The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of atp8 and genome plasticity in flatworms

Non Peer-reviewed author version

MONNENS, Marlies; THIJS, Sofie; Briscoe, Andrew G.; Clark, Miriam; Frost, Emily Joy; Littlewood, D. Tim J.; Sewell, Mary; SMEETS, Karen; ARTOIS, Tom & VANHOVE, Maarten (2020) The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of atp8 and genome plasticity in flatworms. In: International journal of biological macromolecules, 162 , p. 454 -469.

DOI: 10.1016/j.ijbiomac.2020.06.025
Handle: http://hdl.handle.net/1942/31464

The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of atp8 and genome plasticity in flatworms

Marlies Monnens, Sofie Thijs, Andrew G. Briscoe, Miriam Clark, Emily Joy Frost, D. Tim J. Littlewood, Mary Sewell, Karen Smeets, Tom Artois, Maarten P.M. Vanhove

PII: S0141-8130(20)33465-6

DOI: https://doi.org/10.1016/j.ijbiomac.2020.06.025

Reference: BIOMAC 15779

To appear in: *International Journal of Biological Macromolecules*

Received date: 6 December 2019

Revised date: 2 June 2020

Accepted date: 3 June 2020

Please cite this article as: M. Monnens, S. Thijs, A.G. Briscoe, et al., The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of atp8 and genome plasticity in flatworms, *International Journal of Biological Macromolecules* (2020), https://doi.org/10.1016/j.ijbiomac.2020.06.025

The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of *atp8* and genome plasticity in flatworms

Marlies Monnens[a], Sofie Thijs[b], Andrew G. Briscoe[c], Miriam Clark[d], Emily Joy Frost[d], D. Tim J. Littlewood[c], Mary Sewell[d], Karen Smeets[a], Tom Artois[a], Maarten P.M. Vanhove[a,e,f,g]

[a]Hasselt University, Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, Agoralaan Gebouw D, B-3590 Diepenbeek, Belgium. marlies.monnens@uhasselt.be, karen.smeets@uhasselt.be, tom.artois@uhasselt.be, maarten.vanhove@uhasselt.be.

[b]Hasselt University, Centre for Environmental Sciences, Research Group Environmental Biology, Agoralaan Gebouw D, B-3590 Diepenbeek, Belgium. sofie.thijs@uhasselt.be.

[c]Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom. a.briscoe@nhm.ac.uk, t.littlewood@nhm.ac.uk.

[d]School of Biological Sciences, University of Auckland, New Zealand. clarkrmb@gmail.com, efro833@aucklanduni.ac.nz, m.sewell@auckland.ac.nz.

[e]Laboratory of Biodiversity and Evolutionary Genomics, University of Leuven, Charles Deberiotstraat 32, B-3000 Leuven, Belgium.

[f]Zoology Unit, Finnish Museum of Natural History, University of Helsinki, P.O.Box 17, Helsinki FI-00014, Finland.

[g]Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic

Declarations of interest: none.

Corresponding author: Marlies Monnens. Hasselt University, Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, Agoralaan Gebouw D, B-3590 Diepenbeek, Belgium. P: +32 11 85 46. F: +32 11 26 81 99. marlies.monnens@uhasselt.be.

Abstract

The first three mitochondrial (mt) genomes of endosymbiotic turbellarian flatworms are characterised for the rhabdocoels *Graffilla buccinicola, Syndesmis echinorum* and *S. kurakaikina*. Interspecific comparison of the three newly obtained sequences and the only previously characterised rhabdocoel, the free-living species *Bothromesostoma personatum*, reveals high mt genomic variability, including numerous rearrangements. The first intrageneric comparison within rhabdocoels shows that gene order is not fully conserved even at species levels. *Atp8*, until recently assumed absent in flatworms, was putatively annotated in two sequences. Selection pressure was tested in a phylogenetic framework and is shown to be significantly relaxed in this and another protein-coding gene: *cox1*. If present, *atp8* appears highly derived in platyhelminths and its functionality needs to be addressed in future research. Our findings for the first time allude to a large degree of undiscovered (mt) genomic plasticity in rhabdocoels. It merits further attention whether this variation is correlated with a symbiotic lifestyle. Our results illustrate that this phenomenon is widespread in flatworms as a whole and not exclusive to the better-studied neodermatans.

Keywords

Platyhelminthes, Rhabdocoela, genomics, endosymbiosis, mitochondrion

1. Introduction

With an estimated 30.000 described species occurring worldwide, flatworms (Platyhelminthes Minot, 1876) comprise one of the most diverse metazoan taxa on Earth [2,

3]. While easily overlooked, the platyhelminths are found on every continent, in marine and freshwater habitats, where they occupy a wide range of different ecological niches and often occur in substantial numbers [4-7]. As general predators and scavengers, free-living species typically fulfil a key function in meiofaunal communities [7, 8], but flatworms are most widely known for their role as parasites of biomedical and economic importance [9].

Increasingly available sequence data on flatworms have been used successfully to tackle open questions in systematics and for diagnostic purposes [10]. However, progress in this field is partially impeded by the sampling and taxonomical difficulties inherent to almost all of the less known (micro)turbellarian groups [11-13]. Furthermore, obtaining sufficient high-quality genetic material (target DNA) can pose problems in studies of smaller species or life stages [14].

Currently available sequence data show extensive genomic diversity amongst Platyhelminthes (see review by [15]). Indeed, as more flatworm species are incorporated into molecular studies, a greater number of genomic peculiarities are being discovered. Examples include several instances of polyploidy [16], high degrees of genomic polymorphy [17], deviations from the 'standard' genetic codes [18] and 18S rDNA polymorphisms [19]. In the case of mitochondrial (mt) genomes, this variation is well illustrated by technical difficulties in protocols that are relatively straightforward with other taxa. For instance, the *cox1* barcoding region appears highly divergent in flatworms and amplification often requires careful fine-tuning and design of taxon-specific primers [20-24].

Mitochondrial genomic variability in flatworms is further evidenced by the sequence divergence in other protein-coding genes (PCGs), sometimes even at the intraspecific level (*e.g.* [25, 26]). The most-widely-discussed PCG in this respect is *atp8*. This gene is lost in Neodermata [27, 28] and for a long time it was assumed that the same was true for all flatworms [29, 30]. However, this view has been challenged and it has been recently proposed that *atp8* is present in turbellarians, but was not detected in previous (standard) annotation procedures [31, 32].

Taxonomic coverage of Platyhelminthes for genomic data is still limited and most sequencing efforts have been directed towards economically relevant parasitic taxa. Microturbellarians in particular have received little attention in this respect, an exemplary group being Rhabdocoela Ehrenberg, 1831. To date, no reference genome of this group has been published and only a single mt genome has been characterised from *Bothromesostoma personatum* (Schmidt 1848) Braun, 1885, a well-known inhabitant of freshwater pools and brooks [34, 37]. As Rhabdocoela constitutes a particularly species-rich, cosmopolitan clade, with members occurring in many different ecological niches, including marine, brackish, freshwater and even terrestrial habitats [4], there is much to be explored at the molecular level.

Rhabdocoels comprise a promising, but unexplored group to study endosymbiosis. One of the reasons evolutionary biologists have been targeting flatworms is to better comprehend the origins and evolution of parasitism. Research in this regard is largely, if not exclusively, centred on neodermatans (e.g. [38-42]). However, while often dubbed the ‘free-living flatworms’, several turbellarian lineages have also developed a symbiotic lifestyle [43].

Within Rhabdocoela alone over a hundred obligate endosymbiont species have been described (*e.g.* [44]). In 2013, Van Steenkiste, et al. [11] demonstrated that the transition towards endosymbiosis has occurred three times independently in the evolutionary history of rhabdocoels, resulting in three monophyletic families (Umagillidae Wahl, 1910, Pterastericolidae Meixner, 1926 and Graffillidae von Graff, 1904-1908), each of which infects a specific lineage of marine molluscs or echinoderms.

In this study, we target the mt genomes of two species of *Syndesmis* Silliman, 1881 (Umagillidae), the most species-rich genus of endosymbiotic rhabdocoels [49, 50] and of *Graffilla buccinicola* Jameson, 1897, a graffillid parasite of the common whelk [52]. Through a structural comparison and phylogenomic analysis, we assess what these sequences can teach us about (mt genomic) evolution in endosymbionts and flatworms as a whole. As is the case in other platyhelminths, we predict these species to possess AT-rich mt genomes [28]. Based on reported variation in other turbellarian orders (*e.g.* [31, 32, 34]), mt genomic structure and gene order is expected to be comparable, but not completely identical to that of *B. personatum* and to be largely conserved intragenerically. Due to the annotation difficulties associated with identifying and annotating *atp8*, its absence in neodermatans and the derived nature of this gene in some taxa [32, 53], we hypothesise a copy of this gene to be present, though highly divergent in the species under study. We expect a corresponding relaxation of selection pressure in *atp8* compared to other mt protein-coding genes (PCGs). We test this formally, using a newly-constructed mitochondrial phylogenetic topology as a backbone.

2. Results

Mt genomes of three different species of endosymbiotic rhabdocoels are presented. One is derived from the type species of *Syndesmis*, *Syndesmis echinorum* François, 1886 (Umagillidae), an endosymbiont of *Echinus esculentus* Linnaeus, 1758 (Echinidae, Echinoidea) [voucher numbers XXXXXX]. A second sequence belongs to its congener *Syndesmis kurakaikina* Monnens, Vanhove & Artois, 2019, a species recently described from the intestine of the New Zealand sea urchin *Evechinus chloroticus* [57] Mortensen, 1943 (Echinometridae, Echinoidea) [voucher numbers XXXXXX]. The third sequence is obtained from *Graffilla buccinicola* (Graffillidae), which was found infecting the common whelk *Buccinum undatum* Linnaeus, 1758 (Buccinidae, Neogastropoda) [voucher numbers XXXXXX].

2.1 Sequencing and assembly

Mt genomes and ribosomal operons were assembled for all three species. The mt genomes derived from *S. kurakaikina*, *S. echinorum* and *G. buccinicola* are 14,226 bp (average coverage 4906X), 15,053 bp (average coverage 1590X) and 14,369 bp (average coverage 269X) long, respectively. Sequences were deposited in GenBank under accession numbers: xx-xxxx, xx-xxxx and xx-xxxx. While circularised contigs were assembled, observed coverage peaks in non-coding regions (NCRs) indicate unresolved (likely repetitive) segments (indicated by disjunctions in **Fig. 1** and **Fig. 2**).

2.2 Mt genome architecture and gene content

Physical maps of the annotated mt genomes of the three investigated species are given in **Fig. 1**. With some exceptions, the newly assembled mt genomes contain all 12 expected protein-coding genes (PCGs), transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). Predicted secondary structures and free energy levels of tRNAs are displayed in **Fig. A1**. All three sequences are AT-rich, specifically in NCRs. Calculated AT and GC skew values for coding strands are denoted in **Table 1**. An overview of codon usage per amino acid is given in **Table 2**. No PCGs with alternative start codons or stop codons were found, though some were predicted to end in an abbreviated stop codon (T). Numerous rearrangements were found between the four rhabdocoel mt genomes (**Fig. 2**), including transpositions and tandem duplications and random losses in each pairwise comparison.

Mt genomes of *S. kurakaikina* and *S. echinorum* are similar in structure (**Fig. 1** and **Fig. 2**), with some differences: *S. echinorum* contains a single, long NCR (699 bp), compared to three shorter ones in *S. kurakaikina* (219, 349 and 424 bp). In *S. echinorum* two copies of *trnL2* were annotated with strong support (MITOS E-values 3.213E-08 and 4.531E-09 respectively), each carrying a different anticodon sequence (TTG and TTA) and appearing at different positions compared with *S. kurakaikina* (indicated by arrow in **Fig. 2**). In *S. kurakaikina*, the annotated *nad6* is shorter than in *S. echinorum* (225 vs 453 bp) and preceded by an NCR. Alignment of this region with *nad6* in other platyhelminths indicates that the true start codon might have been misassembled (alignment not shown) and in the absence of transcriptomic data we therefore consider this gene as partially annotated.

Gene order in *G. buccinicola* is strikingly different compared with that in both umagillids (**Fig. 2**). As in *S. echinorum*, two copies of *trnL2* were identified (anticodons TTG and TTA). A single, repeat-rich NCR (770 bp) is located between *trnE* and *trnP*. Neither of the automatic software tools detected *nad4L* when employing default settings. Two copies of *nad6* were identified by MITOS, each with similar quality scores and corresponding to an open reading frame (ORF) of 454 bp and 303 bp respectively (cfr. 366 bp in *B. personatum*). The shorter *nad6* gene does not align convincingly to its homolog in other platyhelminths (results not shown) and while three transmembrane regions were detected, no NADH dehydrogenase domains were found. As such, we for now refrain from annotating this region.

2.3 Manual annotation

Automatic annotation tools did not confidently identify *atp8* or (in *G. buccinicola*) *nad4L* in the assembled mt genomes. However, *atp8* poses a well-known annotation challenge due to its short length and high variability. As *nad4L* is similarly short and variable, it is not inconceivable that an analogous issue might have occurred. Therefore, we undertook a series of manual annotation steps to detect these genes.

In *S. kurakaikina*, a region of almost 100 bp was identified as *atp8* by MITOS under default settings, albeit with very low support. An ORF of comparable length is located here. In *S. echinorum*, *G. buccinicola* and the previously published sequence of *B. personatum*, several candidate ORFs were selected for consideration: this included all ORFs with similar length in previously reported flatworm *atp8* genes in otherwise non-coding regions and all (low-support) *atp8*-annotations of MITOS and DOGMA, which were found when lowering

stringency criteria. A promising ORF of 165 bp was found in *S. echinorum*, located downstream of *trnW* in an otherwise-unannotated region. Notably, its amino-acid sequence starts with the 'MPQL' signature, characteristic for metazoan *atp8* proteins [29].

Hydrophobicity patterns of the latter ORF and of the MITOS annotation for *S. kurakaikina* are visualised in **Fig. 3** (top left graphs). For comparison, all (putatively) annotated *atp8* genes of previously published flatworm mt genomes are also depicted. The candidate *atp8* genes in *Syndesmis* appear similar to some (*Macrostomum lignano*, *Stenostomum sthenum* and all polyclads), but not all (*Stenostomum leucops* and all triclads) flatworms. Most (putatively) reported *atp8* genes of flatworms encode a signal peptide and at least one transmembrane domain, usually located at the beginning of the sequence [31, 32, 53]. Scanning of candidate ORFs in SMART [59] revealed a transmembrane region at 7-29 bp in each sequence, but no signal peptide or an ATPase domain was detected.

For now, we have putatively annotated these particular ORFs in the mt genomes of *Syndesmis*. BLASTing these or any other flatworm *atp8* genes against the read pool did not reveal the presence of any possible NUMTs (nuclear mitochondrial DNA segments). In *G. buccinicola* and *B. personatum*, no evaluated ORF contained the MPQL signature or a recognisable hydrophobicity pattern. Some did contain one or multiple transmembrane regions, but no ATPase domain or signal peptide were found. Therefore, we refrain from annotating *atp8* in these two mt genomes.

None of the *nad4L* candidates aligned adequately with other flatworm homologs (results not shown) and comparing hydrophobicity signatures did not provide any additional evidence

for presence of this gene. A SMART search did not reveal the presence of NADH dehydrogenase domains [60] in otherwise unannotated regions. As such, for the time being, we do not annotate *nad4L* in *G. buccinicola*.

2.4 Mt phylogeny and selection tests

Alignment of our sequences with all available platyhelminth mt PCGs and rRNAs resulted in a concatenated alignment of 8886 positions. Best-fitting partitioning scheme and corresponding evolutionary models are listed in **Table 3**. After collapsing weakly supported clades, maximum likelihood and Bayesian analyses yielded identical topologies (**Fig. 4**). The RELAX test was significant in endosymbiotic rhabdocoels for the *atp8* ($p = 0.000$) and *cox1* ($p = 0.000$) datasets (**Table 4**) and corresponding selection intensity parameters (k) were estimated at 0.98 and 0.61 respectively. At the 0.05 significance level, the null hypothesis of the RELAX test could not be rejected for any other PCG. RELAX generated a convergence warning for *atp8*, *cox2*, *cytb*, *nad4* and *nad5*. For these genes, likelihood-ratios (LRs) and differences in AICc (corrected Akaike Information Criterion) between RELAX null and alternative models were considered more reliable than calculated p-values (Pers. Comm. software developer). For *atp8*, the model including a selection intensity parameter K yielded a lower AICc score ($\Delta AICc = 17.4$). With $K < 1$, this gene is therefore considered to evolve under relaxed selection pressure. For the four other alignments, a better fit was recovered for the null model ($\Delta AICc < 0$).

3. Discussion

Despite the diversity and species-richness of Rhabdocoela, only a single mt genome of this group has been characterised prior to this study. As well as adding to the characterisation of the overall molecular diversity in this order, filling this gap in knowledge also allows for in-depth comparisons between closely related species with divergent life history strategies. This study is focused on the particular case of a free-living versus an endosymbiotic way of life. Ultimately, this will contribute to our understanding of the molecular changes when these different ways of life evolve. As a first step towards this, the first mt genomes of endosymbiotic rhabdocoels are characterised here. These are compared among each other and to the single known sequence of a free-living rhabdocoel. In addition, the newly acquired sequences are used in a phylogenetic analysis. Not only does this allow for situating these species in the platyhelminth tree of life, it also enables us to assess selection pressure on the different mt markers under study.

3.1 Mitogenomic composition and structure

The three new mt sequences have the typical flatworm mt genome architecture. Each consists of a single, circular molecule of 14-16 kb and follows a ‘standard’ bilaterian pattern [29]. As is mostly the case in platyhelminths, all genes are transcribed from the same strand and PCGs are translated according to the flatworm and echinoderm mt genetic code [18]. In most cases, all 37 genes are present, including 22 tRNAs, a large and a small ribosomal RNA gene (*rrnL* and *rrnS*), 12 PCGs associated with oxidative phosphorylation, a putative *atp8* gene and at least one NCR [29]. Some exceptions to this pattern were observed, which will be discussed below.

Most PCGs employ canonical start and stop codons, but some are predicted to end in abbreviated stop codons (T), a common phenomenon in metazoans [61]. These sequences are presumed to be completed by polyadenylation after transcription [62]. Truncated stop codons have been reported in flatworms as diverse as catenulids [63], cestodes [53], macrostomorphs [64], monogeneans [65, 66], as well as in *B. personatum* [34]. Completion through posttranscriptional polyadenylation has been confirmed experimentally in the tapeworm *Taenia solium* Linnaeus, 1758 (Taeniidae, Cestoda) [63]. In our sequences and in *B. personatum* [34], only the single base (T) stop codons occur: the alternative TA codon is not observed in any rhabdocoel mt genome.

The newly assembled genomes are characterised by a positive GC skew and a negative AT skew in the coding strand (**Table 1**). This proportion is reversed in most metazoans, but corresponds to what has been published by Castellana, Vicario and Saccone [67] in other flatworms. The skew patterns calculated for our sequences predict a preference for amino acids encoded by GT-rich codons [68], which corresponds to our observations, in particular at the third codon position (**Table 2**).

3.2 Mitochondrial phylogeny

ML and BI analyses of the mt genomic dataset (**Fig. 4**) recovered Rhabdocoela as a monophyletic group (pp = 100, UFBoot = 93.8, SH-aLRT = 79). All three newly obtained sequences cluster together, constituting a well-supported, monophyletic neodalyellid clade (pp = 100, UFBoot = 100, SH-aLRT = 100). Within this group, both species of *Syndesmis* appear as sister taxa (pp = 100, UFBoot = 100, SH-aLRT = 100). Branches are relatively

long for all three newly added sequences. High evolutionary rates have previously been linked to an endosymbiotic or parasitic lifestyle in other metazoans (*e.g.* [69-73]). However, care must be taken to interpret our results in this way as these long branches might also be an artefact related to the still incomplete taxon coverage of rhabdocoels.

Notably, one of the deeper clades in our tree differs from recent flatworm topologies [74, 75]. The clade encompassing all rhabdocoels, triclads and polyclads is well supported in our topology (pp = 100, UFboot = 99.8, SH-aLRT = 80). Taxon coverage for turbellarians is, however, still incomplete and a more representative, densely sampled dataset is imperative before drawing conclusions on the true flatworm tree of life.

3.3 *Atp8* in rhabdocoels and flatworms in general

A point that has attracted considerable attention amongst metazoan molecular systematists, is whether or not platyhelminth mt genomes carry an *atp8* gene [31]. Indeed, *atp8* could not be confidently annotated in any of our sequences through the use of automatic annotation tools. It is well established that neodermatans have lost this gene in the course of evolution [27, 28] and, for a long time, it was assumed that the same was true for all flatworms [29, 30]. More recently published studies have proposed that *atp8* is, in fact, present in all turbellarians, but had previously not been picked up by standard annotation methods [*e.g.* 31, 32].

Besides the fact that *atp8* is highly divergent, difficulties in identifying this gene can, at least in part, be attributed to the fact that *atp8* tends to be particularly short: in most representatives of Lophotrochozoa, *atp8* - if identified at all - is only between 100 and 200

bp long. Considering both MITOS and DOGMA calculate quality scores that are directly derived from E-values of BLASTX hits, which are directly dependent on query lengths, it is logical that these programs encounter difficulties when annotating shorter genes.

Recently, Egger, Bachmann and Fromm [31] identified *atp8* in novel and previously published turbellarian sequences by integrating transcriptomic data in their analysis and through manual curation procedures. Following a similar bio-informatic pipeline (excluding steps relating to transcriptomic data), candidate *atp8* genes were selected and evaluated in some of the rhabdocoel sequences, with varying degrees of confidence. However, at least in *G. buccinicola* and *B. personatum,* these results are ambiguous at best. Indeed, it seems almost impossible to objectively annotate the gene in these mt genomes without additional data, such as those resulting from transcriptomic or proteomic studies.

Because of the annotation challenges inherent to such high degrees of variation, it is not surprising that *atp8* was once presumed to have been lost in all platyhelminths [76]. While we agree with Egger, Bachmann and Fromm [31] that manual curation steps are necessary in any annotation procedure, application of their methodology did not provide unambiguous results for rhabdocoel mt genomes, most notably in the case of *G. buccinicola* and *B. personatum*. This raises the question how confident one can truly be when inferring and annotating *atp8* based solely on genomic data and, for the time being, we would advise interpreting past and future annotations of this gene with caution, especially if no additional evidence is available.

Even disregarding the more problematic cases in rhabdocoels, the combined results from [31, 32, 53] already show the amount of variation displayed in the *atp8* gene. Illustrative examples include the (putative) annotations in triclads, where the gene is almost five times longer than in most other flatworms and the length polymorphisms within the catenulid genus *Stenostomum*, where the two characterised species apparently possess an *atp8* gene differing by more than 750 bp in length. Moreover, even the supposedly conserved 'MPQL' sequence at the N-terminal of the gene shows considerable variation in turbellarians (**Table 5**).

Such divergence raises the question to what extent such a gene can be functional and whether or not the ATP-synthase complex can operate without it. Indeed, genes involved in this complex seem to be lost – and acquired – with surprising ease in metazoans and *atp8* itself has been suggested to be dispensable altogether [29]. This corresponds with its absence in neodermatans and with the finding that *atp8* is one of the few genes under relaxed selection pressure in *Syndesmis*. Paradoxically, *atp8* has been shown to play a crucial role in assembly of the ATP-synthase complex in yeast [78] and mammals [79]. As almost nothing is known of the cellular mechanisms in turbellarians, we refrain from predicting any phenotypical effects of a dysfunctional *atp8* gene.

3.4 Notes on protein-coding genes in endosymbiotic rhabdocoels

In addition to the case of *atp8* described above, *cox1* is also under relaxed selection pressure in endosymbiotic rhabdocoels, when compared with all other flatworms. Moreover, in *G. buccinicola*, no *nad4L* was detected, even after manual curation steps. This gene is known to be rather short and highly variable in flatworms [25], imposing annotation

difficulties similar to *atp8* (see above). Flatworm mt PCGs are also known to occasionally employ alternative start codons [32, 80, 81], hence some candidate ORFs might have been overlooked in our analyses. The fact that assemblies are interrupted in the NCR can also imply that some ORFs were not recovered.

3.5 Rhabdocoela display large genetic rearrangements

Gene order was once assumed to be highly conserved in metazoans [82]. As more and more flatworm taxa are covered in genomic work, it is becoming increasingly clear that gene order has shifted frequently in the evolutionary history of these animals [31, 34, 53, 83]. Our results further reinforce this view: all three mt genomes exhibit a unique gene order (**Fig. 2**) and not a single gene block is conserved between the three rhabdocoel genera.

Nad4 and *nad4L* genes are no longer positioned next to one another in either of the three sequences [31], *nad4L* possibly being completely absent in *G. buccinicola.* In flatworms, *nad4* and *nad4L* predominantly appear as an uninterrupted gene pair (as inferred from the mt genomes available on RefSeq, using '(mitochondrion AND Platyhelminthes[organism])' as a search query), and these genes are generally assumed to constitute a transcriptional unit in metazoans [84]. The same is true for the (putative) *atp8-atp6* pair in *Syndesmis*, which is, however, also disrupted in other turbellarians (see Fig. 1 in [31] for an overview). Separation of these genes seems to imply that *nad4-nad4L* and *atp8-atp6* are no longer transcribed simultaneously in these animals, though transcriptomic data are necessary to validate this finding.

Even within *Syndesmis*, gene order is not fully conserved, as indicated in **Fig. 2** with a rearrangement of the *trnaL1* and *trnaL2* block. To date, there are only two other turbellarian genera for which more than one mt genome has been characterised: *Stenostomum* Schmidt, 1848 (Stenostomidae, Catenulida) and *Imogine* Marcus & Marcus, 1968 (Stylochidae, Polycladida). In the former genus, sequences differ in the orientation of the *cytb-nad6-cox3-nad1-rrnL* block and in the fact that *nad3* and *rrnS* have been duplicated in *S. sthenum.* In *Imogine*, no change in gene order is apparent, as the single difference between the sequenced species is the presence or absence of the *trnaK* gene.

These cases of intrageneric variation are even more noteworthy when comparing these results with Tricladida, where gene order is conserved, even when comparing species separated by much larger evolutionary distances [32]. Notably, intrageneric rearrangements have also been reported on a large scale within the neodermatan *Schistosoma* Weinland, 1858 (Schistosomatidae, Trematoda [87]). In other metazoans, an increased frequency of mt rearrangements has been proposed to correlate with a parasitic lifestyle (*e.g.* [71, 88, 89]). Evaluation of these hypotheses in flatworms requires a more extensive mt genomic dataset of parasitic (or endosymbiotic) taxa and their free-living counterparts. Furthermore, while mt gene order within *Syndesmis* differs only in a single tRNA transposition, our observations also lend further support to the conclusions of Le, et al. [87], that the use of gene order as a phylogenetically relevant trait should be implemented with caution, and, ideally, with inclusion of multiple representatives of each taxon.

4. Conclusions and future perspectives

To date only a single mt genome of rhabdocoels had been published. The sequences presented here increase the available mt genomic data of this highly understudied group fourfold and comprise the first set of complete nuclear, ribosomal operons. Public availability of these data may accelerate sequencing efforts for closely related rhabdocoels, although the vast molecular diversity that this group appears to hold must be considered in future research. Indeed, as more non-neodermatans are included in mitogenomic work, it is becoming clear that flatworm mt genomes are highly dynamic. The sections above have highlighted some unexpected features in mt encoded genes of endosymbiotic rhabdocoel mt genomes, further demonstrating the high degree of (mt) evolutionary genetic plasticity in flatworms. A notable example is the possible absence/dysfunctionality of the *atp8* gene. On the one hand, our findings confirm the conclusion of Egger, Bachmann and Fromm [31], that annotation of highly variable genes requires manual curation. However, without additional evidence from transcriptomic or even proteomic datasets, it remains difficult to unambiguously state that *atp8* is either present or absent in flatworm mt genomes. In addition to the case of *atp8*, several other observations confirm the genetic variability of flatworm (mt) genomes, including relaxed selection pressure on several PCGs related to oxidative phosphorylation. These observations demonstrate the high degree of genomic flexibility in these endosymbionts, or perhaps even in rhabdocoels in general.

While functional experimental approaches are commendable to verify these findings, it is already thought-provoking to place some of these findings in the light of the study of Zarowiecki and Berriman [90]: their meta-analysis of sequenced genomes showed that parasitic flatworms tend to lose metabolic capacities. While these authors focused solely on

Neodermata, a similar trend seems to emerge among endosymbiotic rhabdocoels, illustrated by the seemingly missing *nad4L* and *atp8* genes, in comparison with the more 'standard' mt genome of *B. personatum*. For the *cox1* and *atp8* genes, this is further demonstrated by the significantly relaxed selection pressure in the endosymbionts, compared with their free-living counterparts. Studies on other, more distantly related organisms have suggested similar mechanisms. For instance, Skippington, Barkman, Rice and Palmer [91] demonstrated that several respiratory genes have been lost from the mt genome of hemiparasitic mistletoe and it has been suggested that this reduction may be related to the parasitic lifestyle of this particular group. More mt genomic data is needed to further investigate these claims, and an assessment of a possible link between lifestyle and gene loss will require a thorough comparison of these endosymbionts and their free-living sister taxa.

5. Material and methods

5.1 Sampling

Targeted host species were collected by either dredging or diving. *Echinus esculentus* was collected by dredging in the Gullmar fjord near the Sven Lovèn Centre in Kristineberg, Sweden (summer 2017). *Buccinum undatum* was acquired at the same location by diving (autumn 2017). Specimens of *Evechinus chloroticus* were collected by free-diving in Matheson's Bay in New Zealand (winter 2017). Endosymbionts were obtained from their marine invertebrate hosts by dissection and stored in 99% EtOH for downstream molecular work.

5.2 DNA extraction and sequencing

Total genomic DNA was extracted from each sample using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's standard protocols for tissue. The amount of double-stranded DNA was measured with the Qubit® 2.0 Fluorometer (Life Technologies, Paisley, UK) and indexed libraries were prepared using the Nextera Flex DNA Sample Preparation Kit (Illumina, Inc., San Diego, USA) from 10ng of input DNA. Libraries were run at the DNA Sequencing Facility of the Natural History Museum, London, UK and run simultaneously on a NextSeq 500 sequencing platform (Illumina) using a 300 cycle Mid-output kit. Demultiplexed reads were deposited in the NCBI Sequence Read Archive (ncbi.nlm.nih.gov/sra) under accession numbers xxx.

5.3 Assembly

Read quality was assessed with FastQC v0.11.5 [92]. Mt genomes were assembled using NOVOPlasty v2.7.2 [93]. Initially, the single other available rhabdocoel mt genome (accession number: MF993329), was specified as a reference sequence to guide the assembly. However, as this approach yielded no usable results for any of our read pools, we opted for the *de novo* seed-and-extend algorithm implemented in NOVOPlasty, using a partial *cox1* coding sequence of *Anoplodium hymanae* Shinn, 1983 (Umagillidae) as a seed (accession number: MG256112). As this did not produce a circular sequence for *G. buccinicola*, the *cytb* gene of the mt genome of *S. kurakaikina* was used for this purpose. K-mer values were adjusted stepwise until a single, circular contig was obtained (25 in *S. echinorum,* 30 in *S. kurakaikina,* 60 in *G. buccinicola*).

Coverage depth was assessed with the assembler tool in Geneious v11.1.5 [95], remapping reads against assembled contigs. Custom sensitivity settings were employed: permitting 4% mismatches/read, requiring a minimum overlap of 25 bp and disallowing gaps. Assemblies were checked for contamination through a BLAST search [96] on the NCBI website (ncbi.nlm.nih.gov). Compositional differences between strands were quantified using the GC and AT skew indices [97], by applying the respective formulas [(G-C)/(G+C)] and [(A-T)/(A+T)] on the coding strand of each mt genome.

5.4 Annotation

Automatic annotation was performed using the MITOS and DOGMA online servers, employing the echinoderm and flatworm mt code. Results were matched to ORFs, as identified by Geneious, using default settings and the same genetic code. PCG boundaries were subsequently tweaked by aligning newly obtained sequences to other flatworm mt genes with the MUSCLE v3.8.425 [98] executable, implemented in Geneious. An overview of all sequences used for comparison is given in **Table A1**. Codon frequencies were calculated with the Codon Usage tool in the Sequence Manipulation Suite (bioinformatics.org/sms2/codon_usage.html) [99]. Relative Synonymous Codon Usage (RSCU [100]) was calculated with the CAI python package of Lee [101]. To identify tRNAs, MITOS (MiTFi) results were supplemented with ARWEN v1.2.3 [102], using the -gcflatworm setting, and an online search with tRNAscan-SE v2.0 [103]. In case of conflicts between different software packages, we adopted the solution providing no intergenic overlap, minimising non-coding regions between genes, and (for tRNAs) proposing a 7 bp acceptor stem. Free energy of secondary structures was calculated in the online version of RNAeval [104] (Vienna package [105]) under 17°C (*S. echinorum*), 16°C (*S. kurakaikina*), or 10°C

(*G. buccinicola*) as derived from one-week averages in the Gullmar fjord (from weather.mi.gu.se) and direct measurement in Matheson's Bay upon time of collection. Sequences were scanned for repeats with Geneious (using 5% cut-off in the 'Repeat Finder' plugin) and on the YASS [106] and Tandem Repeats Finder [107] web portals. Most parsimonious genetic rearrangements were inferred with CREx [108]. Annotated mt genomes were visualised in OrganellarGenomeDRAW [109, 110].

In *S. kurakaikina*, initial assembly and annotation uncovered a frameshift in the *cox1* gene at ±200 bp. As physiological implications seem severe, this result was doublechecked through PCR and Sanger sequencing (Macrogen Europe) of this region in four additional specimens (newly designed primers 5'-GTCGCCCTTTAGTAAGCTT and 5'-ATAGTCCAACCAGCCGATA; 98°C 1m, 9X touchdown (98°C 30s, 56°C→50°C 20s, 72°C 30s), 35X (98°C 10s, 49°C 20s, 72°C 30s), 72°C 10m). As the mutation was not confirmed, we consider this initial finding a technical artefact.

5.5 Searching *atp8* and *nad4L*

Atp8 was not readily detected by MITOS or DOGMA, except when severely lowering identity cut-off values. The single exception is *S. kurakaikina*, where the gene was identified by MITOS using default settings. However, this annotation came with low-quality scores (a factor of 10^4 to 10^5 lower than most other PCGs) and accepting this prediction without further validation seemed, therefore, unfit. Similarly, no *nad4L* gene was detected in *G. buccinicola*. However, due to its small size and high variability, *atp8* is known to cause difficulties for automatic annotation tools and, to a lesser degree, the same characteristics can be attributed to *nad4L*. Therefore, a series of extra steps were undertaken to verify the

MITOS prediction in *S. kurakaikina* and to detect the missing genes in the other two sequences. The same procedure was followed to trace *atp8* in the previously published (linear) mt genome of *B. personatum* (accession number: MF993329). We here followed a bioinformatic pipeline similar to [31].

Geneious was used to predict candidate ORFs (transl_table 9) of approximately 100 to 900 bp (for *atp8* [31, 32, 53]) and 240 to 300 bp (for *nad4L*, based on other flatworm mt genomes on RefSeq, **Table A1**). We also included ORFs with the alternative start codons TTG and ATT, as these have been reported to occasionally occur in platyhelminths [32, 80, 81]. We also included the two *nad6* annotations as candidates for *nad4L* in *G. buccinicola*, as these might have been misidentified by MITOS.

Using MUSCLE, translated ORFs were aligned with a dataset of all predicted (hypothetical) *atp8* sequences from [31, 32, 53] (**Table A2**). Similarly, translations of other available flatworm *nad4L* genes were used for comparison (**Table A1**). Hydrophobicity profiles were computed with the ExPASY ProtScale tool [111, 112], employing default settings and the Kyte and Doolittle scale [113]. In *atp8*, special attention was paid to the first four amino acids, as this region is considered most conserved. SMART was used to scan each ORF for the presence of ATPase and NAD-binding domains, signal peptides and transmembrane regions. To check for potential NUMTs, putative genes were BLASTed against the respective SPAdes assemblies they were obtained from.

5.6 Nuclear ribosomal operons

In addition to reconstruction of the mt genomes, ribosomal operons were baited from the read pools. *De novo* assembly was carried out in SPAdes v3.13.0 [114], using the infrastructure of the VSC (Flemish Supercomputer Centre). K-mer lengths were set at 21, 33, 55, 77, 99 and 127. Resulting de Bruijn graphs were visualised with Bandage v0.8.1 [115]. 18S rDNA of *Syndesmis aethopharynx* (accession number: MF574100), was subjected to a BLAST search against the assembly and contigs carrying hits were extracted from the pool. To rule out contamination by host (or other) DNA, each selected sequence was BLASTed on the NCBI webserver.

To predict the respective positions of 18S, 28S and 5.8S rRNA, each sequence was uploaded onto RNAmmer v1.2 [116]. Raw fasta files were fed into ITSx v1.0.11 [117] to identify boundaries of ITS1 and ITS2 regions. For all three sequences, software encountered difficulties in pinpointing the boundary between the 5.8S and ITS2 fragments. To overcome this, and to further verify the predictions proposed by RNAammer, a series of multiple alignments was conducted, comparing our newly obtained sequences with available ribosomal operons of other flatworms. To compile this dataset for comparison, the following search query for complete operons was launched on GenBank: ‘Platyhelminthes[organism] AND ((18S OR "small subunit") AND (28S OR "large subunit") AND (ITS OR "internal transcribed spacer"))’. Alignments were made on the MAFFT v7 online server [118], employing the Q-INS-i algorithm, which accounts for secondary structures, followed by visual inspection.

5.7 Multimarker phylogeny

A dataset was compiled from all flatworm mt genomes available on GenBank. Two gnathostomulids were selected as outgroups [74] (**Table A1**). Both rDNA genes and 12 PCGs were selected as markers. Due to its problematic annotation in turbellarians and its absence in Neodermata (see Introduction), *atp8* was omitted from this analysis. MUSCLE, as implemented in Geneious, was used to compute a codon-based alignment of each PCG. Ribosomal genes were aligned using MAFFT, employing the Q-INS-i algorithm. The online version of Gblocks v0.91b [119] was used to eliminate problematic regions from all obtained alignments, specifying options for a less stringent selection [120]. The 14 trimmed alignments were concatenated in Geneious.

An initial partitioning scheme was constructed, subdividing the concatenated alignment in genes and (for PCGs) codon positions. Alignment and partition file were then fed into the ModelFinder tool of IQ-TREE [121], enabling partition merging [122]. The latter feature determines the best-fit partitioning scheme for a particular alignment, while also calculating the most suitable evolutionary models for each selected subset. Model fit was evaluated with the Bayesian Information Criterion and the best fit partially-merged partitioning scheme was specified in maximum likelihood (ML) and Bayesian (BI) model-based analyses (**Table 3**). Gaps were treated as missing data. MrBayes v3.2.6 [123] was run on XSEDE on the CIPRES Science Gateway [124]. Two parallel analyses were carried out simultaneously for 10 million generations, following the Metropolis coupled Markov chain Monte Carlo algorithm. Each run consisted of one cold and three heated chains and trees were sampled every 1000th generation. Convergence was assumed once split frequencies dropped beneath 0.01 and Potential Scale Reduction Factors approached 1. The initial 25% of

inferred trees were discarded as burn-in, and remaining topologies were summarised in a 50% majority-rule consensus tree. Node support was evaluated with posterior probabilities (pp).

The ML phylogeny was estimated with the stochastic tree reconstruction algorithm [125] on the W-IQ-TREE server [126] An edge-linked partition model was specified, allowing proportional branch lengths. Support was assessed through 1000 ultrafast bootstraps (UFboot) [127] and the SH-aLRT statistic (Shimodaira-Hasegawa approximate likelihood-ratio test [128]). Inferred BI and ML phylogenies were visualised and rooted in TreeGraph v2 [129] and weakly supported clades (pp < 95%) were collapsed.

5.8 Test for relaxed selection pressure

Using the inferred topology as a backbone, nucleotide alignments of all PCGs were subjected to the RELAX test [130]. RELAX tests for changes in selective strength in a predefined subset of branches (foreground), compared with the remaining reference branches in the topology (background). RELAX computes a null model where selection pressure is kept constant between foreground and background lineages, and an alternative model allowing selective intensification and relaxation. The latter is then compared with the null model through a likelihood-ratio test and AICc values.

As RELAX only supports usage of a single genetic code, catenulids and gnathostomulids were cut from the dataset, as these taxa employ the invertebrate mt code [18]. For *atp8*, a truncated dataset was used, including only those species of which a (putative) *atp8* gene has been annotated (**Table A2**). Neodermata were completely excluded here, as *atp8* is not

known from this taxon. As we found multiple candidate *atp8* genes in *B. personatum* and *G. buccinicola*, none of which could be annotated with adequate confidence, these species were also omitted from this analysis. Final datasets were aligned and processed using MUSCLE and Gblocks, with the same settings described above. All removed taxa were also pruned from the inferred topology in Mesquite v3.51 [131]. Topology and alignments were then used as input for the RELAX tool on the Datamonkey v2.0 interface [132]. Designated test branches are marked in green in **Fig. 4**, and the echinoderm/flatworm mt code was specified.

In five PCGs, RELAX produced a convergence warning impeding reproducibility of the results. Convergence issues can occur when handling small test datasets and/or low divergence alignments. Following the developer's suggestions, the same test was run locally in RELAX v3.1 under a minimal model and using fewer rate classes, minimising the risk of overparameterization. As the issue persisted, we followed the developer's recommendation to consider LRs (and corresponding AICc values) instead of calculated p-values for these datasets (Pers. Comm.).

Acknowledgements

We thank an anonymous reviewer for their careful reading and recommendations for improvement of this work. Dr Geert Jan Bex is cordially thanked for his support in VSC related topics and Dr Stephan Greiner for his helpful advice on use of OGDRAW. The authors are grateful to the staff at the Sven Lovén Centre for Marine Sciences in Kristineberg and Tjärnö for hosting M.M. during her research stays in Sweden and to Dr Ulf Jondelius for organising sampling campaigns here and his help in species collection and

identification. Dr Sam Dupont is thanked for donating his specimens of *Echinus esculentus*, Dr Maria Asplund for host collection by diving and Lisa Huijgen for her valuable assistance in the lab in Tjärnö. Finally, we wish to thank Jeff Vandeweyer for his help in troubleshooting bioinformatic work.

Funding sources

This research was supported by the Research Foundation – Flanders (FWO, grant number 1141817N to M.M.). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the FWO and the Flemish Government – department EWI. The research was carried out with infrastructure funded by the European Marine Biological Resource Centre (EMBRC Belgium - FWO project GOH3817N). Dredging in the Gullmar fjord was financially supported by Dr Andreas Hejnol. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] C.S. Minot, Studien an Turbellarien. Beitrage zur Kenntnis der Plathelminthen, Arbeiten aus dem Zoologisch-Zootomischen Institut in Würzburg, III 3 (1876) 405-471, Fig 16-20.

[2] T.M. Goater, C.P. Goater, G.W. Esch, Platyhelminthes: the flatworms, in: T.M. Goater, C.P. Goater, G.W. Esch (Eds.), Parasitism: the diversity and ecology of animal parasites, Cambridge University Press, Cambridge, United Kingdom, 2014.

[3] J.N. Caira, D.T.J. Littlewood, Worms, Platyhelminthes, in: S.A. Levin (Ed.), Encyclopedia of Biodiversity, Volume 7, Academic Press, Waltham, Massachusetts, United States of America, 2013, pp. 437-469.

[4] T. Artois, D. Fontaneto, W.D. Hummon, S.J. McInnes, M.A. Todaro, M.V. Sørensen, A. Zullini, Ubiquity of microscopic animals? Evidence from the morphological approach in species identification, in: D. Fontaneto (Ed.), Biogeography of microscopic organisms: is everything small everywhere?, Cambridge University Press, Cambridge, United Kingdom, 2011, pp. 244-283.

[5] E.R. Schockaert, M. Hooge, R. Sluys, S. Schilling, S. Tyler, T. Artois, Global diversity of free living flatworms (Platyhelminthes, "Turbellaria") in freshwater, Hydrobiologia 595 (2008) 41-48.

[6] D.T.J. Littlewood, The evolution of parasitism in flatworms, in: A.G. Maule, N.J. Marks (Eds.), Parasitic flatworms: molecular biology, biochemistry, immunology and physiology, Biddles Ltd, King's Lynn, United Kingdom, 2006, pp. 1-36.

[7] V.G. Fonseca, G.R. Carvalho, W. Sung, H.F. Johnson, D.M. Power, S.P. Neill, M. Packer, M.L. Blaxter, P.J.D. Lambshead, W.K. Thomas, S. Creer, Second-generation environmental sequencing unmasks marine metazoan biodiversity, Nat Commun 1 (2010).

[8] O. Giere, Meiofauna taxa: a systematic account, Meiobenthology. The microscopic motile fauna of aquatic sediments, Springer-Verlag, Berlin Heidelberg, Germany, 2009.

[9] S. Lustigman, R.K. Prichard, A. Gazzinelli, W.N. Grant, B.A. Boatin, J.S. McCarthy, M.-G. Basáñez, A research agenda for helminth diseases of humans: the problem of helminthiases, PLoS Negl Trop Dis 6(4) (2006) e1582.

[10] D.T.J. Littlewood, Platyhelminth systematics and the emergence of new characters., Parasite 15 (2008) 333-341.

[11] N. Van Steenkiste, Phylogeny of Dalytyphloplanida (Platyhelminthes: Rhabdocoela): single escape from the marine environment? Zoology: Biodiversity and Toxicology, Hasselt University, Diepenbeek, Belgium, 2012.

[12] E.R. Schockaert, Turbellarians: the importance of turbellarians in ecosystems, in: G. Hall (Ed.), Methods for the examination of organismal diversity in soils and sediments, CAB International in association with UNESCO and the IUBS, Wallingford, UK, 1996, pp. 211-225.

[13] J. Wit, J.S. Gilleard, Resequencing helminth genomes for population and genetic studies, Trends Parasitol 33(5) (2017) 388-399.

[14] C. Hahn, B. Fromm, L. Bachmann, Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparastic Neodermata, Genome Biol Evol 6(5) (2014) 1105-1117.

[15] D.A. Johnston, Genomes and genomics of parasitic flatworms, in: A.G. Maule, N.J. Marks (Eds.), Parasitic flatworms: molecular biology, biochemistry, immunology and physiology, Biddles Ltd, King's Lynn, United Kingdom, 2006, pp. 37-80.

[16] S. Lorch, D. Zeuss, R. Brandl, M. Brändle, Chromosome numbers in three species groups of freshwater flatworms increase with increasing latitude, Ecol Evol 6(5) (2016) 1420-1429.

[17] M.A. Grohme, S. Schloissnig, A. Rozanski, M. Pippel, G.R. Young, S. Winkler, H. Brandl, I. Henry, A. Dahl, S. Powell, M. Hiller, E. Myers, J.C. Rink, The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms, Nature 554 (2018) 56-61.

[18] 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 U S A 97(21) (2000) 11359-11364.

[19] S. Carranza, G. Giribet, C. Ribera, J. Baguña, M. Riutort, Evidence that two types of 18S rDNA coexist in the genome of *Dugesia* (*Schmidtea*) *mediterranea* (Platyhelminthes, Turbellaria, Tricladida), Mol Biol Evol 13(6) (1996) 824-832.

[20] M.P.M. Vanhove, B. Tessens, C. Schoelinck, U. Jondelius, D.T.J. Littlewood, T. Artois, T. Huyse, Problematic barcoding in flatworms: A case-study on monogeneans and rhabdocoels (Platyhelminthes), Zookeys 265 (2013) 355-379.

[21] J. Geller, C. Meyer, M. Parker, H. Hawk, Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys, Mol Ecol Resour 13(5) (2013) 851-861.

[22] N. Van Steenkiste, S.A. Locke, M. Castelin, D.J. Marcogliese, C.L. Abbott, New primers for DNA barcoding of digeneans and cestodes (Platyhelminthes), Mol Ecol Resour 15 (2015) 945-952.

[23] D. Sanna, T. Lai, P. Francalacci, M. Curini-Galletti, M. Casu, Population structure of the *Monocelis lineata* (Proseriata, Monocelididae) species complex assessed by phylogenetic analysis of the mitochondrial Cytochrome c Oxidase subunit I (*COI*) gene, Genet. Mol. Biol. 32(4) (2009) 864-867.

[24] A. Moszczynska, S.A. Locke, J.D. McLaughlin, D.J. Marcogliese, T.J. Crease, Development of primers for the mitochondrial cytochrome c oxidase I gene in digenetic trematodes (Platyhelminthes) illustrates the challenge of barcoding parasitic helminths, Mol Ecol Resour 9(Suppl. 1) (2009) 75-82.

[25] T.H. Le, D. Blair, D.P. McManus, Complete DNA sequence and gene organization of the mitochondrial genome of the liverfluke, *Fasciola hepatica* L. (Platyhelminthes; Trematoda), Parasitology 123 (2001) 609-621.

[26] T.H. Le, M.S. Pearson, D. Blair, N. Dai, L.H. Zhang, D.P. McManus, Complete mitochondrial genomes confirm the distinctiveness of the horse-dog and sheep-dog strains of *Echinococcus granulosus*, Parasitology 124 (2002) 97-112.

[27] M. Hardman, L.M. Hardman, Comparison of the phylogenetic performance of neodermatan mitochondrial protein-coding genes, Zool Scr 35(6) (2006) 655-665.

[28] D.P. McManus, T.H. Le, D. Blair, Genomics of parasitic flatworms, Int J Parasitol 34 (2004) 153-158.

[29] C. Gissi, F. Iannelli, G. Pesole, Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species, Heredity (Edinb) 101 (2008) 301-320.

[30] L. Podsiadlowski, A. Mwinyi, P. Lesný, T. Bartolomaeus, Mitochondrial gene order in Metazoa - theme and variations, in: J.W. Wägele, T. Bartolomaeus (Eds.), Deep metazoan phylogeny: the backbone of the tree of life. New insights from analyses of molecules, morphology, and theory of data analysis, De Gruyter, Inc., Berlin, Germany, 2014, pp. 459-472.

[31] B. Egger, L. Bachmann, B. Fromm, *Atp8* is in the ground pattern of flatworm mitochondrial genomes, BMC Genomics 18(414) (2017) 1-10.

[32] E. Ross, D. Blair, Guerrero-Hernández, A.S. Alvarado, Comparative and transcriptome analyses uncover key aspects of coding- and long noncoding RNAs in flatworm mitochondrial genomes, G3 (Bethesda) 6(5) (2016) 1191-1200.

[33] C.G. Ehrenberg, Phytozoa Turbellaria africana et asiatica, in: C.G. Ehrenberg (Ed.), Hemprich und Ehrenberg "Symbolae physicae." Animalia evertebrata exclusis insectis. Series prima cum tabularum decade prima., Berolini, Fol. Phytozoa Turbellaria 1831, pp. folia a-d, T4-5.

[34] N.J. Kenny, C. Noreña, C. Damborenea, C. Grande, Probing recalcitrant problems in polyclad evolution and systematics with novel mitochondrial genome resources, Genomics 111(3) (2019) 343-355.

[35] E.O. Schmidt, Die rhabdocoelen Strudelwürmer (Turbellaria Rhabdocoela) des süssen Wassers, Friedrich Mauke, Jena, Germany, 1848.

[36] M. Braun, Die Rhabdocoeliden Turbullarien Livlands, Archiv für die Naturkunde Liv-, Ehst- und Kurlands. Zweite Serie. Biologische Naturkunde X(2) (1885) 131-251.

[37] C. Noreña, A. Eitam, L. Blaustein, “Microturbellarian” flatworms (Platyhelminthes) from freshwater pools: New species and records from Israel, Zootaxa 1705 (2008) 1-20.

[38] J.-K. Park, K.-H. Kim, S. Kang, W. Kim, K.S. Eom, D.T.J. Littlewood, A common origin of complex life cycles in parasitic flatworms: evidence from the complete mitochondrial genome of *Microcotyle sebastis* (Monogenea: Platyhelminthes), BMC Evol Biol 7(11) (2007) 1-13.

[39] S. Fontenla, G. Rinaldi, P. Smircich, J.F. Tort, Conservation and diversification of small RNA pathways within flatworms, BMC Evol Biol 17(215) (2017).

[40] D.T.J. Littlewood, T.H. Cribb, P.D. Olson, R.A. Bray, Platyhelminth phylogenetics - a key to understanding parasitism?, Belg J Zool 131(Supplement 1) (2001) 35-46.

[41] E.M. Perkins, S.C. Donnellan, T. Bertozzi, I.D. Whittington, Closing the mitochondrial circle on paraphyly of the Monogenea (Platyhelminthes) infers evolution in the diet of parasitic flatworms., Int J Parasitol 40(11) (2010) 1237-1245.

[42] International Helminth Genomes Consortium, Comparative genomics of the major parasitic worms, Nat Genet 51 (2019) 163-174.

[43] K. Rohde, The origins of parasitism in the Platyhelminthes, Int J Parasitol 24(8) (1994) 1099-1115.

[44] G. Doignon, T. Artois, Annotated checklist of the umagillid turbellarians infesting echinoids (Echinodermata), Belg J Zool 136(1) (2006) 101-106.

[45] N. Van Steenkiste, B. Tessens, W. Willems, T. Backeljau, U. Jondelius, T. Artois, A comprehensive molecular phylogeny of Dalytyphloplanida (Platyhelminthes: Rhabdocoela) reveals multiple escapes from the marine environment and origins of symbiotic relationships, PLoS One 8(3) (2013) 1-13.

[46] B. Wahl, Beitrage zur Kenntnis der Dalyelliiden und Umagilliden, Gustav Fischer, Jena, Germany, 1910.

[47] J. Meixner, Beitrag zur Morphologie und zum System der Turbellaria-Rhabdocoela: 2. Über *Typhlorhynchus nanus* Laidlaw und die parasitischen Rhabdocölen nebst nachträgen zu den Calyptorhnchia, Zeitschrift fur Morphologie und Okologie der Tiere 5 (1926) 577-624.

[48] L.v. Graff, Klassen und Ordnungen des Thier-Reichs, wissenschaftlich dargestellt in Wort und Bild: Turbellaria - Acoela und Rhabdocoelida, Leipzig und Heidelberg, C. F. Winter1904-1908.

[49] S. Tyler, T. Artois, S. Schilling, M. Hooge, L.F. Bush, World List of turbellarian worms: Acoelomorpha, Catenulida, Rhabditophora. *Syndesmis* Silliman, 1881, 2006-2019. World Register of Marine Species at http://marinespecies.org/aphia.php?p=taxdetails&id=142304. (Accessed 2019-07-22).

[50] WoRMS Editorial Board, World Register of Marine Species, 2019. http://www.marinespecies.org at VLIZ. (Accessed 2019-02-07).

[51] H.L. Jameson, Additional notes on the Turbellaria of the L.M.B.C. district, Proc. Liverpool Biol. Soc. 11 (1897) 160-178.

[52] J.B. Jennings, Physiological adaptations to entosymbiosis in three species of graffillid rhabdocoels, Hydrobiologia 84 (1981) 147-153.

[53] M.T. Rosa, D.S. Oliveira, E.L.S. Loreto, Characterization of the first mitochondrial genome of a catenulid flatworm: *Stenostomum leucops* (Platyhelminthes), J Zool Syst Evol Res 55(2) (2017) 98-105.

[54] P. François, Sur le *Syndesmis*, nouveau type de Turbellariés décrit par M. W.-A. Sillimann, C. r. hebd. séances Acad. sci. 103 (1886) 752-757.

[55] C. Linnaeus, Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, Laurentius Salvius, Stockholm, Sweden, 1758.

[56] M. Monnens, E.J. Frost, M. Clark, M.A. Sewell, M.P.M. Vanhove, T. Artois, Description and ecophysiology of a new species of *Syndesmis* Silliman, 1881 (Rhabdocoela: Umagillidae) from the sea urchin *Evechinus chloroticus* (Valenciennes, 1846) Mortensen, 1943 in New Zealand, Int J Parasitol Parasites Wildl 10 (2019) 71-82.

[57] A. Valenciennes, Zoophytes, in: A.-A. Du Petit Thouars (Ed.), Voyage autour de monde sur la frégate La Vénus, pendant les années 1836-1839, Atlas de Zoologie Gide et Cie, Paris, France, 1846.

[58] T. Mortensen, A monograph of the Echinoidea. III. 3. Camarodonta. II. Echinidæ, Strongylocentrotidæ, Parasaleniidæ, Echinometridæ, C.A. Reitzel, Copenhagen, Denmark, 1943.

[59] J. Schultz, F. Milpetz, P. Bork, C.P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains, Proc Natl Acad Sci U S A 95(11) (1998) 5857-5864.

[60] H.R. Bridges, J.A. Birrell, J. Hirst, The mitochondrial-encoded subunits of respiratory complex I (NADH:ubiquinone oxidoreductase): identifying residues important in mechanism and disease, Biochem Soc Trans 39 (2011) 799-806.

[61] D.R. Wolstenholme, Animal mitochondrial DNA: structure and evolution, Int Rev Cytol 141 (1992) 173-216.

[62] D. Ojala, J. Montoya, G. Attardi, tRNA punctuation model of RNA processing in human mitochondria, Nature 290 (1981) 470-474.

[63] M. Nakao, Y. Sako, A. Ito, The mitochondrial genome of the tapeworm *Taenia solium*: a finding of the abbreviated stop codon U, J Parasitol 89(3) (2003) 633-635.

[64] I. Ruiz-Trillo, M. Riutort, H.M. Fourcade, J. Baguña, J.L. Boore, Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyly of the Platyhelminthes, Mol Phylogenet Evol 33(321-332) (2004).

[65] D. Zhang, H. Zou, S.G. Wu, M. Li, I. Jakovlíc, J. Zhang, R. Chen, G.T. Wang, W.X. Li, Sequencing, characterization and phylogenomics of the complete mitochondrial genome of *Dactylogyrus lamellatus* (Monogenea: Dactylogyridae), J Helminthol 92(4) (2017) 455-466.

[66] M.P.M. Vanhove, A.G. Briscoe, M.W.P. Jorissen, D.T.J. Littlewood, T. Huyse, The first next-generation sequencing approach to the mitochondrial phylogeny of African monogenean parasites (Platyhelminthes: Gyrodactylidae and Dactylogyridae), BMC Genomics 19(520) (2018).

[67] S. Castellana, S. Vicario, S. Saccone, Evolutionary patterns of the mitochodnrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes, Genome Biol Evol 3 (2011) 1067-1079.

[68] X.J. Min, D.A. Hickey, DNA asymmetric strand bias affects the amino acid composition of mitochondrial proteins, DNA Res 14 (2007) 201-206.

[69] M. Dowton, A.D. Austin, Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera, J Mol Evol 41 (1995) 958-965.

[70] R.D.M. Page, P.L.M. Lee, S.A. Becher, R. Griffiths, D.H. Clayton, A different tempo of mitochondrial DNA evolution in birds and their parasitic lice, Mol Phylogenet Evol 9(2) (1998) 276-293.

[71] L.R. Castro, A.D. Austin, M. Dowton, Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera, Mol Biol Evol 19(7) (2002) 1100-1113.

[72] K.P. Johnson, J.M. Allen, B.P. Olds, L. Mugisha, D.L. Reed, K.N. Paige, B.R. Pittendrigh, Rates of genomic divergence in humans, chimpanzees and their lice, Proc R Soc Lond B Biol Sci 281(1777) (2014) 2013-2174.

[73] S. Koblmüller, C. Sturmbauer, E. Verheyen, A. Meyer, W. Salzburger, Mitochondrial phylogeny and phylogeography of East African squeaker catfishes (Siluriformes: *Synodontis*), BMC Evol Biol 6(49) (2006) 1471-2148.

[74] C.E. Laumer, A. Hejnol, G. Giribet, Nuclear genomic signals of the 'microturbellarian' roots of platyhelminth evolutionary innovation, eLIFE 4 (2015) e05503.

[75] B. Egger, F. Lapraz, B. Tomiczek, S. Müller, C. Dessimoz, J. Girstmair, N. Škunca, 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 (2015) 1347-1353.

[76] T.H. Le, D. Blair, D.P. McManus, Mitochondrial genomes of parasitic flatworms, Trends Parasitol 18(5) (2002) 206-213.

[77] E.O. Schmidt, Neue Beiträge zur Naturgeschichte der Würmer gesammelt auf einer Reise nach den Färör im Frühjahr 1848, Friedrich Mauke, Jena, Germany, 1848.

[78] A.N. Stephens, M.A. Khan, X. Roucou, P. Nagley, R.J. Devenish, The molecular neighborhood of subunit 8 of yeast mitochondrial F_1F_0-ATP synthase probed by cysteine scanning mutagenesis and chemical modification, J Biol Chem 278(20) (2003) 17867-17875.

[79] J. Lee, S. Ding, T.B. Walpole, A.N. Holding, M.G. Montgomery, I.M. Fearnley, J.E. Walker, Organisation of subunits in the membrane domain of the bovine F-ATPase revealed by covalent cross-linking, J Biol Chem 290 (2015) 13308-13320.

[80] F. Ye, S.D. King, D.K. Cone, P. You, The mitochondrial genome of *Paragyrodactylus variegatus* (Platyhelminthes: Monogenea): differences in major non-coding region and gene order compared to *Gyrodactylus*, Parasit Vectors 7(377) (2014).

[81] L. Bachmann, B. Fromm, L. Patella de Azambuja, W.A. Boeger, The mitochondrial genome of the egg-laying flatworm *Aglaiogyrodactylus forficulatus* (Platyhelminthes: Monogenoidea), Parasit Vectors 9(1) (2016).

[82] A.R. Wey-Fabrizius, L. Podsiadlowski, H. Herlyn, T. Hankeln, Platyzoan mitochondrial genomes, Mol Phylogenet Evol 69(2) (2013) 365-375.

[83] M.T. Aguado, C. Grande, M. Gerth, C. Bleindorn, C. Noreña, Characterization of the complete mitochondrial genomes from Polycladida (Platyhelminthes) using next-generation sequencing, Gene 575 (2016) 199-205.

[84] M. Bernt, A. Braband, B. Schierwater, P.F. Stadler, Genetic aspects of mitochondrial genome evolution, Mol Phylogenet Evol 69 (2013) 328-338.

[85] E.D.B.-R. Marcus, E. Marcus, Polycladida from Curaçao and faunistically related regions, Studies on the fauna of Curaçao and other Caribbean Islands 101 (1968) 1-134.

[86] D.F. Weinland, Human cestoïdes. An essay on the tapeworms of man, Metcalfe and company, Cambridge, United States, 1858.

[87] T.H. Le, D. Blair, T. Agatsuma, P.-F. Humair, N.J.H. Campbell, M. Iwagami, D.T.J. Littlewood, B. Peacock, D.A. Johnston, J. Bartley, D. Rollinson, E.A. Herniou, D.S. Zarlenga, D.P. McManus, Phylogenies inferred from mitochondrial gene orders - A cautionary tale from the parasitic flatworms, Mol Biol Evol 17(7) (2000) 1123-1125.

[88] R. Shao, N.J.H. Campbell, E.R. Schmidt, S.C. Barker, Increased rate of gene rearrangment in the mitochondrial genomes of three orders of hemipteroid insects, Mol Biol Evol 18(9) (2001) 1828-1832.

[89] J.-H. Xiao, J.-G. Jia, R.W. Murphy, D.-W. Huang, Rapid evolution of the mitochondrial genome in chalcidoid wasps (Hymenoptera: Chalcidoidea) driven by parasitic lifestyles, PLoS One 6(11) (2011) e26645.

[90] M. Zarowiecki, M. Berriman, What helminth genomes have taught us about parasite evolution, Parasitology 142 (2014) S85-S97.

[91] E. Skippington, T.J. Barkman, D.W. Rice, J.D. Palmer, Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes, PNAS 112(7) (2015) E3515-3524.

[92] S. Andrews, FASTQC: a quality control tool for high throughput sequence data, 2010.

[93] N. Dierckxsens, P. Mardulyn, G. Smits, NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data, Nucleic Acids Res 45(4) (2017) e18.

[94] G.L. Shinn, *Anoplodium hymanae* sp. n., an umagillid turbellarian from the coelom of *Stichopus californicus*, a Northeast Pacific holothurian, Can. J Zool. 61(4) (1983) 750-760.

[95] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Mentjies, A. Drummond, Geneious Basic: an integrated and extendable dekstop software platform for the organization and analysis of sequence data, Bioinformatics 28(12) (2012) 1647-1649.

[96] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J Mol Biol 215 (1990) 403-410.

[97] N.T. Perna, T.D. Kocher, Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes, J Mol Evol 41 (1995) 353-358.

[98] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res 32(5) (2004) 1792-1797.

[99] P. Stothard, The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences, Biotechniques 28 (2000) 1102-1104.

[100] P.M. Sharp, T.M.F. Tuohy, K.R. Mosurski, Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes, Nucleic Acids Res 14(13) (1986) 5125-5143.

[101] B.D. Lee, Python implementation of codon adaptation index, J Open Source Softw 3(30) (2018) 905.

[102] D. Laslett, B. Canbäck, ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences, Bioinformatics 24 (2008) 172-175.

[103] T.M. Lowe, P.p. Chen, tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes, Nucleic Acids Res 44(Web Server issue) (2016) W54-W57.

[104] A.R. Gruber, R. Lorenz, S.H. Bernhart, R. Neuböck, I.L. Hofacker, The Vienna RNA websuite, Nucleic Acids Res 46 (2008) W70-W74.

[105] R. Lorenz, S.H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algorithms for Molecular Biology 6 (2011) 26.

[106] L. Noe, G. Kucherov, YASS: enhancing the sensitivity of DNA similarity search, Nucleic Acids Res 33(2) (2005) W540-W543.

[107] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res 27(2) (1999) 573-580.

[108] M. Bernt, D. Merkle, K. Ramch, G. Fritzch, M. Perseke, D. Bernhard, M. Schlegel, P. Stadler, M. Middendorf, CREx: inferring genomic rearrangements based on common intervals, Bioinformatics 23(21) (2007) 2957-2958.

[109] M. Lohse, O. Drechsel, R. Bock, OrganellarGenomeDRAW (OGDRAW) - a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes, Curr Genet 52 (2007) 267-274.

[110] M. Lohse, O. Drechsel, S. Kahlau, R. Bock, OrganellarGenomeDRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets, Nucleic Acids Res 41 (2013) W575-W581.

[111] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Res 31 (2003) 3784-3788.

[112] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, Protein identification and analysis tools on the ExPASy server, in: J.M. Walker (Ed.), The proteomics protocols handbook, Humana Press 2005, pp. 571-607.

[113] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J Mol Biol 157 (1982) 105-132.

[114] A. Bankevich, S. Nurk, D. Antipov, A. Gurevich, M. Dvorkin, A.S. Kulikov, V. Lesin, S. Nikolenko, S. Pham, A. Prjibelski, A. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J Comput Biol 19(5) (2012) 455-477.

[115] R.R. Wick, M.B. Schultz, J. Zobel, K.E. Holt, Bandage: interactive visualization of *de novo* genome assemblies., Bioinformatics 31(20) (2015) 3350-3352.

[116] K. Lagesen, P. Hallin, E.A. Rødland, H.H. Staerfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, Nucleic Acids Res 35(9) (2007) 3100-3108.

[117] J. Bengtsson-Palme, V. Veldre, M. Ryberg, M. Hartmann, S. Branco, Z. Wang, A. Godhe, Y. Bertrand, P. De Wit, M. Sanchez, I. Ebersberger, K. Sanli, F. de Souza, E. Kristiansson, K. Abarenkov, K.M. Eriksson, R.H. Nilsson, ITSx: Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for use in environmental sequencing, Methods Ecol Evol 4 (2013) 914-919.

[118] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, Brief Bioinform (2017) 1-7.

[119] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol Biol Evol 17 (2000) 540-552.

[120] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst Biol 56(4) (2007) 564-577.

[121] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat Methods 14 (2017) 587-589.

[122] O. Chernomor, A. von Haeseler, B.Q. Minh, Terrace aware data structure for phylogenomic inference from supermatrices, Syst Biol 65 (2016) 997-1008.

[123] F. Ronquist, M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, MrBayes 3.2 Efficient Bayesian phylogenetic inference and model choice across a large model space, Syst Biol 61(3) (2012) 539-542.

[124] M.A. Miller, W. Pfeiffer, T. Schwartz, Creating the CIPRES Science Gateway for inference of large phylogenetic trees, Proceedings of the Gateway Computing Environments Workshop (GCE) (2010) 1-8.

[125] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol Biol Evol 32(1) (2015) 268-274.

[126] J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, B.Q. Minh, W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis, Nucleic Acids Res 44(W1) (2016) W232-235.

[127] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: Improving the ultrafast bootstrap approximation, Mol Biol Evol 35(2) (2017) 518-522.

[128] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Fascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0 Syst Biol 59(3) (2010) 307-321.

[129] B.C. Stöver, K.F. Müller, TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses, BMC Bioinformatics 11(7) (2010).

[130] J.O. Wertheim, B. Murrell, M.D. Smith, S.L. Kosakovsky Pond, K. Scheffler, RELAX: detecting relaxed selection in a phylogenetic framework, Mol Biol Evol 32(3) (2015) 820-832.

[131] W.P. Maddison, D.R. Maddison, Mesquite: a modular system for evolutionary analysis, 2018.

[132] S. Weaver, S.D. Shank, S.J. Spielman, M. Li, S.V. Muse, S.L. Kosakovsky Pond, Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes, Mol Biol Evol 35(3) (2018) 773-777.

Appendices

Table A1. List of flatworm species and corresponding GenBank accession numbers used to align protein-coding genes for annotation purposes (excluding outgroup taxa), as well as in the mitochondrial phylogenetic analysis.

Species	Accession number
NEODERMATA	
Digenea	
Alaria americana	MH536507
Artyfechinostomum sufrartyfex	NC_037150
Atractolytocestus huronensis	NC_035635
Brachycladium goliath	NC_029757
Calicophoron microbothrioides	NC_027271
Cardiocephaloides medioconiger	MH536508
Clinostomum complanatum	NC_027082
Clonorchis sinensis	NC_012147
Cotylurus marcogliesei	MH536509
Cyathocotyle prussica	NC_039780
Dactylogyrus lamellatus	NC_035610
Dicrocoelium chinensis	NC_025279
Dicrocoelium dendriticum	NC_025280
Echinochasmus japonicus	NC_030518
Echinostoma hortense	NC_028010
Echinostoma paraensei	KT008005
Eurytrema pancreaticum	KP241855
Explanatum explanatum	NC_027958
Fasciola gigantica	NC_024025
Fasciola hepatica	NC_002546
Fasciola jacksoni	KX787886
Fasciola sp.	KF543343
Fascioloides magna	NC_029481
Fasciolopsis buski	NC_030528
Fischoederius cobboldi	NC_030529
Fischoederius elongatus	NC_028001
Gastrothylax crumenifer	NC_027833
Haplorchis taichui	NC_022433
Homalogaster paloniae	NC_030530
Hysteromorpha triloba	MH536511
Metagonimus yokogawai	NC_023249
Metorchis orientalis	NC_028008
Ogmocotyle sikae	NC_027112

Ogmocotyle sp.	KR006935
Opisthorchis felineus	NC_011127
Opisthorchis viverrini	JF739555
Orthocoelium streptocoelium	NC_028071
Paragonimus ohirai	NC_032032
Paragonimus westermani	NC_002354
Paramphistomum cervi	KT198987
Posthodiplostomum centrarchi	MH536512
Schistosoma haematobium	NC_008074
Schistosoma japonicum	HM120848
Schistosoma mansoni	NC_002545
Schistosoma mekongi	NC_002529
Schistosoma spindale	NC_008067
Trichobilharzia regenti	NC_009680
Trichobilharzia szidati	NC_036411
Tylodelphis immer	MH536513
Cestoda	
Anoplocephala magna	NC_031801
Anoplocephala perfoliata	NC_028425
Atractolytocestus huronensis	NC_035635
Breviscolex orientalis	NC_035634
Caryophyllaeus brachycollis	NC_035430
Cladotaenia vulturi	NC_032067
Cloacotaenia megalops	NC_032295
Digramma interrupta	NC_039446
Diphyllobothrium latum	NC_008945
Diphyllobothrium nihonkaiense	NC_009463
Diplogonoporus balaenopterae	NC_017613
Diplogonoporus grandis	NC_017615
Drepanidotaenia lanceolata	NC_028164
Echinococcus canadensis	NC_011121
Echinococcus equinus	NC_020374
Echinococcus granulosus	NC_008075
Echinococcus multilocularis	NC_000928
Echinococcus oligarthrus	NC_009461
Echinococcus ortleppi	NC_011122
Echinococcus shiquicus	NC_009460
Echinococcus vogeli	NC_009462
Homalogaster paloniae	NC_030530
Hydatigera kamiyai	NC_037071
Hydatigera krepkogorski	NC_021142
Hydatigera parva	NC_021141
Hymenolepis diminuta	NC_002767
Hymenolepis nana	NC_029245

Khawia sinensis	NC_034800
Ligula intestinalis	NC_039445
Moniezia benedeni	NC_036218
Moniezia expansa	NC_036219
Paragonimus heterotremus	NC_039430
Paruterina candelabraria	MH282837
Pseudanoplocephala crawfordi	NC_028334
Raillietina tetragona	KP057580
Schyzocotyle acheilognathi	NC_030316
Schyzocotyle nayarensis	NC_030317
Senga ophiocephalina	NC_034715
Spirometra decipiens	NC_026852
Spirometra erinaceieuropaei	NC_011037
Taenia arctos	NC_024590
Taenia asiatica	NC_004826
Taenia crassiceps	NC_002547
Taenia crocutae	NC_024591
Taenia hydatigena	NC_012896
Taenia multiceps	NC_012894
Taenia pisiformis	NC_013844
Taenia regis	NC_024589
Taenia saginata	NC_009938
Taenia solium	NC_004022
Taenia taeniaeformis	NC_014768
Testudotaenia sp.	KU761587
Versteria mustelae	NC_021143
Monogenea	
Aglaiogyrodactylus forficulatus	NC_030339
Benedenia hoshinai	NC_014591
Benedenia seriolae	NC_014291
Cichlidogyrus halli	MG970255
Cichlidogyrus mbirizei	MG970257
Dactylogyrus lamellatus	NC_035610
Eudiplozoon sp.	MG458328
Gyrodactylus brachymystacis	NC_031337
Gyrodactylus derjavinoides	NC_010976
Gyrodactylus gurleyi	KU659806
Gyrodactylus kobayashii	NC_030050
Gyrodactylus nyanzae	NC_038214
Gyrodactylus parvae	NC_031438
Gyrodactylus salaris	NC_008815
Lamellodiscus spari	MH328204
Lepidotrema longipenis	MH328203
Macrogyrodactylus karibae	MG970258

Microcotyle sebastis	NC_009055
Neobenedenia melleni	JQ038228
Paradiplozoon opsariichthydis	MG458327
Paragyrodactylus variegatus	NC_024754
Paratetraonchoides inermis	NC_036305
Polylabris halichoeres	NC_016057
Pseudochauhanea macrorchis	NC_016950
Sindiplozoon sp.	MG458326
Tetrancistrum nebulosi	NC_018031
'TURBELLARIA'	
Catenulida	
Stenostomum leucops	KX553929
Stenostomum sthenum	NC_035256
Macrostomorpha	
Macrostomum lignano	NC_035255
Polycladida	
Crassiplana albatrossi	MF993330
Cryptocelis alba	MF993331
Discocelis tigrina	MF993332
Enchiridium sp.	NC_028199
Eurylepta cornuta	MF993334
Hoploplana elisabelloi	NC_028200
Imogine fafai	MF993335
Imogine stellae	MF993336
Notocomplana palta	MF993337
Planocera reticulata	NC_036051
Prosthiostomum siphunculus	NC_028201
Stylochoplana maculata	KP965863
Tricladida	
Crenobia alpina	KP208776
Dugesia japonica	NC_016439
Dugesia ryukyuensis	AB618488
Girardia sp.	KP090061
Obama sp.	KP208777
Phagocata gracilis	KP090060
Schmidtea mediterranea	NC_022448
Rhabdocoela	
Bothromesostoma personatum	MF993329
Graffilla buccinicola	This study
Syndesmis kurakaikina	This study
Syndesmis echinorum	This study
Gnathostomulida (outgroup)	
Gnathostomula armata	NC_026983
Gnathostomula paradoxa	NC_026984

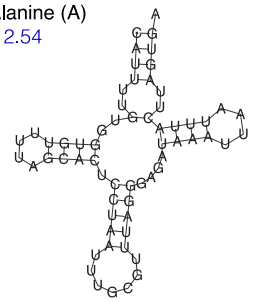
Table A2. Overview of sequences included in *atp8* alignment and selection test.

Species	**Accession number**
Crenobia alpina	KP208776
Dugesia japonica	NC_016439
Dugesia ryukyuensis	AB618488
Enchiridium sp.	NC_028199
Girardia sp.	KP090061
Hoploplana elisabelloi	NC_028200
Macrostomum lignano	NC_035255
Phagocata gracilis	KP090060
Prosthiostomum siphunculus	NC_028201
Schmidtea mediterranea	NC_022448
Stylochoplana maculata	KP965863
Syndesmis kurakaikina	This study
Syndesmis echinorum	This study

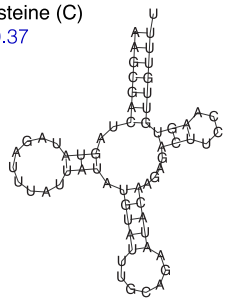
Fig. A1 (next pages). Predicted secondary structures of mt tRNA in a. *Syndesmis kurakaikina*, b. *Syndesmis echinorum*, c. *Graffilla buccinicola*, as visualised in MiTFi. Values in blue represent free energy levels of secondary structures as calculated in RNAeval (expressed in kcal/mol). If alternative predictions (tRNAscan-SE/ARWEN) were adopted, vector images were edited accordingly in Adobe Illustrator CS5.1.

a. *Syndesmis kurakaikina*

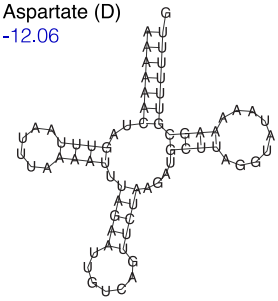
Alanine (A)
-12.54



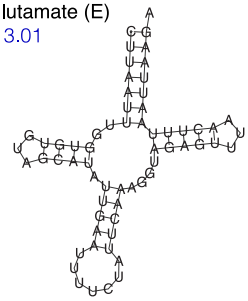
Cysteine (C)
-10.37



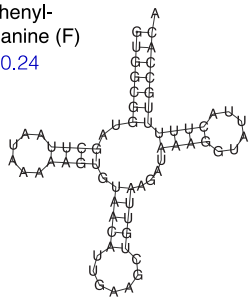
Aspartate (D)
-12.06



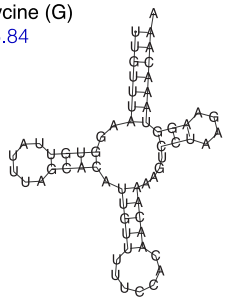
Glutamate (E)
-13.01



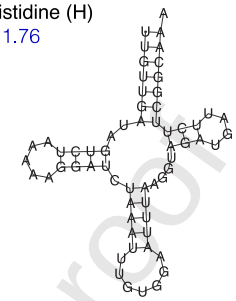
Phenyl-
alanine (F)
-20.24



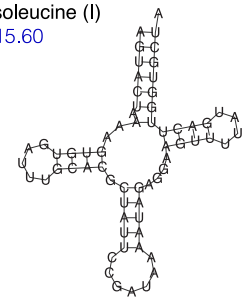
Glycine (G)
-15.84



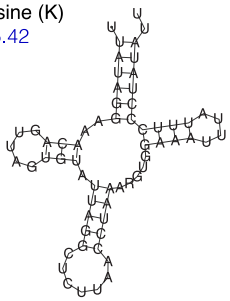
Histidine (H)
-11.76



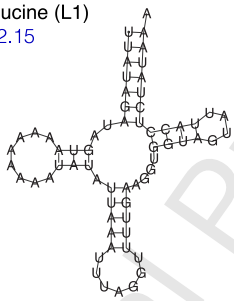
Isoleucine (I)
-15.60



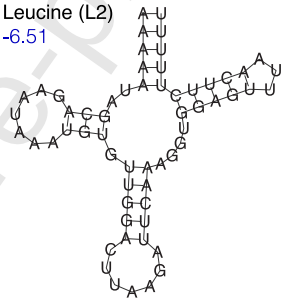
Lysine (K)
-15.42



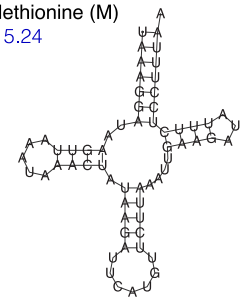
Leucine (L1)
-12.15



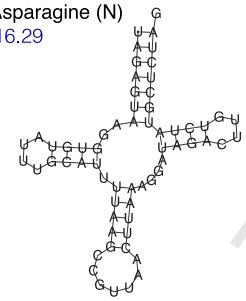
Leucine (L2)
-6.51



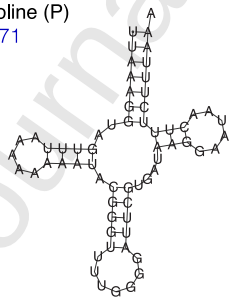
Methionine (M)
-15.24



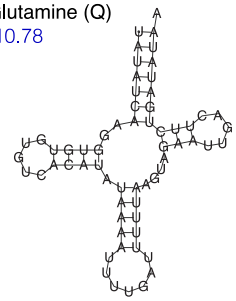
Asparagine (N)
-16.29



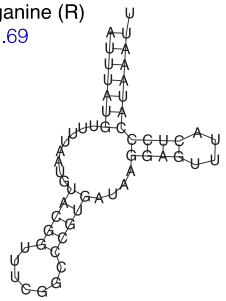
Proline (P)
-9.71



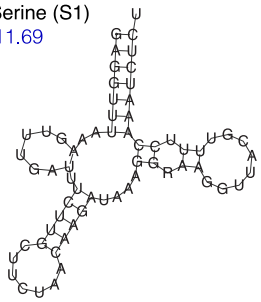
Glutamine (Q)
-10.78



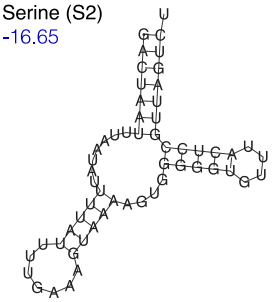
Arganine (R)
-16.69



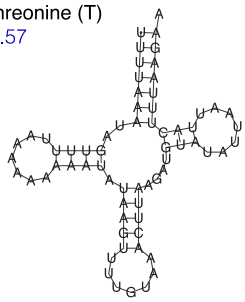
Serine (S1)
-11.69



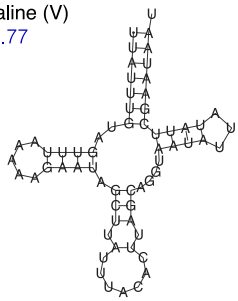
Serine (S2)
-16.65



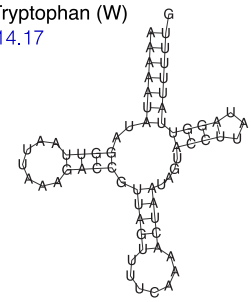
Threonine (T)
-6.57



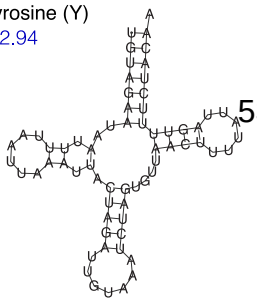
Valine (V)
-4.77



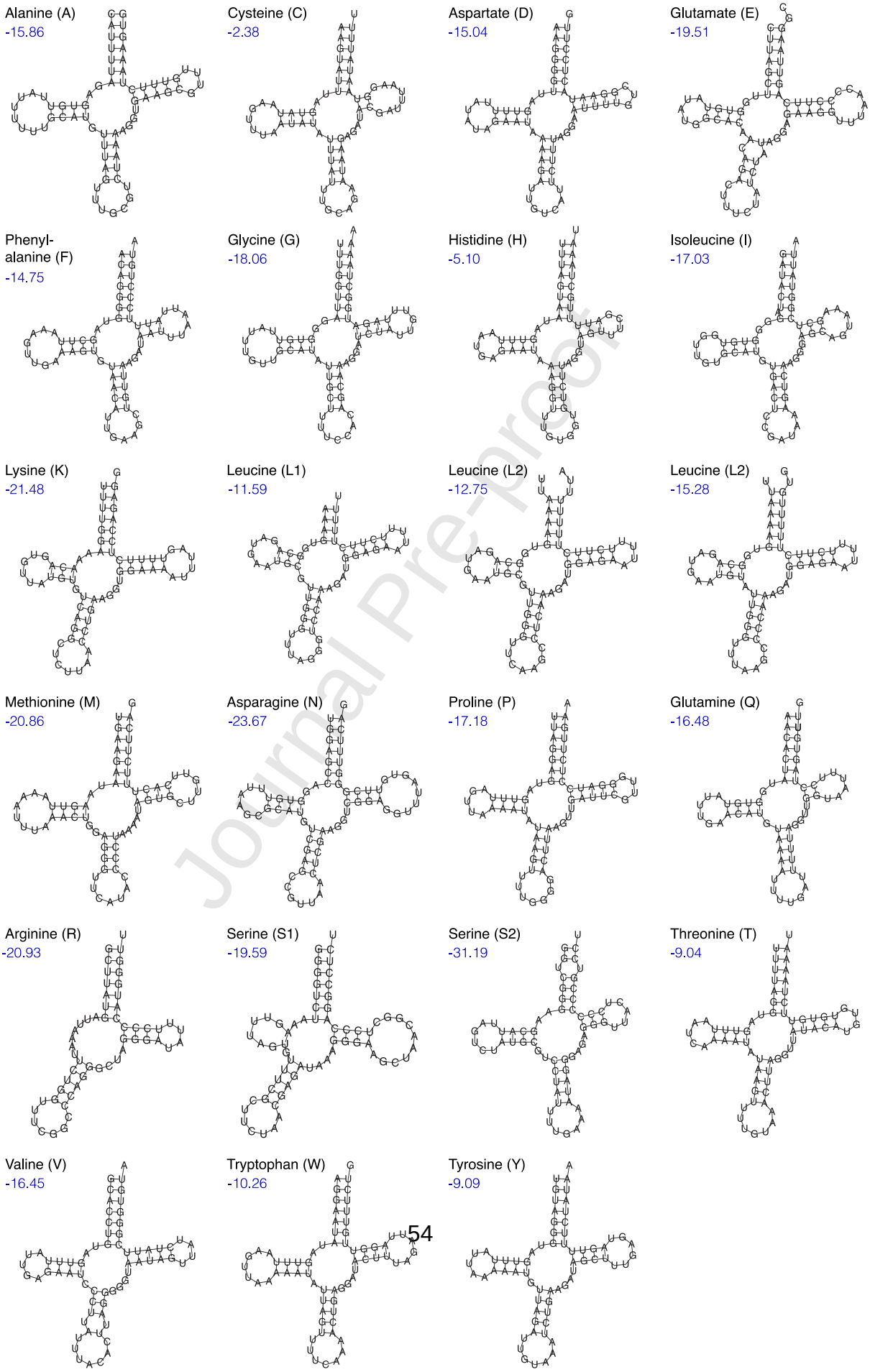
Tryptophan (W)
-14.17



Tyrosine (Y)
-12.94



b. *Syndesmis echinorum*



c. *Graffilla buccinicola*

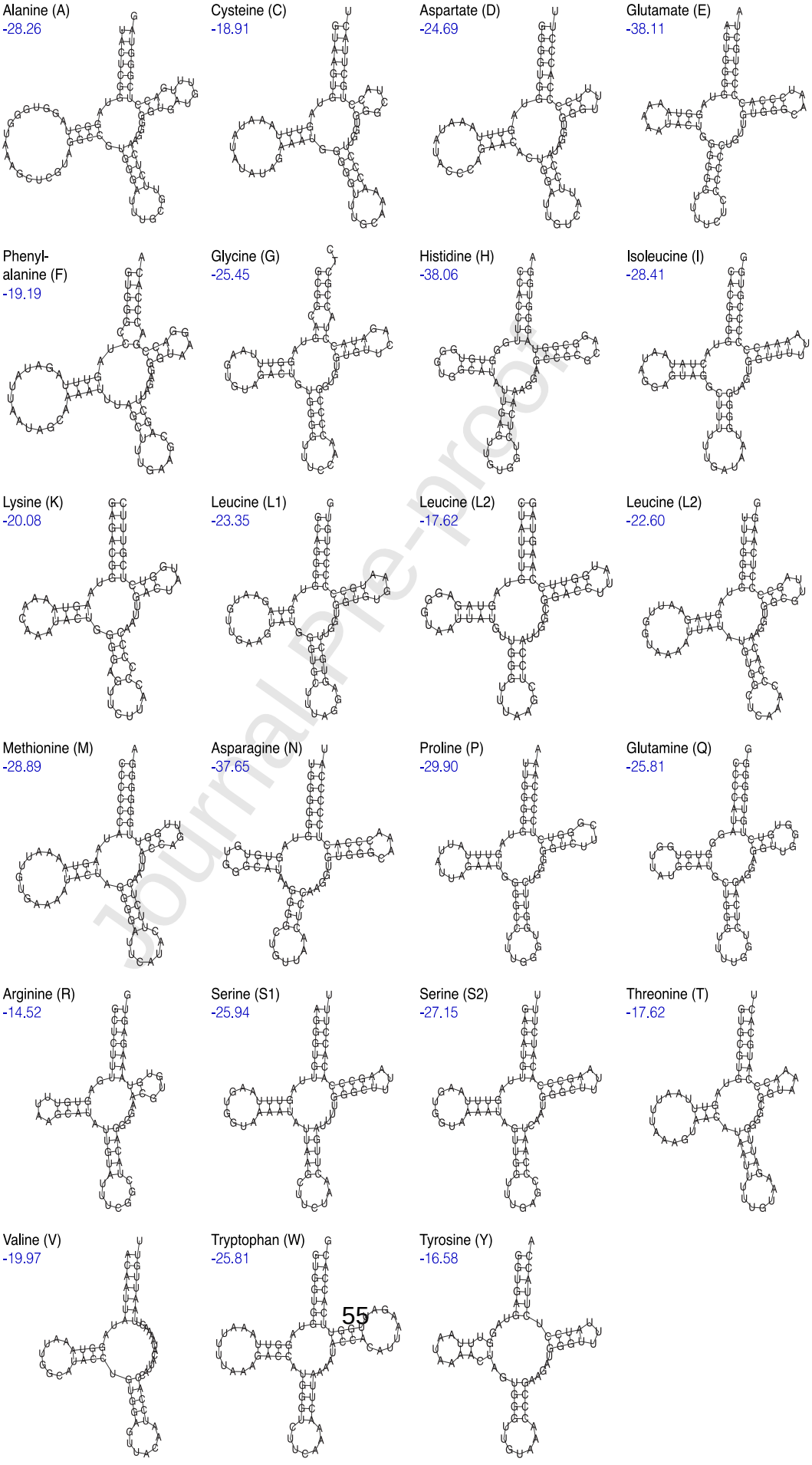


Table 1. GC content and compositional asymmetry in the three newly-assembled sequences of *Syndesmis kurakaikina*, *Syndesmis echinorum* and *Graffilla buccinicola*. Skew values were calculated manually following the formulas [(G-C)/(G+C)] and [(A-T)/(A+T)].

	Syndesmis kurakaikina	***Syndesmis echinorum***	***Graffilla buccinicola***
GC%	27.3	38.0	44.7
AT skew	-0.161	-0.297	-0.357
GC skew	0.384	0.402	0.480

Table 2. Codon frequency in *Syndesmis kurakaikina*, *Syndesmis echinorum* and *Graffilla buccinicola*, expressed as a fraction of occurrences in each synonymous codon family (Codon Usage tool in the Sequence Manipulation suite) and as relative codon usage (RSCU, python implementation of Lee, et al. [110]). GTG and ATG can operate as start codons and are marked with (i). Most-frequently-used codons per species are indicated in bold for each amino acid.

		S. kurakaikina		***S. echinorum***		***G. buccinicola***	
Amino acid	Triplet	Fraction	RSCU	Fraction	RSCU	Fraction	RSCU
Ala	GCG	0.08	0.32	0.19	0.74	0.18	0.70
	GCA	0.28	1.12	0.17	0.69	0.19	0.77
	GCT	**0.62**	**2.49**	**0.53**	**2.14**	**0.43**	**1.73**
	GCC	0.02	0.07	0.11	0.43	0.20	0.80
Cys	**TGT**	**0.98**	**1.96**	**0.85**	**1.69**	**0.93**	**1.86**
	TGC	0.02	0.04	0.15	0.31	0.07	0.13
Asp	**GAT**	**0.82**	**1.63**	**0.80**	**1.61**	**0.86**	**1.71**
	GAC	0.18	0.37	0.23	0.39	0.14	0.29
Glu	GAG	0.20	**0.40**	**0.91**	**1.54**	**0.83**	**1.66**
	GAA	**0.80**	1.60	0.09	0.46	0.17	0.34
Phe	**TTT**	**0.95**	**1.91**	**0.91**	**1.81**	**0.89**	**1.79**
	TTC	0.05	0.09	0.09	0.19	0.11	0.21
Gly	**GGG**	0.19	0.76	**0.56**	**2.25**	**0.45**	**1.81**
	GGA	0.37	1.47	0.08	0.33	0.08	0.33
	GGT	**0.42**	**1.70**	0.29	1.17	0.37	1.49
	GGC	0.02	0.07	0.07	0.26	0.09	0.37
His	**CAT**	**0.96**	**1.92**	**0.79**	**1.57**	**0.61**	**1.22**

	CAC	0.04	0.08	0.21	0.43	0.39	0.78
Ile	ATA	0.38	1.13	0.23	0.73	0.20	0.60
	ATT	**0.61**	**1.84**	**0.68**	**2.04**	**0.76**	**2.29**
	ATC	0.01	0.04	0.08	0.23	0.04	0.11
Lys	**AAG**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Leu	**TTG**	0.10	0.59	**0.39**	**2.33**	**0.48**	**2.90**
	TTA	**0.76**	**4.55**	0.36	2.21	0.34	2.03
	CTG	0.00	0.02	0.04	0.23	0.03	0.18
	CTA	0.04	0.21	0.04	0.24	0.04	0.21
	CTT	0.10	0.58	0.13	0.81	0.10	0.60
	CTC	0.01	0.04	0.03	0.17	0.01	0.08
Met	**ATG (i)**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**
Asn	AAA	0.55	1.66	0.39	1.15	0.13	0.38
	AAT	**0.42**	**1.25**	**0.48**	**1.46**	**0.65**	**1.95**
	AAC	0.03	0.09	0.13	0.39	0.22	0.67
Pro	CCG	0.11	0.44	0.18	0.71	0.16	0.63
	CCA	0.20	0.80	0.17	0.67	0.29	1.18
	CCT	**0.67**	**2.67**	**0.45**	**1.79**	**0.34**	**1.35**
	CCC	0.02	0.09	0.20	0.82	0.21	0.84
Gln	**CAG**	0.05	0.11	**0.50**	**1.02**	**0.52**	**1.04**
	CAA	**0.95**	**1.89**	0.50	0.98	0.48	0.96
Arg	**CGG**	0.12	0.46	**0.46**	**1.85**	0.41	1.65
	CGA	**0.52**	**2.06**	0.17	0.69	0.18	0.71
	CGT	0.37	1.45	0.29	1.15	**0.29**	**1.18**
	CGC	0.00	0.04	0.08	0.31	0.12	0.47
Ser	**AGG**	**0.14**	**1.10**	**0.26**	**2.11**	0.27	2.20
	AGA	0.30	2.40	0.09	0.66	0.10	0.82
	AGT	0.18	1.44	0.17	1.40	**0.23**	**1.87**
	AGC	0.03	0.24	0.07	0.60	0.04	0.33
	TCG	0.03	0.24	0.05	0.39	0.05	0.39
	TCA	0.13	1.06	0.07	0.57	0.06	0.49
	TCT	0.18	1.47	0.23	1.81	0.21	1.70
	TCC	0.01	0.05	0.06	0.46	0.02	0.20
Thr	ACG	0.04	0.16	0.12	0.49	0.14	0.58
	ACA	0.43	1.70	0.28	1.15	0.17	0.67
	ACT	**0.52**	**2.10**	**0.48**	**1.90**	**0.51**	**2.02**
	ACC	0.01	0.04	0.11	0.45	0.18	0.72
Val	GTG (i)	0.11	0.43	0.32	1.28	**0.42**	**1.66**
	GTA	0.36	1.42	0.10	0.39	0.16	0.65
	GTT	**0.51**	**2.05**	**0.51**	**2.05**	**0.40**	**1.58**
	GTC	0.02	0.09	0.07	0.28	0.03	0.10
Trp	**TGG**	**0.34**	**0.68**	**0.81**	**1.60**	**0.83**	**1.65**
	TGA	0.66	1.32	0.19	0.40	0.17	0.35

Tyr	**TAT**	**0.94**	**1.88**	**0.77**	**1.54**	**0.75**	**1.49**
	TAC	0.06	0.11	0.23	0.46	0.25	0.51
Stop	**TAG**	0.13	0.25	**0.63**	**1.25**	**0.90**	**1.80**
	TAA	**0.88**	**1.75**	0.38	0.75	0.10	0.20

Table 3. Best-fitting evolutionary models for each partition according to the Bayesian Information Criterion as inferred by IQ-TREE.

Partition	Evolutionary model
rrnS, rrnL	GTR+G4+I
First codon positions of *cox1, cox2* Second codon positions of *cox1, cox2, cox3*	GTR+G4+I
Second codon positions of *cytb, nad1, nad2, nad3, nad4, nad5, nad6*	GTR+G4+I
Third codon positions of *cox1, cox2, cox3, cytb, nad1, nad2, nad3, nad4, nad4L, nad5, nad6*	GTR+G4+ASC
First codon positions of *atp6*, *cox3*, *cytb*, *nad1, nad2, nad3, nad4L, nad5, nad6* Second codon positions of *atp6, nad4L* Third codon positions of *atp6*	GTR+G4+I

Table 4. Estimated selection intensity parameters (K), likelihood ratios (LR) and corresponding p-values derived from RELAX tests (*p < 0.05). ΔAICc corresponds with the difference between RELAX null and alternative models. P-values of datasets for which a convergence warning was generated are greyed out.

PCG	K	LR	p-value	ΔAICc ($AICc_{null} - AICc_{alternative}$)
atp6	0.84	-0.07	1.000	-2.2
atp8	0.98	19.80	0.000*	17.4
cox1	0.61	14.49	0.000*	12.4
cox2	1.01	0.00	0.944	-2.1
cox3	1.27	2.42	0.119	0.4
cytb	1.02	0.14	0.710	-1.9
nad1	0.88	-1.20	1.000	-3.2
nad2	1.10	1.25	0.264	-0.9
nad3	0.81	0.21	0.646	-1.9
nad4	1.00	-0.05	1.000	-2.1
nad4L	**0.79**	**-1.08**	1.000	-3.2
nad5	1.00	-3.08	1.000	-5.1
nad6	0.58	-0.06	1.000	-2.2

Table 5. Overview of the four N-terminal amino acids in previously-annotated *atp8* genes in turbellarian mt genomes. Included sequences were mined from GenBank, and translated according to their respective genetic codes. Translation tables are numbered according to the recommendations found at ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi.

Species	Accession number	Genetic code	Translation
Crenobia alpina	KP208776	9	MIFS
Dugesia japonica	NC_016439	9	MFFF
Dugesia ryukyuensis	AB618488	9	MFVL
Enchiridium sp.	NC_028199	9	MPQM
Girardia sp.	KP090061	9	MCCY
Hoploplana elisabelloi	NC_028200	9	LPHM
Macrostomum lignano	NC_035255	9	IPQL
Phagocata gracilis	KP090060	9	LVDV
Prosthiostomum siphunculus	NC_028201	9	MPQM
Schmidtea mediterranea	NC_022448	9	MVHT
Stenostomum leucops	KX553929	5	MNQF
Stenostomum sthenum	NC_035256	5	MYQM
Stylochoplana maculata	KP965863	9	LPQM

Captions to illustrations

Fig. 1. Physical maps of the three newly-assembled mitochondrial (mt) genomes of endosymbiotic rhabdocoels, as constructed in OGDRAW. GC content is displayed in the centre of each mt genome.

Fig. 2. Schematic representation of gene order changes in rhabdocoel mitochondrial (mt) genomes, based on the three newly-annotated sequences in this study and the annotation for *Bothromesostoma personatum* (accession number: MF993329) as proposed by Kenny et al. [34]. Sequences are displayed linearised and *cox1* was chosen as an arbitrary origin for comparability. Scenarios for pairwise rearrangements were calculated in CREx and are indicated by black (transposition) and white (tandem duplication random loss) squares. Figure was created using Adobe Illustrator CS v5.1.

Fig. 3. Hydrophobicity patterns of the candidate *atp8* gene found in *Syndesmis kurakaikina* and *Syndesmis echinorum* (top left graphs), in comparison with the previously-published (putative) annotations of this gene in other flatworms. Patterns were computed using the ProtScale tool on the ExPASy server (window size = 9), employing the Kyte and Doolittle (1982) hydropathic scoring system for amino acids.

Fig. 4. Majority-rule consensus tree from the Bayesian analysis of the concatenated mt dataset. Topology is congruent with the inferred ML tree. Clades with pp values below 95% have been collapsed. Symbols above branches indicate support values and the corresponding legend is displayed in the box at the right. Neodermata are summarised in a single clade and the root branch has been cut for visibility. Branch lengths denote the

number of expected nucleotide substitutions per site. Branches marked in green were formally tested for shifts in selection pressure.

CRediT author statement

Marlies Monnens: Data Curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Validation, Writing – Original Draft. **Sofie Thijs:** Formal analysis, Writing – Review & Editing. **Andrew Briscoe:** Data Curation, Investigation, Writing – Review & Editing. **Miriam Clark:** Resources. **Emily Joy Frost:** Resources. **Tim Littlewood:** Data Curation, Investigation, Writing – Review & Editing. **Mary Sewell:** Resources, Writing – Review & Editing. **Karen Smeets:** Writing – Review & Editing. **Tom Artois:** Conceptualization, Funding acquisition, Supervision, Writing – Review & Editing. **Maarten P.M. Vanhove:** Conceptualization, Funding acquisition, Supervision, Writing – Review & Editing.