseudocerotoidea; Pericelidae) [42]., who additionally included *Pericelis* in their study of nuclear *28S rRNA* gene sequences, were unable to show Cotylean monophyly, and instead recovered *Pericelis* as the sister taxon to the Acotylea. On the other hand, Bahia et al. [5] include *Pericelis* in their analysis with *28S* rRNA and recover this taxon within Cotylea, but isolated and as sister-group to the rest of Cotylea (excluding Cestoplanoidea). Faced with these seemingly contradictory results, we therefore recommend the sequencing of the mtDNA of *Pericelis* at the first possible opportunity.

5. Conclusions

Free-living flatworm relationships are still under debate, but the addition of novel molecular data is greatly increasing our understanding. Here we have described the mitochondrial genome sequences of nine species of platyhelminth, adding significantly more data to available resources in these under-explored clades. Particularly, these samples will allow us to more firmly understand Polycladida taxonomy, and provide a firm base for future, targeted exploration in that clade. Furthermore, the description of the complete mitochondrial genome of the rhabdocoel *B. personatum* represents a significant addition to published knowledge.

We have used these resources to investigate platyhelminth

relationships in general and polycladid relationships in particular. Our results strongly suggest that presently assumed cotylean superfamilial relationships need revision, and will provide the basis for further testing of Acotylean monophyly, as well as the relationships within platyhelminthes. Our results do not provide any evidence for a "Neoophoran" clade, relating Rhabdocoela, Tricladida and Neodermata (excluding Polycladida), a finding which contrasts with general understanding of Platyhelminth inter-relationships. However, this finding must be tested more stringently when new evidence becomes available. The resources presented here, alongside other novel resources, will help to give a final answer to old questions, such as the relationship of platyhelminths to other lophotrochozoans, those within in the platyhelminths, and the of the exact history of free-living species with regard to the parasitic Neodermata. With the advent of sufficient molecular data, these recalcitrant questions will finally be categorically addressed.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2018.02.009.

Acknowledgements

The authors thank the members of their laboratories for helpful comments and aid in the preparation of this manuscript. Sequencing was performed by AllGenetics & Biology, Coruña, Spain. Francisco Brusa, Fernando Angel Fernandez Alvarez, Bruno Almon and Jacinto Perez provided specimens. We thank the artist who provided animal silhouettes for the Phylopic resource used in our graphical abstract (Mali'o Kodis, National Museums of Ireland) and acknowledge the licence that allowed us to use it, CC BY-NC-SA 3.0. The authors thank the reviewers and editors for their aid and helpful suggestions in preparing this work for publication.

Funding

CG was for a portion of the time spent on this project a "Ramon y Cajal" postdoctoral fellow supported by the Spanish MEC (Ministerio de Economia y Competividad) and the UAM and funded by project no CGL2011-29916. NJK is currently supported by a H2020 MSCA grant [IF750937]. Funding sources had no involvement in the decision to submit the article for publication.

References

[1] M.A. Aguado, C. Grande, M. Gerth, C. Bleidorn, C. Noreña, Characterization of the complete mitochondrial genomes from Polycladida (Platyhelminthes) using next-generation sequencing, Gene 575 (2016) 199–205.
[2] M.T. Aguado, C. Noreña, L. Alcaraz, D. Marquina, F. Brusa, C. Damborenea, B. Almon, C. Bleidorn, C. Grande, Phylogeny of Polycladida (Platyhelminthes) based on mtDNA data, Org. Divers. Evol. 17 (4) (2017) 767–778, http://dx.doi.org/10.1007/s13127-017-0344-4.
[3] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data, Available online at, 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc, .
[4] P. Ax, Multicellular Animals. A New Approach to the Phylogenetic Order in Nature, Vol. I Springer, Berlin, Heidelberg, New York, 1996.
[5] J. Bahia, V. Padula, M. Schrödl, Polycladida phylogeny and evolution: integrating evidence from 28S rDNA and morphology, Org. Divers. Evol. 17 (3) (2017) 653–678.
[6] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Putz, M. Middendorf, P.F. Stadler, MITOS: Improved *de novo* metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2) (2013) 313–319.
[7] Y. Bessho, T. Ohama, S.J. Osawa, Planarian mitochondria. II. The unique genetic code as deduced from cytochrome c oxidase subunit I gene sequences, J. Mol. Evol. 34 (1992) 331–335.
[8] D. Blair, A. Campos, M.P. Cummings, J.P. Laclette, Evolutionary biology of parasitic platyhelminthes: the role of molecular phylogenetics, Parasitol. Today 12 (1995) 66–71.
[9] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (15) (2014) 2114–2120.
[10] J.L. Boore, The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals, Comp. Genom. (2000) 133–147 (Springer, Dordrecht).
[11] J.T. Cannon, B.C. Vellutini, J. Smith, F. Ronquist, U. Jondelius, A. Hejnol, Xenacoelomorpha is the sister group to Nephrozoa, Nature 530 (7588) (2016) 89–93.
[12] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (4) (2000) 540–552.
[13] D. Darriba, G.L. Taboada, R. Doallo, D. Posada, ProtTest 3: fast selection of best-fit models of protein evolution, Bioinformatics 27 (2011) 1164–1165.
[14] D. Darriba, G.L. Taboada, R. Doallo, D. Posada, jModelTest 2: more models, new heuristics and parallel computing, Nat. Methods 9 (8) (2012) 772.
[15] B. Egger, F. Lapraz, B. Tomiczek, S. Müller, C. Dessimoz, J. Girstmair, N. Skunca, K.A. Rawlinson, C.B. Cameron, E. Beli, M.A. Todaro, M. Gammoudi, C. Noreña, M.J. Telford, A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms, Curr. Biol. 25 (10) (2015) 1347–1353.
[16] U. Ehlers, Das Phylogenetische System der Plathelminthes, G. Fischer, Stuttgart and New York, 1985.
[17] A. Faubel, The Polycladida, Turbellaria; Proposal and Establishment of a New System. Part 1. The Acotylea, Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut 80 (1983) 17–121.
[18] A. Faubel, The Polycladida, Turbellaria; Proposal and Establishment of a New System. Part 2. The Cotylea, Mitteilungen aus dem Hamburgischen Zoologischen Museum und Institut 81 (1984) 189–259.
[19] A.C. Fröbius, P. Funch, Rotiferan Hox genes give new insights into the evolution of metazoan bodyplans, Nat. Commun. 8 (1) (2017) 9.
[20] J.R. Garey, T.J. Near, M.R. Nonnemacher, S.A. Nadler, Molecular evidence for Acanthocephala as a subtaxon of Rotifera, J. Mol. Evol. 43 (3) (1996) 287–292.
[21] A. Golombek, S. Tobergte, T.H. Struck, Elucidating the phylogenetic position of Gnathostomulida and first mitochondrial genomes of Gnathostomulida, Gastrotricha and Polycladida (Platyhelminthes), Mol. Phylogenet. Evol. 86 (2015) 49–63.
[22] L.H. Hyman, The Invertebrates, Platyhelminthes and Rhynchocoela. The Acoelomate Bilateria, Vol. II McGraw-Hill, New York, 1951.
[23] K. Katoh, K. Misawa, K.I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Res. 30 (14) (2002) 3059–3066.
[24] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780.
[25] R.M. Kristensen, P. Funch, Micrognathozoa: a new class with complicated jaws like those of Rotifera and Gnathostomulida, J. Morphol. 246 (1) (2000) 1–49.
[26] P. Kück, K. Meusemann, FASconCAT: convenient handling of data matrices, Mol. Phylogenet. Evol. 56 (3) (2010) 1115–1118.
[27] A. Lang, Die Polycladen (Seeplanarien) des Golfes von Neapel und der angrenzenden Meeresabschnitte, Eine Monographie. Fauna Flora Golfes v, Neapel, Leipzig, 1884.
[28] K. Larsson, U. Jondelius, Phylogeny of Catenulida and support for Platyhelminthes, Org. Divers. Evol. 8 (2008) 378–387.
[29] N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating, Bioinformatics 25 (2009) 2286–2288.
[30] C.E. Laumer, N. Bekkouche, A. Kerbl, F. Goetz, R.C. Neves, M.V. Sørensen, R.M. Kristensen, A. Hejnol, C.W. Dunn, G. Giribet, K. Worsaae, Spiralian phylogeny informs the evolution of microscopic lineages, Curr. Biol. 25 (15) (2015) 2000–2006.
[31] C.E. Laumer, A. Hejnol, G. Giribet, Nuclear genomic signals of the 'micro-turbellarian' roots of platyhelminth evolutionary innovation, elife 4 (2015) e05503.
[32] T.H. Le, D. Blair, D.P. McManus, Mitochondrial genomes of parasitic flatworms, Trends Parasitol. 1 (2002) 206–213.
[33] T.H. Le, D.P. McManus, D. Blair, Codon usage and bias in mitochondrial genomes of parasitic Platyhelminthes, Korean J. Parasitol. 42 (2004) 159–167.
[34] D.T.J. Littlewood, K. Rohde, R.A. Bray, E.A. Herniou, Phylogeny of the Platyhelminthes and the evolution of parasitism, Biol. J. Linn. Soc. 68 (1–2) (1999) 257–287, http://dx.doi.org/10.1111/j.1095-8312.1999.tb01169.x.
[35] D.T.J. Littlewood, K. Rohde, K.A. Clough, The interrelationships of all major groups of Platyhelminthes: phylogenetic evidence from morphology and molecules, Biol. J. Linn. Soc. 66 (1) (1999) 75–114, http://dx.doi.org/10.1111/j.1095-8312.1999.tb01918.x.
[36] D.T.J. Littlewood, Platyhelminth systematics and the emergence of new characters, Parasite 15 (3) (2008) 333–341.
[37] M. Litvaitis, L. Newman, A molecular framework for the phylogeny of the Pseudocerotidae (Platyhelminthes, Polycladida), Hydrobiologia 444 (2001) 177–182.
[38] M.B. Lohse, A.D. Hernday, P.M. Fordyce, L. Noiman, T.R. Sorrells, V. Hanson-Smith, et al., Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains, Proc. Natl. Acad. Sci. 110 (2013) 7660–7665.
[39] S. Prudhoe, A monograph on polyclad Turbellaria, British Museum (Natural History), Oxford University Press, Oxford, New York, 1985.
[40] A. Rambaut, Figtree v1.4.3, Available Online, 2016. http://tree.bio.ed.ac.uk/software/figtree/, .
[41] K. Rawlinson, M. Litvaitis, Cotylea (Polycladida): a cladistic analysis of morphology, Invertebr. Biol. 127 (2008) 121–138.
[42] K.A. Rawlinson, J.A. Gillis, J. Billings, E.H. Borneman, Taxonomy and life history of the *Acropora*-eating flatworm *Amakusaplana acroporae* nov. sp. (Polycladida: Prosthiostomidae), Coral Reefs 30 (2011) 693–705.
[43] K. Rawlinson, The diversity, development and evolution of polyclad flatworm larvae, EvoDevo 5 (9) (2014) 1–13.
[44] M. Riutort, M. Alvarez-Presas, E. Lazaro, E. Solá, J. Paps, Evolutionary history of the Tricladida and the Platyhelminthes: an up-to-date phylogenetic and systematic

account, Int. J. Dev. Biol. 56 (2012) 5–17.
[45] M. Sakai, M. Sakaizumi, The complete mitochondrial genome of *Dugesia japonica* (Platyhelminthes; order Tricladida), Zool. Sci. 29 (2012) 672–680.
[46] D.R. Smith, Updating our view of organelle genome nucleotide landscape, Front. Genet. 3 (2012) 175.
[47] E. Solà, M. Álvarez-Presas, C. Frías-López, D.T.J. Littlewood, J. Rozas, M. Riutort, Evolutionary analysis of mitogenomes from parasitic and free-living flatworms, PLoS One 10 (3) (2015) e0120081.
[48] A. Stamatakis, P. Hoover, J. Rougemont, A rapid bootstrap algorithm for the RaxML web servers, Syst. Biol. 57 (2008) 758–771.
[49] P. Stothard, The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences, BioTechniques 28 (2000) 1102–1104.
[50] M.J. Telford, E.A. Herniou, R.B. Russell, D.T.J. Littlewood, Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms, Proc. Natl. Acad. Sci. 97 (21) (2000) 11359–11364.
[51] A.R. Wey-Fabrizius, H. Herlyn, B. Rieger, D. Rosenkranz, A. Witek, D.B.M. Welch, I. Ebersberger, T. Hankeln, Transcriptome data reveal syndermatan relationships and suggest the evolution of endoparasitism in Acanthocephala via an epizoic stage, PLoS One 9 (2) (2014) e88618.
[52] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res. 18 (5) (2008) 821–829.
[53] M.K. Litvaitis, D.M. Bolanos, S.Y. Quiroga, When names are wrong and colours deceive: unravelling the *Pseudoceros bicolor* species complex (Platyhelminthes: Polycladida), J. Nat. Hist. 44 (2010) 829–845, http://dx.doi.org/10.1080/00222930903537074.
[54] A.R. Wey-Fabrizius, L. Podsiadlowski, H. Herlyn, T. Hankeln, Platyzoan mitochondrial genomes, Mol. Phylogenet. Evol. 69 (2) (2013) 365–375, http://dx.doi.org/10.1016/j.ympev.2012.12.015.
[55] T.H. Struck, A.R. Wey-Fabrizius, A. Golombek, L. Hering, A. Weigert, C. Bleidorn, S. Klebow, N. Iakovenko, B. Hausdorf, M. Petersen, P. Kück, H. Herlyn, T. Hankeln, Platyzoan paraphyly based on phylogenomic data supports a non-coelomate ancestry of Spiralia, Mol. Biol. Evol. 31 (2014) 1833–1849.