Signatures of Endosymbiosis in Mitochondrial Genomes of Rhabdocoel Flatworms

Peer-reviewed author version

MONNENS, Marlies; ARTOIS, Tom; Briscoe, A.; DIEZ GARCIA, Yander; Fraser, K. P. P.; Leander, B. S.; Littlewood, D. T. J.; Santos, M. J.; SMEETS, Karen; Van Steenkiste, N. W. L. & VANHOVE, Maarten (2025) Signatures of Endosymbiosis in Mitochondrial Genomes of Rhabdocoel Flatworms. In: Molecular Ecology,.

DOI: 10.1111/mec.70015
Handle: http://hdl.handle.net/1942/46455

Signatures of endosymbiosis in mitochondrial genomes of rhabdocoel flatworms

M. Monnens[1,2,*], T. Artois[1], A. Briscoe[3], Y. L. Diez[1,4], K. P. P. Fraser[5], B. S. Leander[6], D. T. J. Littlewood[3], M. J. Santos[7,8], K. Smeets[1], N. W. L. Van Steenkiste[6,9], M. P. M. Vanhove[1]

[1] Hasselt University, Centre for Environmental Sciences, Research Group 'Zoology: Biodiversity and Toxicology', Agoralaan Gebouw D, B-3590 Diepenbeek, Belgium

[2] Royal Belgian Institute of Natural Sciences, OD Taxonomy and Phylogeny, Vautierstraat 29, B-1000 Brussels, Belgium

[3] Natural History Museum, London SW7 5BD, United Kingdom

[4] Museum of Nature Hamburg – Zoology, Leibniz Institute for the Analysis of Biodiversity Change (LIB), Martin-Luther-King-Platz 3, D-20146 Hamburg, Germany

[5] University of Plymouth, Marine Station, Artillery Place, Coxside, Plymouth PL4 OLU, United Kingdom

[6] Departments of Botany and Zoology, University of British Columbia, 3156-6270 University Blvd, Vancouver, BC, V6T 1Z4, Canada

[7] CIIMAR, Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, s/n, Edifício FC4, 4169-007 Porto, Portugal

[8] CIIMAR, Universidade do Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos, s/n, 4450-208 Matosinhos, Portugal

[9] Hakai Institute, 303-1100 Island Hwy, Campbell River, BC, V9W 8C6, Canada

Running title: Genomic signatures of endosymbiosis

[*] Corresponding author. Email: marlies.monnens@uhasselt.be

Abstract

The transition from a free-living lifestyle to endosymbiosis represents a large evolutionary shift, impacting various aspects of any organism's biology, including its molecular-genetic groundwork. So far, it has been impossible to generalise the impact this lifestyle shift has on genomic architecture. This study explores this phenomenon using a new model system: neodalyellid flatworms (Rhabdocoela), a diverse assemblage of free-living and independently evolved endosymbiotic lineages. A uniquely comprehensive mitochondrial genomic dataset, consisting of 50 complete or partial mitogenome sequences (47 of which are new to science), is constructed, increasing the genomic resources available for rhabdocoel flatworms over tenfold. A robust phylogenomic framework is built, enabling an in-depth exploration of the molecular-genetic signatures associated with evolutionary shifts towards endosymbiosis. To understand speciation influenced by host phylogeny, first steps are taken to unravel the host-switching history of the largest endosymbiotic group of neodalyellids. We test several hypotheses regarding the potential consequences of a symbiotic lifestyle, and find marginally heightened AT content, more pronounced GC skew, and relaxed selection on specific protein-coding genes in endosymbionts compared to their free-living counterparts. Numerous substitutions have accumulated in certain endosymbiotic lineages; however, the correlation with lifestyle remains uncertain. A high frequency of genetic rearrangements across all studied lineages is observed. Our findings affirm the variable nature of rhabdocoel mitogenomes and, for the first time, reveal distinct signatures of an endosymbiotic lifestyle in neodalyellid flatworms. This effort lays the groundwork for future research into evolutionary and genomic consequences of a symbiotic lifestyle in this and other animal systems.

Keywords: molecular evolution, comparative genomics, turbellarian flatworms, parasitology, phylogenomics, mitochondrion

Introduction

The shift from a free-living to a symbiotic lifestyle is one of the most radical evolutionary transitions. Throughout metazoan evolution, many lineages have independently made this shift (Poulin 2011; Poulin and Randhawa 2015) towards host(ile) niches markedly distinct from those of their free-living ancestors. This profound transition often manifests through changes in body plan (e.g. Medvedev 2017), behaviour (e.g. Mikheev et al. 2015), reproductive strategy (e.g. Baeza 2015), and/or physiology (Nguyen 2020). Although symbiotic lifestyles encompass a wide variety of ecological strategies, they also represent striking cases of convergent evolution across morphological, functional, and life history traits: many phylogenetically distinct organisms have independently developed similar phenotypic features, following parallel evolutionary trajectories to meet the challenges posed by an obligate symbiotic lifestyle (Poulin, 2011).

On a molecular level, this shift is also associated with substantial effects. Apart from the functional modifications intrinsically tied to a symbiotic lifestyle (Jackson 2014 and references therein; Poulin and Randhawa 2015), a range of genetic ramifications has been identified. For instance, among symbionts as diverse as hymenopterans, bats, fishes, and birds, mitochondrial (mt) genome divergence rate is increased compared to their free-living counterparts (Hafner et al. 1994; Dowton and Austin 1995; Sorenson and Payne 2001; Castro et al. 2002; Koblmüller et al. 2006; Xiao et al. 2011; Botero-Castro et al. 2018). Furthermore, evidence from the mitochondrial genome suggests a connection between parasitism and nucleotide compositional bias, and more specifically, a trend toward higher AT content and an accelerated rate in gene order changes (Dowton and Austin 1995, 1999; Xiao et al. 2011; Li et al. 2020 and references therein).

The number of taxa studied in this respect remains limited, and findings sometimes conflict across different taxonomic groups. For instance, no evidence was found in

dipterans for increased divergence or composition bias linked to parasitism, and gene order appears conserved between free-living and parasitic taxa (Castro et al. 2002; Li et al. 2020). Furthermore, mitogenomic rearrangements seem prevalent in free-living insects (Shao et al. 2001; Silvestre and Arias 2006; Chen et al. 2011). In planarians (Platyhelminthes), three out of four free-living species studied so far displayed an increased mitochondrial AT content compared to parasitic neodermatans (Solà et al. 2015), and parasitic vampire bats exhibit a mutational trend favouring cytosine over thymine (Botero-Castro et al. 2018). Given these varied results, whether there is a definitive correlation between these facets of mitogenomic evolution and a symbiotic lifestyle remains uncertain, and it is difficult to disentangle lifestyle from other factors that may be at play, such as phylogenetic signal, differences in body size and generation times (Zhang et al., 2021), as well as mutation bias and nutrient availability (Foerstner et al. 2005; Long et al., 2018).

We present Rhabdocoela (Platyhelminthes) as a prime model clade to investigate the mitogenomic effects of symbiosis. Rhabdocoela stands out as the most ecologically diverse and species-rich group of non-neodermatan flatworms (WoRMS 2024). In recent years, this group of microturbellarians has been extensively studied in terms of speciation patterns and habitat transitions (Houben 2013; Van Steenkiste et al. 2013; Stephenson et al. 2018; Tessens et al. 2021). Within Rhabdocoela, an endosymbiotic lifestyle has evolved independently in three clades: Umagillidae, Pterastericolidae, and Graffillinae, all three belonging to the subtaxon Neodalyellida (Van Steenkiste et al. 2013; Stephenson et al. 2018). These lineages differ greatly in species numbers and inhabit a wide variety of marine hosts, including echinoderms (crinoids, sea urchins, sea cucumbers, starfish), sipunculids, bivalves, gastropods, and even some fish species (Jennings 1971; Justine et al. 2009). Reported feeding behaviours are highly diverse, ranging from endozoic predation of cosymbiotic protists to full endoparasitism (Jennings 1997; Doignon and Artois 2006; Cavaleiro et al. 2018). Studies on these animals' life history traits, host specificity, and microhabitat choices are rare, leaving much to be explored (Jennings 1997; Monnens et al. 2019).

The well-established free-living sister lineages of these three endosymbiotic groups of rhabdocoels offer a solid foundation for comparative analyses (Van Steenkiste et al. 2013; Stephenson et al. 2018), permitting multiple tests of hypotheses about the (mitochondrial) genomic implications of lifestyle transitions. Such an approach allows results to be interpreted in a more general framework rather than in an anecdotal one. However, current mitogenomic data on rhabdocoels are limited, covering only one free-living species (Kenny et al. 2019) and three endosymbiotic ones (Monnens et al. 2020). In the latter study, we highlighted the variability in the mt genomes of rhabdocoels concerning gene content and order. We also noted relaxed evolutionary constraints in several protein-coding genes (PCGs). Nevertheless, given the still very small dataset on which this study was based, whether and how these findings relate to these organisms' respective lifestyles remains to be elucidated.

In this study, we expand on the existing mitogenomic dataset by adding representative sequences from umagillids, pterastericolids, and graffillins, along with their sister lineages and closest relatives. Alongside newly acquired nuclear ribosomal data, these sequences are used to position the respective species in the current neodalyellid phylogeny. From this phylogenomic framework, we aim to unravel the evolutionary patterns of these organisms tied to an endosymbiotic lifestyle, focusing primarily on their mitogenomic content and architecture. In light of the referenced studies above, our hypotheses include: (1) an altered AT content, (2) an accelerated substitution rate, (3) a relaxed selection pressure on one or more PCGs, and (4) an increase in gene order changes in endosymbionts compared to the free-living neodalyellids. Finally, we hypothesize that different parasite speciation mechanisms contributed to the diversification of rhabdocoel endosymbionts, and that their diversification patterns are linked to host phylogeny. Therefore, we explore speciation patterns through the first-ever cophylogenetic analysis of non-neodermatan endosymbiotic flatworms.

Material and methods

Sampling and preservation of neodalyellid flatworms

Endosymbiotic neodalyellids were collected from their marine hosts over the course of several sampling campaigns between 2016 and 2019. Field expeditions were carried out in collaboration with the marine station of the University of Plymouth (UK), the Station Biologique de Roscoff (Sorbonne Université, France), the Università degli Studi di Sassari (Sardinia, Italy), the Sven Lovén Centre for Marine Sciences in Kristineberg and Tjärnö (University of Gothenburg, Sweden), and the University of Porto (Porto, Portugal). Host organisms were acquired through SCUBA diving, free-diving, and dredging, or purchased fresh from local fish markets. Host animals were measured and photo-vouchered prior to dissection. Post organ examination, the host's coelomic cavity was flushed with seawater to retrieve any remaining symbionts. Host vouchers were preserved in 70% EtOH and DNA samples were stored in EtOH (99%) or RNAlater™ Stabilization Solution for molecular work.

Free-living neodalyellids were sourced from the intertidal zone in Wimereux, France (summer 2019), Cuba (winter 2017 and spring 2017), and Canada (autumn 2015 and summer 2016). We also made use of specimens previously collected during field trips to Finland (2008) and the Netherlands (2010). Specimens were extracted from sandy sediment or algae using the $MgCl_2$ decantation method, or from mud using the oxygen depletion method (Schockaert 1996). Recovered flatworms were studied and identified alive, photo-vouchered, and stored in EtOH (99%) or RNAlater. When multiple specimens were available, at least one was whole-mounted in lactophenol for morphological study. All flatworm and host vouchers are deposited in the collections of the research group 'Zoology: Biodiversity and Toxicology' at Hasselt University. Sampling locations are provided in **Fig. 1** and a complete overview of species and localities included in this work is available in **Table S1**.

Fig. 1 Map showing sampling locations of neodalyellid flatworms, from A. New Zealand, B. Canada, C. Cuba, and D. Europe. Worms collected from the locality labelled in orange were collected in previous work (Monnens et al. 2019). Details on species and localities are provided in **Table S1.** Country outlines are not to scale.

DNA extraction, library preparation, and sequencing

DNA extractions were performed using a QIAamp DNA Micro Kit (Qiagen) or a custom salting-out protocol tailored for low-input tissues (Dr Chris Laumer, pers. comm.): individual specimens were submerged in TNES buffer containing 0.5 mg/ml Proteinase K (Invitrogen), at 55°C for ± 30 minutes, or until tissue was fully lysed. To the lysate, 65 µl 5M NaCl and 290 µl 96% EtOH were added, using 1.5 µl yeast tRNA (Invitrogen) as a coprecipitant. After 1h storage at -20°C, the precipitated DNA was purified via two EtOH (70%) wash steps and eluted overnight at 4°C in 0.1X TE buffer with 0.02% Tween™ 20 (Thermofisher).

Given the constraints posed by limited DNA quantities, an aliquot of each extract was used as input for whole genome amplification (WGA) with either illustra GenomiPhi™ V2 or illustra TempliPhi™ kits (GE Healthcare/Cytiva). Nucleotide concentrations were quantified on a Qubit™ fluorometer (Life Technologies). Extracts were processed into libraries either in-house using the Illumina Nextera DNA Flex or with the Nextera DNA XT kits at Macrogen Europe. In cases where the initial quality checks were unsatisfactory, the WGA products were used as an alternative. Sequencing was carried out commercially on the Illumina HiSeq X platform (Macrogen Europe), producing 150 bp paired-end reads. The procedure followed for each sample is summarised in **Table S2**. Read pools are publicly available at the SRA repository under Bioprojects PRJNA606139 (flatworms) and PRJNA692190 (hosts) at ncbi.nlm.nih.gov.

Assembly of mitochondrial genomes and ribosomal operons

Obtained reads were trimmed and quality-checked in fastp (Chen et al. 2018) with default settings for paired reads. A seed file was constructed by extracting publicly available mitochondrial sequences from GenBank (Benson et al. 2012). Using this seed file, *de novo* assembly of mt genomes was carried out in Novoplasty v2.7.2 (Dierckxsens et al. 2017) on the Flemish Supercomputer Centre (VSC) Genius cluster. K-mer values were incrementally increased until assemblies were circularised, or until no further improvement was observed. Linear contigs were imported in Unipro UGENE v35.0 (Okonechikov et al. 2012) and scanned for large (>300 bp) overlaps between the 3' and 5' ends. When overlaps were found, termini were merged within this region, hence circularising the assembly. When no or only short (<10 kb) contigs were obtained in Novoplasty, a second *de novo* assembly was performed in SPAdes v3.14.1 (Bankevich et al. 2012) on the VSC cluster with k-mers set at 21, 33, 55, 77, 99, 127. Resulting De Bruijn graphs were visualised in Bandage v0.8.1 (Wick et al. 2015) and mt contigs were baited out by BLASTing (Altschul et al. 1990), using the initial seed file as query.

Ribosomal operons were obtained with the mirabait utility from MIRA v4.0.2 (Chevreux et al. 1999; Cock et al. 2013): trimmed fastq files were filtered according to matches with a precompiled bait file, sourced from GenBank with the search query '(18S OR 28S) AND Neodalyellida'. Positive matches were used as input for a new *de novo* assembly in SPAdes, mirroring the earlier approach. Resulting mt and ribosomal assemblies were subjected to a BLAST search on the NCBI webserver (ncbi.nlm.nih.gov) to check for signs of contamination.

At the time of this study, we were not in possession of molecular-grade samples of *Evechinus chloroticus* (Valenciennes, 1846). Therefore, mt and ribonuclear genes were extracted from the publicly available transcriptome of this species (SRR1014618), guided by the annotations provided by Gillard et al. (2014).

Automatic and manual annotation of mitochondrial and nucleoribosomal genes

Mt PCGs, transfer RNA (tRNAs), and ribosomal RNA (rRNAs) genes were predicted on the MITOS webserver, employing the Echinoderm-Flatworm genetic code (for endosymbionts and echinoderm hosts), or Invertebrate (for mollusc hosts) genetic code (Bernt et al. 2013). Annotations were evaluated according to their respective Expect (E) values and congruence scores within each search. Open reading frames (ORFs) were predicted in Geneious v11.1.5 (Kearse et al. 2012) (transl_table = 9), with the inclusion of two alternative start codons (TTG and ATT) previously reported in platyhelminths (Ye et al. 2014; Bachmann et al. 2016; Ross et al. 2016). Annotation boundaries were tweaked to match ORFs and to minimise non-coding segments. Annotations were verified by comparison with homologous genes in closely related species.

Unidentified ORFs in seemingly non-coding regions were investigated using pBLAST protein-protein comparisons and HMMER searches in SMART (Schultz et al. 1998; Letunic et al. 2015; Letunic and Bork 2017). PCGs anticipated to be present in

flatworm mitogenomes (Wey-Fabrizius et al. 2013), but not detected through automated annotation were sought by pBLASTing candidate ORFs of the expected length. Once positive matches were identified, an amino acid (AA) alignment with homologous rhabdocoel sequences was constructed using MUSCLE v3.8.425 in Geneious (Edgar 2004a, b). These alignments were then examined visually for patterns of similarity and hydrophobicity.

Annotating *atp8* in flatworms often poses challenges due to its high divergence and inherent aligning challenges (**Fig. 2**) (Egger et al. 2017; Monnens et al. 2020). Recognizing that small and variable genes may be overlooked during annotation procedures, additional efforts were made to identify *atp8*, *nad4L*, and *nad6* in assemblies where their presence was not immediately evident. To this end, a custom AA database was built in Geneious, including publicly accessible and newly annotated versions of these PCGs. Assemblies were translated in all six frames and subsequently subjected to a pBLAST search against this tailored database, using default settings and a max E-value of 0.01 as cut-off. Hits matching with predicted ORFs were evaluated and tweaked as described above.

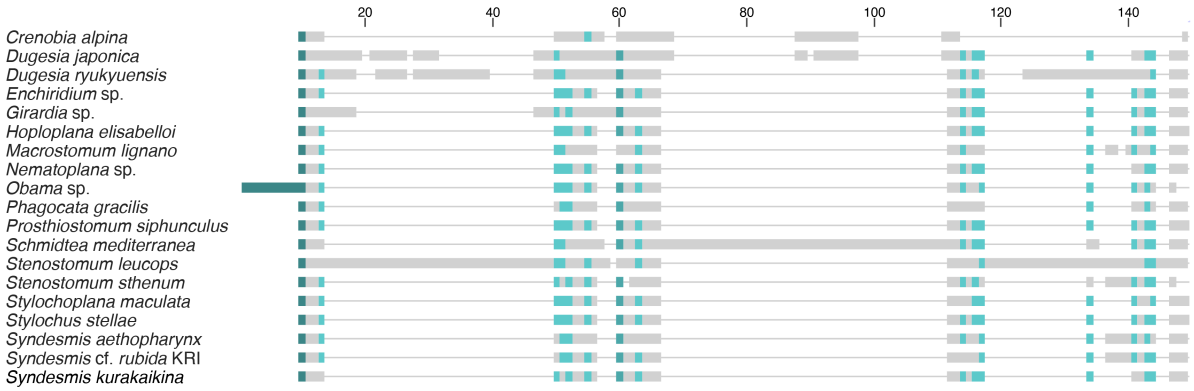


Fig. 2 Amino acid alignment showcasing previously annotated (putative) *atp8* sequences within Platyhelminthes (Ross et al. 2016; Egger et al. 2017; Rosa et al. 2017; Monnens et al. 2020). Only the initial 150 bp region is displayed. The alignment was constructed in MUSCLE v3.8.425 (Edgar 2004a, b) as implemented in Geneious

v11.1.5 (Kearse et al. 2012). Colour key: ■ 100% similarity, ■ 80–100% similarity, ■ 60–80% similarity, ■ below 60% similarity.

Mitos (MiTFi) tRNA predictions were juxtaposed with outputs from tRNAscan-SE v2.0 (Lowe and Chen 2016) and ARWEN (Laslett and Canbäck 2008) and assessed using E-values (MiTFi) and COVE scores (tRNAscan-SE). In instances of conflicting outputs, the secondary structure with the lowest free energy was selected, as determined by the RNAeval tool from the Vienna package (Lorenz et al. 2011). Energy parameters were adjusted to the nearest estimated environmental temperature available (**Table S3**). Repetitive regions were detected using the YASS (Noe and Kucherov 2005) and Tandem Repeats Finder (Benson 1999) platforms. Annotated mt genomes were visualised in Geneious. The origin of each sequence was arbitrarily set at the 5' end of *cox*1. Gene boundaries in nuclear ribosomal operons were predicted in ITSx v1.1.2 (Bengtsson-Palme et al. 2013) and RNAmmer v1.2 (Lagesen et al. 2007) and corroborated by aligning to related species.

All annotated assemblies are publicly available on GenBank (**Tables S2, S4,** and **S6**).

Comparative analysis of mitochondrial genome composition and gene order variation across neodalyellids

Average GC content for each mt genome was calculated in Geneious. Compositional differences between strands were derived using the [(G-C)/(G+C)] and [(A-T)/(A+T)] indices (Perna and Kocher 1995). Variations in nucleotide makeup, along with GC and AT skew degrees, were contrasted between lifestyles using the MCMCglmmm package in RStudio v2023.06.0+421 (Hadfield, 2010; R Core Team 2022; RStudio Team 2022), with the newly inferred phylogenetic tree (see further) specified as a covariance structure to account for shared ancestry among taxa. To minimise potential biases, we excluded sequences obtained through WGA and restricted this analysis to (nearly) complete assemblies (cut-off at 13 kb; Wey-Fabrizius et al. 2013). The most parsimonious scenario for genetic rearrangements in the mt genomes for every neodalyellid subgroup was calculated in CREx (Bernt et al. 2007), considering

circular data. In cases where only fragments of the mitogenome could be assembled, each segment was individually used as input. Nucleotide diversity was visualised for free-living and endosymbiotic species using a sliding window analysis in SVARAP which can be applied for interspecific datasets up to 100 sequences (Khamis et al. 2003; Colson et al. 2006). A 25 bp window was applied. Concatenated mt alignments of PCGs and RNA genes were segmented in 4,000 bp sections to comply with SVARAP's maximum input restriction. The resulting graphs were manually merged in Adobe Illustrator CC.

Phylogenetic inference and testing for relaxation of selection pressure in neodalyellids

Newly annotated and publicly available neodalyellid sequences were used to compile datasets for phylogenetic analyses. The limnotyphloplanid rhabdocoel *Bothromesostoma personatum* (Schmidt, 1848) Braun, 1885 (Typhloplanidae, Rhabdocoela) and three triclads were selected as outgroups. All sequences and corresponding accession numbers are documented in **Tables S2, S4**, and **S6**. A translational alignment was built for 12 PCGs (excluding *atp8*) using the MUSCLE algorithm in Geneious with default settings. Ribosomal sequences were aligned in the online version of MAFFT v7 with the Q-INS-i algorithm, which accounts for secondary structures of RNAs (Katoh and Standley 2013; Katoh et al. 2019). Ambiguously aligned regions were excised from each alignment in Gblocks v0.91b (Castresana 2000; Talavera and Castresana 2007), with settings for a less stringent selection. Trimmed alignments were concatenated in Geneious.

Phylogenetic congruence between mt and nucleoribosomal datasets was evaluated with a hierarchical likelihood-ratio test in Concaterpillar v1.7.2 (Leigh et al. 2008). As congruence was rejected (Weibull-smoothed p-value < 0.01), a partitioned analysis was performed: an initial partitioning scheme was made to subdivide the alignment in genes and, where relevant, codon positions. Best-fitting substitution models were inferred according to the Bayesian Information Criterion in ModelFinder

(Kalyaanamoorthy et al. 2017) on the W-IQ-TREE server (Trifinopoulos et al. 2016), enabling partition merging to determine best-fitting partitioning schemes (Chernomor et al. 2016). A second search was performed to include only those models compatible with MrBayes.

Following the output of ModelFinder (**Table 1**), a Maximum Likelihood (ML) tree search was conducted in the online version of IQ-TREE (Nguyen et al. 2015; Minh et al. 2020). An edge-linked partition model was specified, allowing proportional branch lengths. For support values, 1000 ultrafast bootstrap (UF) iterations and 1000 Shimodaira-Hasegawa approximate likelihood ratio tests (SH) replicates were computed (Guindon et al. 2010; Hoang et al. 2017).

Table 1 Best-scoring partition schemes and associated substitution models according to the Bayesian Information Criterion (BIC) in ModelFinder (Kalyaanamoorthy et al. 2017). Two models are reported for each partition: one represents the best-fitting model among all models implemented in ModelFinder (second column), and the other signifies the best model among those supported by MrBayes (last column).

Partition	**Optimal model**	**Optimal model (MrBayes only)**
rrnS, rrnL	TVM+F+I+G4	GTR+F+I+G4
1st codon positions of *atp6, cox3, cytb, nad1, nad2, nad3, nad4, nad4L, nad5, nad6*		
2nd codon positions of *atp6, cox3, cytb, nad1, nad2, nad3, nad4, nad4L, nad5, nad6*	GTR+F+I+G4	GTR+F+I+G4
3rd codon positions of *atp6, cox1, cox2, cox3, cytb, nad1, nad2, nad3, nad4, nad4L, nad5, nad6*	TIM2+F+I+G4	GTR+F+I+G4
1st codon positions of *cox1, cox2*	GTR+F+I+G4	GTR+F+I+G4
2nd codon positions of *cox1, cox2*		
18S and 28S rDNA	TVM+F+I+G4	GTR+F+I+G4

353 Bayesian inference was executed using MrBayes v3.2.7a (Huelsenbeck and Ronquist
354 2001; Ronquist and Huelsenbeck 2003; Ronquist et al. 2012), specifying linked
355 parameters and best-fitting partitions and substitution models according to **Table 1**.
356 Two independent analyses were undertaken using the Metropolis-coupled Markov
357 chain Monte Carlo algorithm, each comprised of one cold chain and three hot chains.
358 Each analysis was run for 1,000,000 generations, with trees sampled every 1000th
359 generation. The initial 25% of samples were discarded as burn-in. Convergence was
360 assumed when the standard deviation of split frequencies dropped below 0.02. If
361 convergence was not reached after 1,000,000 generations, an additional 500,000
362 generations were performed until convergence was reached. Resultant topologies
363 were summarised in a 50% majority-rule consensus tree. Node support was
364 evaluated using posterior probabilities (pp).
365
366 Inferred topologies were visualised and rooted in FigTree v1.4.4 (Rambaut 2006–2021)
367 and processed in Adobe Illustrator CC. Weakly supported clades were collapsed (UF
368 < 95; SH < 80; pp < 0.90). To assess relative evolutionary rates of mitogenome
369 sequences, the obtained mt topology was imported into TreeGraph v2.15.0 (Stöver
370 and Müller 2010), where branch colours and width were configured to vary in function
371 of the total number of substitutions accumulated per branch.
372
373 To assess the potential relaxation of selection pressure of PCGs among
374 endosymbionts, individual gene trees for each PCG were constructed using IQ-TREE,
375 following the methodology described above. Outgroups were pruned via the iTOL
376 webserver (Letunic and Bork 2021). RELAX tests from the HyPhy package (Wertheim
377 et al. 2015) were used to evaluate relaxation of selection pressure. Through the HyPhy
378 Phylotree portal (https://phylotree.hyphy.org/) endosymbiotic lineages were
379 designated as test branches and free-living neodalyellids as reference groups. The
380 null hypothesis of no relaxation of selection pressure was rejected when p-values <
381 0.05 were inferred. For *nad2*, *nad3*, and *nad4L*, RELAX produced a convergence
382 warning, affecting the reproducibility of the results. Following the developers'
383 suggestion, we considered the likelihood ratios and corresponding AICc scores for

these cases instead of the calculated p-values (Pers. Comm., see also methodology in Monnens et al. 2020).

Host phylogeny and cophylogenetic analyses

As taxonomic coverage of pterastericolids and graffillids remains comparatively low, the cophylogenetic part of this study is focused on umagillids and their hosts. With our sampling focus, we were able to include primarily echinoid- and holothuroid-inhabiting neodalyellids, along with a single sipunculid-infecting species. The neodalyellid topology was pruned from all other taxa and a host topology was constructed using newly acquired and pre-existing (GenBank) sequences, following the same methodology described above (accession numbers listed in **Table S5**). Host and symbiont trees were used as input files for downstream analyses.

Several methods exist to formally assess congruence between symbionts and hosts, each based on different data, algorithms and assumptions (de Vienne, Refrégier et al. 2013). An event-based reconciliation analysis was performed in CoRe-PA v0.5.2 (Merkle, Middendorf et al. 2010), as this software allows for the automatic assignment of evolutionary costs of speciation events. Cost values were calculated using the simplex method on the quality function and the algorithm was run for 5,000 random cycles. Root-to-root mapping was enforced, host switches were permitted, and the chronological consistency of events was checked.

Concerns have been raised as to the extent to which event-based methods assume and maximise co-speciation (de Vienne, Refrégier et al. 2013, Hoberg and Brooks 2008). As such, we also carried out a distance-based test for correlation. This ‘global fit’ test was performed using the ‘paco’ package in R (Balbuena, Míguez-Lozano et al. 2013, Hutchinson, Cagua et al. 2017). This algorithm tests the null hypothesis that host and symbiont phylogenies are randomly associated. Distance matrices for host and symbiont datasets were built from Newick files on the T-rex web server (Boc,

414 Diallo et al. 2012), and a host-symbiont link file was manually constructed. A total of
415 1000 permutations were performed and default settings were employed.

Results

Composition and architecture of mitochondrial genomes

In this study, 47 new mitochondrial genomes were assembled. Of these, 28 were either complete or nearly so, with an expected length of at least 13,000 bp and containing all anticipated PCGs as per Wey-Fabrizius et al. (2013). Every assembly is AT-rich, reaching up to 75%, and displays a positive GC and negative AT skew (**Table S4**).

Endosymbiotic neodalyellids exhibit a marginally higher AT% compared to free-living taxa (posterior mean = 0.066); however, this difference was not statistically significant at the 0.05 level (pMCMC = 0.1). Similarly, endosymbionts show no significant difference in mt genomic AT skew compared to free-living taxa (posterior mean = -0.005; pMCMC = 0.918). Endosymbiotic neodalyellids exhibit significantly higher mt genomic GC skew compared to their free-living counterparts (posterior mean = 0.1012; pMCMC < 0.05).

Nucleotide diversity patterns between endosymbionts and free-living taxa are largely similar (**Fig. 3**). However, there seems to be a subtle trend towards increased variation in *rrnS*, *cox2*, and genes coding for NADH dehydrogenases among the free-living groups. The genes *rrnS, atp6, cox2, cox3, nad2, nad4, nad4L,* and *nad6* contain regions with peak variation, while the most conserved regions in the alignment are observed in *rrnL, cox1, cytb, nad4,* and *nad5.*

Full-length mt genomes comprise 12 PCGs, along with a small and large rRNA subunit gene (*rrnS* and *rrnL*) and 22 tRNAs, all transcribed in the same direction. Putative *atp8* genes were identified in sequences of representatives of several umagillids (*Anoplodium, Anoplodiopsis, Syndesmis, Umagilla*), as well as within members of the free-living family Provorticidae, including *Provortex*, *Pogaina*, and *Vejdovskya.* The corresponding amino acid alignment is visually presented in **Fig. 4**. Multiple genes

were predicted to end in an incomplete stop codon (T). In the case of *Vejdovskya ignava* Ax, 1951, the alternative start codon ATT was discerned in *nad3*, *cox3*, *nad6*, *cytb*, and *nad4*. A total of 24 unique neodalyellid gene orders were inferred, differentiated by a combination of predicted duplications or deletions, reversals, transpositions, reverse transpositions, and tandem-duplication-random-loss events (TDRLs) (**Fig. 5**).

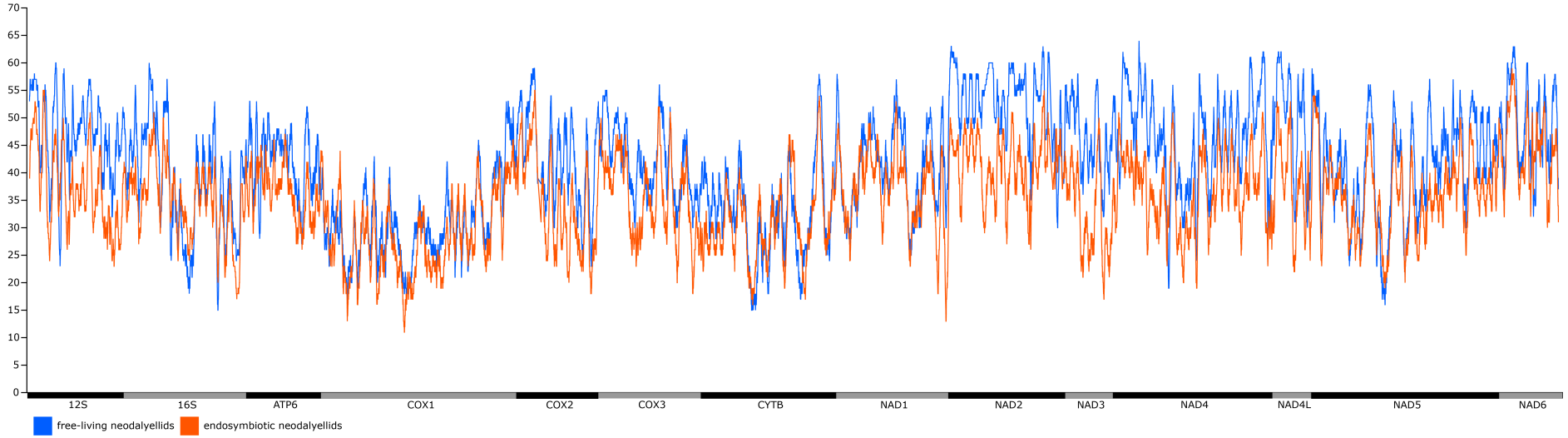


Fig. 3 Nucleotide diversity in mitochondrial protein-coding genes in representatives of Neodalyellida. Graphs were generated separately for free-living and endosymbiotic neodalyellids in SVARAP (Colson et al. 2006) with a window size of 25 bp. Values on the y-axis represent the mean variability (%) per sliding window. SVARAP graphs were processed and annotated in Adobe Illustrator CC.

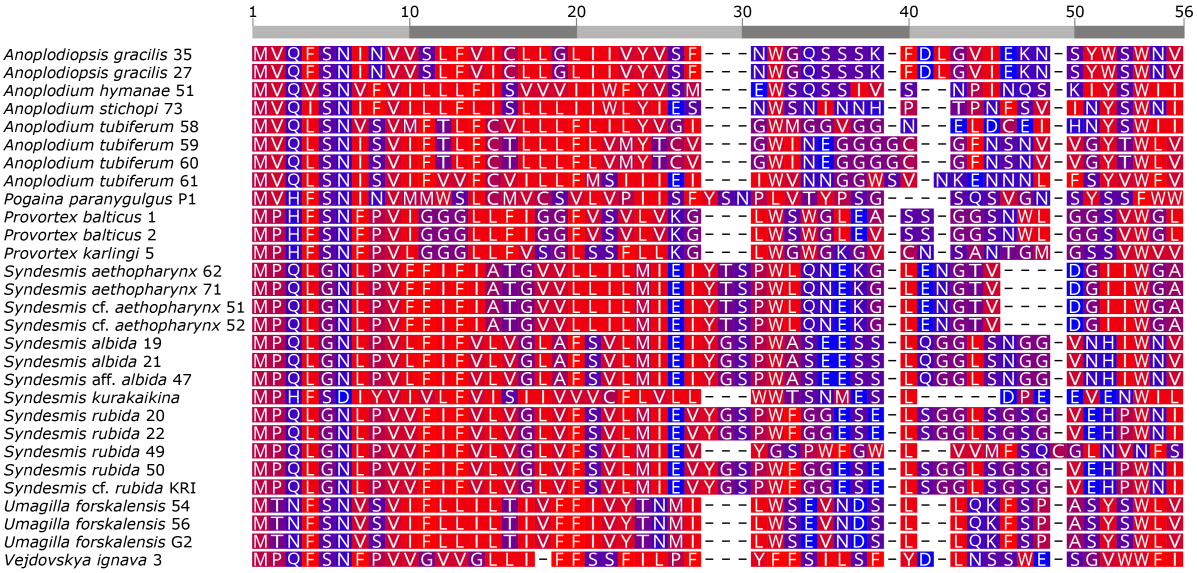


Fig. 4 Amino acid alignment of putative *atp8* genes in mitochondrial genomes of Neodalyellida. The alignment was constructed using MUSCLE v3.8.425 (Edgar 2004a, b) in Geneious v11.1.5 (Kearse et al. 2012). Residues are colour-coded based on hydrophobicity.

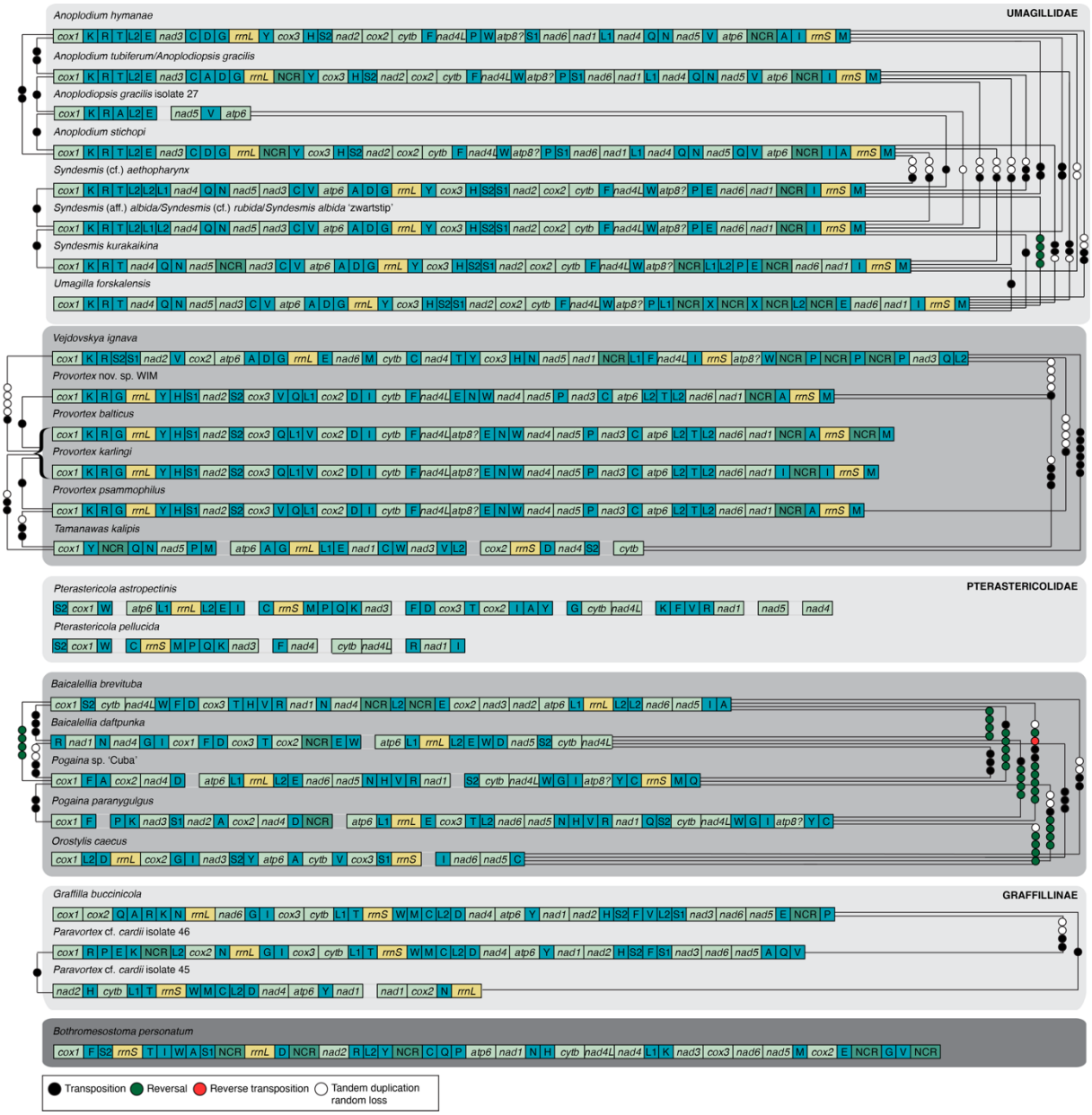


Fig. 5 Relative arrangement of mitochondrial gene order in Rhabdocoela. The origin of each mitogenome was arbitrarily set at the starting position of *cox1*. An X denotes transfer RNA genes with unidentified anticodons. Endosymbiotic taxa are highlighted in light grey and positioned above their phylogenetically closest free-living relatives (in dark grey). *Bothromesostoma personatum*, not belonging to Neodalyellida, is shown separately in the darkest grey. Circles on linking lines represent the most parsimonious scenario for genetic rearrangements, as predicted in CREx (Bernt et al. 2007).

Gene order varies throughout Rhabdocoela, as illustrated in **Fig. 5**, and predicted rearrangements will be detailed per clade in subsequent sections. Observed variations in the genetic content of new mitogenomes from the 'standard' set of mitochondrial genes are primarily due to transfer RNA duplications. Several mitogenomes exhibit multiple *trnL2* genes (**Fig. 5, Tables 2, S4**). This duplication is evident in all isolates of the genera *Graffilla, Provortex*, and *Syndesmis*, as well as the largest obtained contig of *Paravortex* cf. *cardii*. Each of these assemblies contains one copy with a TTG and another with a TTA anticodon. The mitogenome of *Baicalellia brevituba* (Luther, 1918) Nasonov, 1930 contains three *trnL2* copies, each with a TTA anticodon. Two copies of *trnF* were retrieved in one isolate of *Syndesmis rubida* Kozloff & Westervelt, 1990 from Plymouth (isolate TA; E-values 1.571e-10, 5.16e-10), but this was not observed in conspecific sequences. In *Anoplodium tubiferum* Westblad, 1953 and *Anoplodiopsis gracilis* (Wahl, 1906) Westblad, 1953, two NCRs of variable length are apparent: one downstream of *atp6* and one between *rrnL* and *trnY*.

In many assemblies, the non-coding region (NCR) is repeat rich. In certain species, such as *Paravortex* cf. *cardii, Umagilla forskalensis* Wahl, 1909, and *Vejdovskya ignava*, several tRNA-like cloverleaf structures were detected within the NCR. However, retrieval of these structures is inconsistent across software packages, and when obtained, these annotations exhibit relatively low support compared to the other "canonical" tRNA structures in the same assemblies. Considering the potential influence of repeats on the reliability of sequencing results and the known tendency of the Multiple Displacement Amplification (MDA) technique—employed in some of these samples—to introduce repetitive artefacts in assemblies, we opt not to address this aspect for the remainder of this study.

Genetic rearrangements throughout the newly obtained neodalyellid phylogeny

After the removal of ambiguously aligned positions, the concatenated mt and ribosomal alignments contained 11,979 bp and 3,440 bp respectively. Well-supported clades are identical in the inferred ML and BI topologies (**Fig. 6**). Separate ribosomal and mt topologies are provided in supplementary files (**Fig. S1**). The per branch evolutionary rate of the mt topology is visualised in **Fig. 7**.

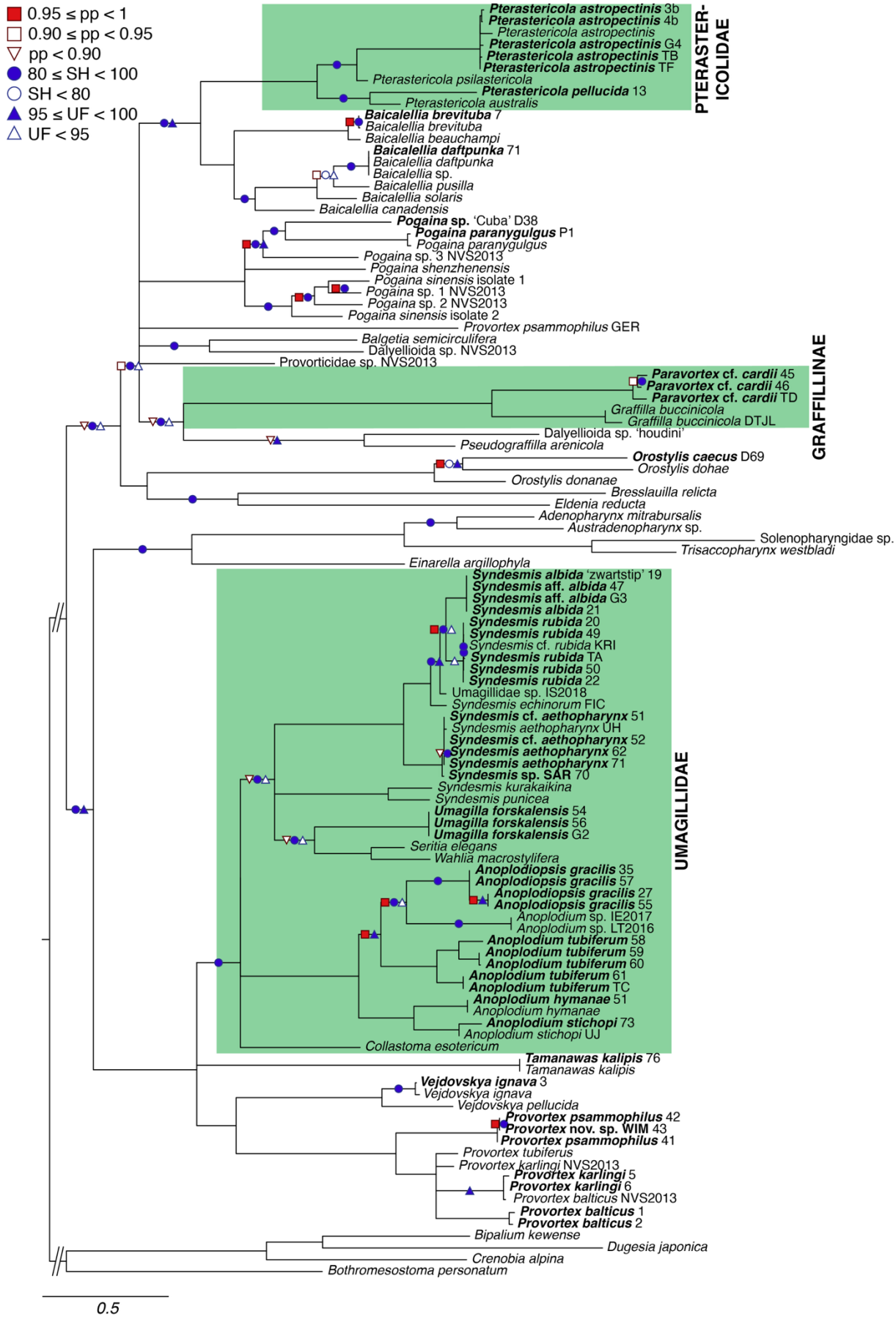

0.95 ≤ pp < 1
0.90 ≤ pp < 0.95
pp < 0.90
80 ≤ SH < 100
SH < 80
95 ≤ UF < 100
UF < 95
Pterastericola astropectinis 3b
Pterastericola astropectinis 4b
Pterastericola astropectinis
Pterastericola astropectinis G4
Pterastericola astropectinis TB
Pterastericola astropectinis TF
Pterastericola psilastericola
Pterastericola pellucida 13
Pterastericola australis
PTERASTER-ICOLIDAE
Baicalellia brevituba 7
Baicalellia brevituba
Baicalellia beauchampi
Baicalellia daftpunka 71
Baicalellia daftpunka
Baicalellia sp.
Baicalellia pusilla
Baicalellia solaris
Baicalellia canadensis
Pogaina sp. 'Cuba' D38
Pogaina paranygulgus P1
Pogaina paranygulgus
Pogaina sp. 3 NVS2013
Pogaina shenzhenensis
Pogaina sinensis isolate 1
Pogaina sp. 1 NVS2013
Pogaina sp. 2 NVS2013
Pogaina sinensis isolate 2
Provortex psammophilus GER
Balgetia semicirculifera
Dalyellioida sp. NVS2013
Provorticidae sp. NVS2013
Paravortex cf. cardii 45
Paravortex cf. cardii 46
Paravortex cf. cardii TD
Graffilla buccinicola
Graffilla buccinicola DTJL
GRAFFILLINAE
Dalyellioida sp. 'houdini'
Pseudograffilla arenicola
Orostylis caecus D69
Orostylis dohae
Orostylis donanae
Bresslauilla relicta
Eldenia reducta
Adenopharynx mitrabursalis
Austradenopharynx sp.
Solenopharyngidae sp.
Trisaccopharynx westbladi
Einarella argillophyla
Syndesmis albida 'zwartstip' 19
Syndesmis aff. albida 47
Syndesmis aff. albida G3
Syndesmis albida 21
Syndesmis rubida 20
Syndesmis rubida 49
Syndesmis cf. rubida KRI
Syndesmis rubida TA
Syndesmis rubida 50
Syndesmis rubida 22
Umagillidae sp. IS2018
Syndesmis echinorum FIC
Syndesmis cf. aethopharynx 51
Syndesmis aethopharynx UH
Syndesmis cf. aethopharynx 52
Syndesmis aethopharynx 62
Syndesmis aethopharynx 71
Syndesmis sp. SAR 70
Syndesmis kurakaikina
Syndesmis punicea
Umagilla forskalensis 54
Umagilla forskalensis 56
Umagilla forskalensis G2
Seritia elegans
Wahlia macrostylifera
Anoplodiopsis gracilis 35
Anoplodiopsis gracilis 57
Anoplodiopsis gracilis 27
Anoplodiopsis gracilis 55
Anoplodium sp. IE2017
Anoplodium sp. LT2016
Anoplodium tubiferum 58
Anoplodium tubiferum 59
Anoplodium tubiferum 60
Anoplodium tubiferum 61
Anoplodium tubiferum TC
Anoplodium hymanae 51
Anoplodium hymanae
Anoplodium stichopi 73
Anoplodium stichopi UJ
Collastoma esotericum
UMAGILLIDAE
Tamanawas kalipis 76
Tamanawas kalipis
Vejdovskya ignava 3
Vejdovskya ignava
Vejdovskya pellucida
Provortex psammophilus 42
Provortex nov. sp. WIM 43
Provortex psammophilus 41
Provortex tubiferus
Provortex karlingi NVS2013
Provortex karlingi 5
Provortex karlingi 6
Provortex balticus NVS2013
Provortex balticus 1
Provortex balticus 2
Bipalium kewense
Dugesia japonica
Crenobia alpina
Bothromesostoma personatum
0.5

Fig. 6 Topology of Neodalyellida inferred from Bayesian and Maximum Likelihood analyses of the combined nuclear and mitochondrial genomic dataset. Clades with posterior probabilities below 0.90 were collapsed. Symbols on branches represent support values, as indicated in the upper left corner. Nodes without labels have maximum support. Branch lengths denote the number of expected nucleotide substitutions per site. Taxa newly sequenced in this work are marked in bold and endosymbiotic lineages are labelled in green. Two root branches were cut short for visibility purposes, as indicated by the // symbol in the tree. Abbreviations: pp, posterior probabilities. SH, Shimodaira-Hasegawa approximate likelihood ratio values. UF, ultrafast bootstrap values. Please note that, based on the genus revision proposed by Cavaleiro et al. (2018), and considering the results of the current molecular work, the species previously identified as *S. echinorum* in our earlier study (Monnens et al. 2020) should likely be reclassified as *S. rubida*. Unfortunately, the voucher material for this species has been lost, so this sequence is labelled as *Syndesmis* cf. *rubida* KRI in this topology, and throughout this manuscript.

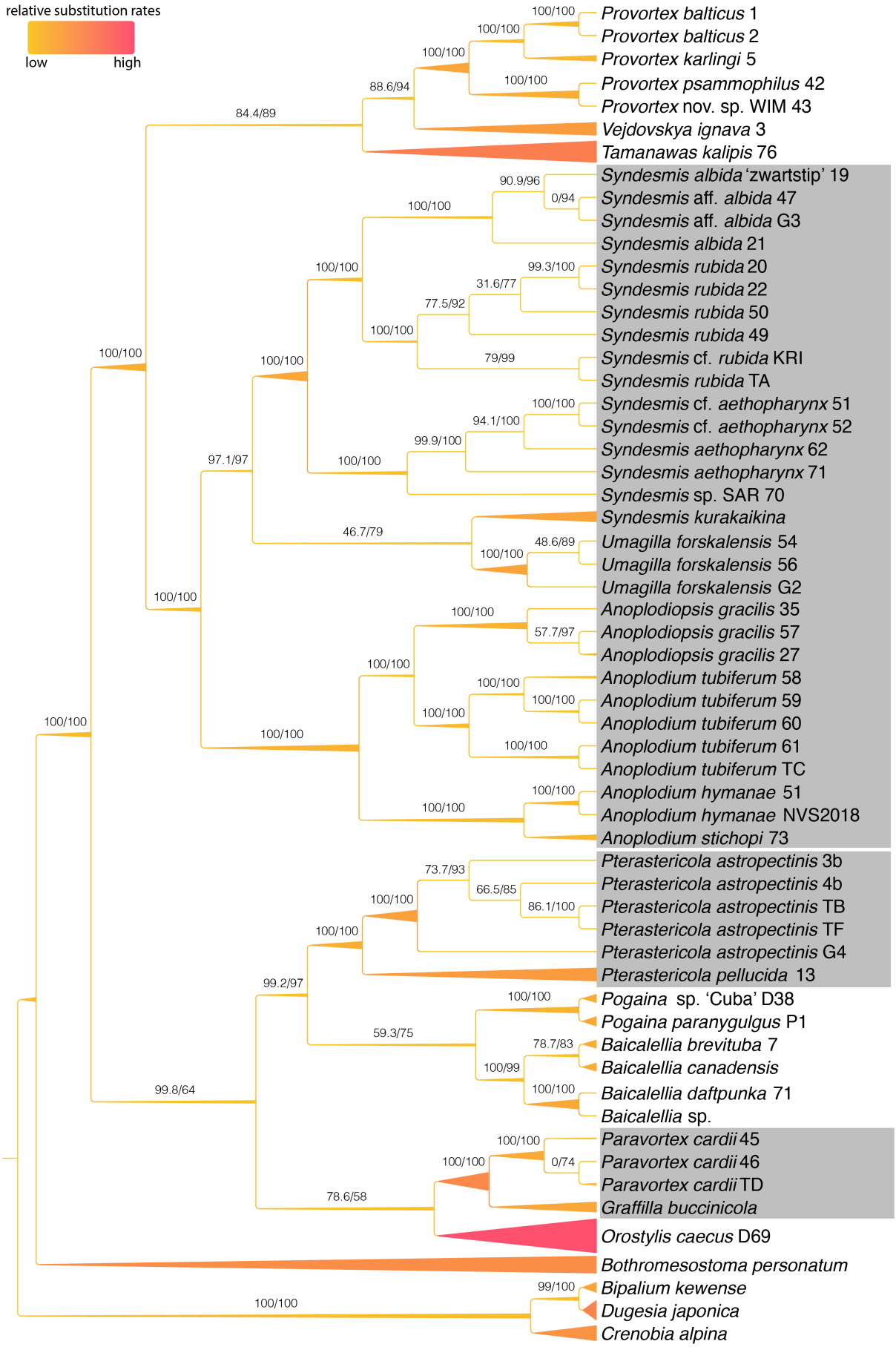

relative substitution rates
low
high
Provortex balticus 1
Provortex balticus 2
Provortex karlingi 5
Provortex psammophilus 42
Provortex nov. sp. WIM 43
Vejdovskya ignava 3
Tamanawas kalipis 76
Syndesmis albida 'zwartstip' 19
Syndesmis aff. albida 47
Syndesmis aff. albida G3
Syndesmis albida 21
Syndesmis rubida 20
Syndesmis rubida 22
Syndesmis rubida 50
Syndesmis rubida 49
Syndesmis cf. rubida KRI
Syndesmis rubida TA
Syndesmis cf. aethopharynx 51
Syndesmis cf. aethopharynx 52
Syndesmis aethopharynx 62
Syndesmis aethopharynx 71
Syndesmis sp. SAR 70
Syndesmis kurakaikina
Umagilla forskalensis 54
Umagilla forskalensis 56
Umagilla forskalensis G2
Anoplodiopsis gracilis 35
Anoplodiopsis gracilis 57
Anoplodiopsis gracilis 27
Anoplodium tubiferum 58
Anoplodium tubiferum 59
Anoplodium tubiferum 60
Anoplodium tubiferum 61
Anoplodium tubiferum TC
Anoplodium hymanae 51
Anoplodium hymanae NVS2018
Anoplodium stichopi 73
Pterastericola astropectinis 3b
Pterastericola astropectinis 4b
Pterastericola astropectinis TB
Pterastericola astropectinis TF
Pterastericola astropectinis G4
Pterastericola pellucida 13
Pogaina sp. 'Cuba' D38
Pogaina paranygulgus P1
Baicalellia brevituba 7
Baicalellia canadensis
Baicalellia daftpunka 71
Baicalellia sp.
Paravortex cardii 45
Paravortex cardii 46
Paravortex cardii TD
Graffilla buccinicola
Orostylis caecus D69
Bothromesostoma personatum
Bipalium kewense
Dugesia japonica
Crenobia alpina

Fig. 7 Visualisation of relative substitution rates in the neodalyellid mt phylogeny, as plotted on the uncollapsed mitochondrial Maximum Likelihood tree in TreeGraph v2.15.0 (Stöver and Müller 2010). Endosymbiotic taxa are marked in grey. Divergence is visualised by branch widths and colours, the scale of which is indicated in the top left. Values above branches respectively represent SH, Shimodaira-Hasegawa approximate likelihood ratio values and UF, ultrafast bootstrap values.

The monophyly of Neodalyellida is maximally supported, with a deep dichotomy at its root. All three endosymbiotic lineages are recovered as monophyletic, with robust support (SH/UF/pp 94.4/100/1 for Umagillidae and maximal support for Pterastericolidae and Graffillinae).

Two major lineages within Umagillidae correspond to (1) a clade encompassing *Anoplodium* and *Anoplodiopsis* and (2) a partially resolved group of remaining umagillid genera. Three similar, but unique gene orders appear in clade (1), attributed to predicted transpositions of *trnA* and *trnP*, as well as an apparent duplication of *trnQ* in *A. stichopi*. In (2), gene order is more or less consistent, but variations exist due to transpositions of *trnP*, *trnL1*, and *trnL2*, with the latter being duplicated in all isolates of *Syndesmis* except for *S. kurakaikina*. *Syndesmis*, the largest genus of endosymbiotic neodalyellids, had the most sequences included, though its interspecific relationships remain unresolved. Gene order within this genus varies due to duplications and/or transpositions of *trnL1* and *trnL2*.

Among the closest free-living relatives of Umagillidae, variation in mt gene order is also apparent, with four TDRLs and one transposition distinguishing *Provortex* from *Vejdovskya*. Gene order within *Provortex* is also variable, requiring at least one tRNA transposition per included species, as well as an apparent *trnA* deletion for *P. karlingi*. Due to the distant position of a previously published *P. psammophilus* sequence (AY157162), we suspect misidentification and exclude it from further discussion here. The obtained assembly of *Tamanawas* is fragmented but shows distinct gene order, with at least five transpositions separating it from *Vejdovskya* and at least two

553 transpositions and one TDRL separating it from *Provortex* in the most parsimonious
554 scenario.
555
556 Deep interrelationships between the endosymbiotic taxa Graffillinae and
557 Pterastericolidae, and related free-living neodalyellids are not fully resolved. Two
558 genera of Graffillinae were included here, with *Paravortex* exhibiting gene order
559 variation across two isolates: one fully assembled mt genome and a partial sequence
560 from a second isolate. The variation is primarily due to a transposition involving the
561 *nad2-trnH* block, with additional variation related to the position of *nad1* in *Paravortex*
562 cf. *cardii* isolate 45. At least two transpositions and two TDLRs separate the mt
563 genome of *Paravortex* from that of *G. buccinicola*.
564 Several free-living neodalyellids closely related to Graffillinae were included in the
565 analysis, but interrelationships remain unresolved due to a deep polytomy. Gene order
566 varies between taxa, including at the intrageneric level, as seen in *Pogaina*, where mt
567 genomic fragments from two included species show variability in the positions of *trnA*,
568 *trnE*, *trnT*, and *trnQ*.
569
570 Also Pterastericolidae is monophyletic with maximal support, with well-resolved
571 interspecific relationships. Despite obtaining several mt sequences from all isolates,
572 the fragmented assemblies hinder the reconstruction of genetic rearrangements.
573 *Baicalellia* comprises the free-living sister lineage to Pterastericolidae with maximal
574 support. Gene order within *Baicalellia* varies, and although we could not circularise
575 the mt genome of all species, at least three transpositions are predicted.
576

577 ## Testing for relaxed selection pressure in endosymbionts

578 A significant ($p < 0.05$) reduction in selection pressure was observed in five PCGs
579 (*atp6*, *cox1*, *cox2*, *cytb*, and *nad1*) of endosymbiotic lineages when compared to their
580 free-living counterparts (**Table 2**). Apart from *nad3* ($K = 1.01$), all PCGs exhibit trends
581 towards relaxation to varying degrees, as indicated by K values < 1, although these

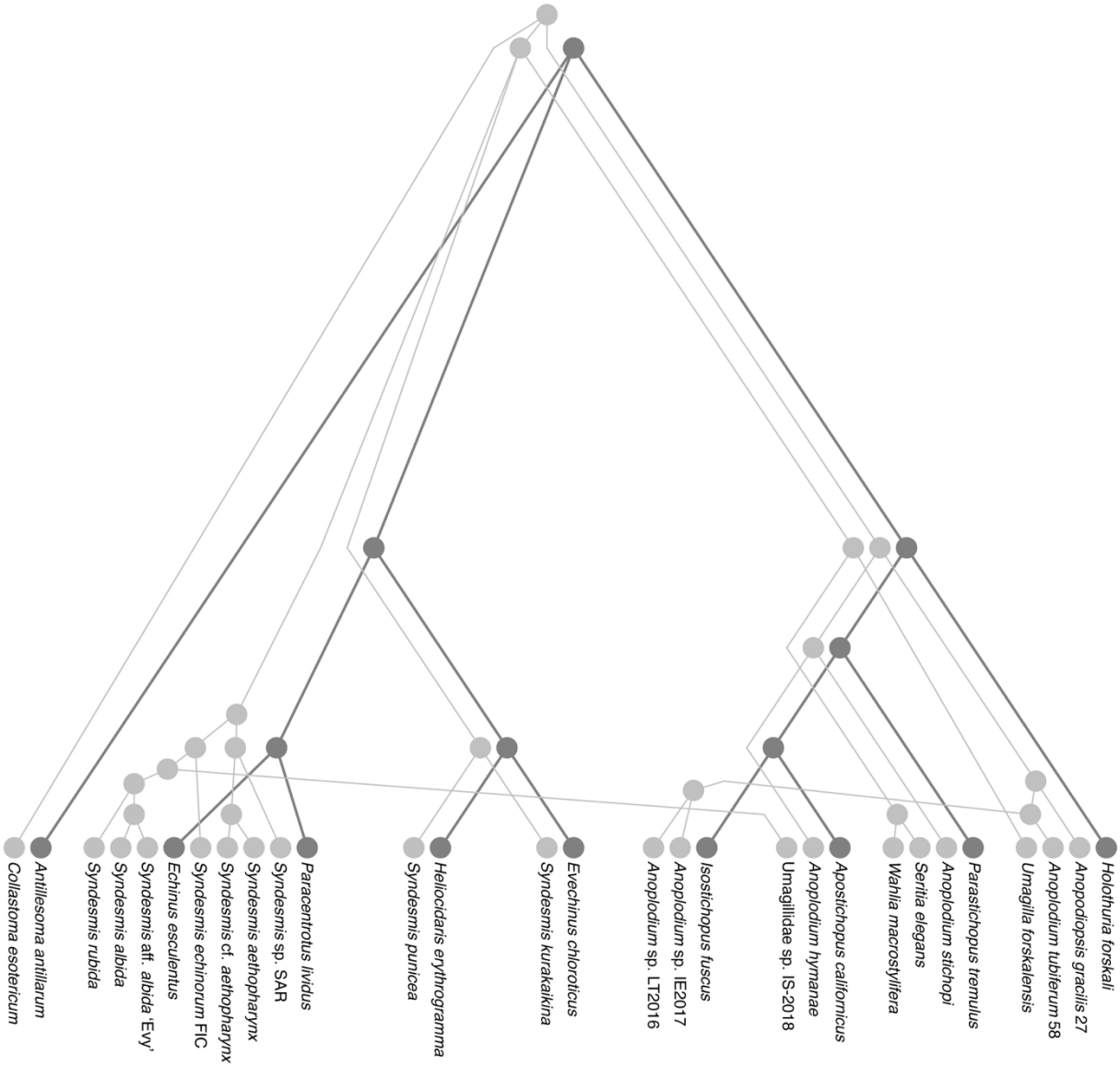
outcomes did not reach statistical significance for all genes ($p > 0.05$). No significant intensification of selection pressure was inferred.

Table 2. Comparison of selection pressure in protein-coding genes (PCG) of neodalyellid mitochondrial genomes between endosymbiotic (test branches) and free-living taxa (reference branches). Analyses were performed in RELAX (Wertheim et al. 2015). The table presents the inferred selection intensity parameters (K), log-likelihood scores, delta AICc values for null and relaxed models, and corresponding p-values. Statistically significant results are indicated with asterisks (*$p < 0.05$, **$p < 0.001$). P-values of datasets for which a convergence warning was generated are greyed out.

PCG	K	$Log(L)_{null}$	$Log(L)_{relax}$	$\Delta AICc_{(null - relax)}$	p-value
atp6	0.70	-12108.12	-12101.17	11.86	0.0002**
cox1	0.42	-24621.46	-24601.36	38.17	0.0000**
cox2	0.67	-12135.18	-12127.69	12.94	0.0001**
cox3	0.61	-14674.99	-14673.85	0.25	0.1307
cytb	0.74	-21292.41	-21283.67	15.47	0.0000**
nad1	0.64	-17330.9	-17328.60	2.66	0.0303*
nad2	0.96	-14736.80	-14736.80	-2.02	0.9586
nad3	1.01	-6512.51	-6512.52	-2.05	0.8683
nad4	0.84	-24894.49	-24893.81	-0.67	0.2435
nad4L	0.98	-4831.33	-4831.33	-2.09	0.9849
nad5	0.85	-29718.58	-29716.97	1.21	0.0722
nad6	0.97	-8184.35	-8184.28	-1.92	0.7070

Cophylogenetic analyses

CoRe-PA inferred a total of twelve solutions for the co-evolutionary history of umagillids and their hosts (**Table S7**). The best-scoring scenario from this analysis, which had a minimised quality score of less than 0.001, is presented in **Fig. 8**. The total cost for this scenario is 4.53, comprising six cospeciation events (cost 0.189), eight lineage sortings (cost 0.142), 11 duplications (cost 0.103), and two host switches (cost 0.566). A significant global fit between host and symbiont phylogenies was established (sum of squared residuals 3.17; $p < 0.001$). Individual contributions of interactions are quantified in **Table 3** in terms of observed residuals of the Procrustean superimposition and Jackknife estimates. Links with lower values of these metrics reflect high contributions to cophylogenetic matching.



606

607 **Fig. 8** Cophylogenetic reconciliations of the umagillid (light grey) and corresponding
608 echinoderm (dark grey) tree as predicted in CoRe-PA.

609

610 **Table 3** Interaction-specific host(H)-symbiont(S) contributions to the cophylogenetic
611 signal expressed in raw residuals of the Procrustean superimposition and a
612 jackknifing procedure.

	Interaction	Residuals	Jackknife
H	*Antillesoma antillarum*		
S	*Collastoma esotericum*	7.926	16.426
H	*Echinus esculentus*		

S	*Syndesmis* aff. *albida* 'Evy'	0.894	1.663
S	*Syndesmis albida*	0.895	1.666
S	*Syndesmis rubida*	0.920	1.726
H	*Paracentrotus lividus*		
S	*Syndesmis* sp. SAR	1.050	2.077
S	*Syndesmis aethopharynx*	1.047	2.070
S	*Syndesmis* cf. *aethopharynx*	1.047	2.070
S	*Syndesmis echinorum* FIC	1.145	2.246
H	*Heliocidaris erythrogramma*		
S	*Syndesmis punicea*	2.236	4.490
H	*Evechinus chloroticus*		
S	*Syndesmis kurakaikina*	2.268	4.590
H	*Isostichopus fuscus*		
S	*Anoplodium* sp. IE2017	0.780	0.943
S	*Anoplodium* sp. LT2016	0.780	0.943
H	*Apostichopus californicus*		
S	*Anoplodium hymanae*	1.213	2.228
S	Umagillidae sp. IS-2018	3.855	7.792
H	*Parastichopus tremulus*		
S	*Anoplodium stichopi*	1.597	3.009
S	*Seritia elegans*	2.254	4.775
S	*Wahlia macrostylifera*	2.478	5.308
H	*Holothuria forskali*		
S	*Anoplodiopsis gracilis*	2.149	4.509
S	*Anoplodium tubiferum*	2.132	4.489
S	*Umagilla forskalensis*	3.557	7.473

613

Discussion

In this study, we investigated the molecular-genetic patterns associated with the transition from a free-living to an endosymbiotic lifestyle in neodalyellid flatworms. To investigate several previously proposed hypotheses concerning the changes in mitochondrial genomic architecture linked with such an evolutionary transition, we constructed a uniquely extensive dataset comprising 50 complete or partial mitogenomic sequences, with 47 being novel to science. Leveraging this dataset, we established a robust phylogenomic framework, allowing for an in-depth exploration of molecular-genetic signatures associated with lifestyle shifts. In the following section, we explore the intricacies of the newly obtained neodalyellid mitogenomes and contextualise these with our five initial hypotheses.

Mitogenomic architecture and potential for new rhabdocoel markers

The novel sequence data presented in this study dramatically expand the molecular record of rhabdocoel flatworms, augmenting the publicly available mitochondrial genomes by over tenfold and more than doubling the number of published neodalyellid ribosomal sequences. All newly assembled mt genomes display the characteristic platyhelminth feature of being AT-rich, accompanied by a positive GC and negative AT skew (Wey-Fabrizius et al. 2013). Although most samples resulted in large contigs (up to 18 kb), a minority of assemblies yielded only one or several incomplete fragments (**Table S4**), containing only a portion of the expected mt genes. The short reads generated by the HiSeq X platform (150 bp, Illumina, Inc.) may have been insufficient for assembling some of the more repeat-rich mitogenomes in this context, and this limitation could be addressed by using long-read technologies, such as those offered by PacBio or Ion Torrent platforms.

All sequences were translated according to the echinoderm-flatworm mt genetic code. Many PCGs were predicted to end in the abbreviated stop codon T and are

presumably converted to TAA by post-transcriptional polyadenylation, a well-established phenomenon in metazoan mt genomes (Wolstenholme 1992). Additionally, we observed indications of the usage of the non-canonical start codon ATT in five PCGs of *Vejdovskya ignava*, a peculiarity anecdotally reported in monogeneans (Bachmann et al. 2016; Zhang 2019). While transcriptomic data are needed to verify this, if confirmed, this discovery would represent the first documentation of an ATT start codon in turbellarian flatworms.

Sliding window analyses revealed largely congruent patterns and comparable nucleotide diversity levels between endosymbionts and free-living taxa (**Fig. 3**). Exceptions are found in two NADH dehydrogenase genes, as *nad2* and the starting region of *nad4* appear somewhat more conserved among endosymbiotic lineages. For the most part, mitogenomic signatures of lifestyle shifts seem, however, not to be occurring at the level of additional or differing nucleotide substitutions, but rather in traits related to genomic architecture (see further). The largest peaks are found in *cox2*, *nad4L*, *nad6*, *rrnS*, and the starting region of *rrnL*. For free-living taxa, this also includes the *nad4* and *nad2* genes. The observed variability patterns are largely congruent with mt genomic characterisations in monogeneans (Vanhove et al. 2018; Zhang et al. 2018), where *nad2* and *nad6* (as well as *atp6* and *nad5*) were established as fastest evolving markers.

The high variability in these regions renders them compelling candidate markers for exploring neodalyellid interrelationships at the finest taxonomic level. Some of these peaks are flanked by more conserved segments (*cox2, rrnL*), making them easily accessible targets for sequencing. Public availability of our newly assembled mt genomes will enable future researchers to develop lineage-specific primers for these animals, hence surpassing 'universal' barcoding primers, which often underperform in platyhelminths (Vanhove et al. 2013).

Mitochondrial gene composition and variation in neodalyellids

The vast majority of (nearly) complete assemblies carry the 'standard' mt genes of flatworms: 12 PCGs (excluding *atp8*, see further) encoding enzyme subunits of the oxidative phosphorylation chain, a small and large ribosomal RNA gene (*rrnS* and *rrnL*), and 22 transfer RNA genes (Wey-Fabrizius et al. 2013). Deviations from this pattern are primarily attributed to duplicated tRNA genes, including *trnP* in *Vejdovskya ignava, trnQ* in *Anoplodium stichopi* Bock, 1925, *trnI* in *Provortex karlingi*, and an unidentified (*trnX*) gene in *Umagilla forskalensis* (**Table S4**). By far the most duplicated gene is *trnL2*, and this is observed throughout various free-living and endosymbiotic lineages, including *Baicalellia* (with three copies), *Paravortex, Provortex, Syndesmis*, and *Umagilla*. The position of *trnL2* also varies between all lineages (**Fig. 5**). No instances of duplicated protein-coding genes were identified, although some unidentified ORFs exist, including one in *Anoplodium tubiferum* (located downstream of *rrnL*) and another in *Tamanawas kalipis* Stephenson, Van Steenkiste & Leander, 2019 (situated downstream of *trnY*).

Putative copies of the ATP synthase gene *atp8* were identified in all umagillids and in several free-living neodalyellids through a combination of automatic annotation and reverse pBLASTing (**Figs. 5, 6**). Detection of this short, highly variable gene is facilitated by its putative annotation in two previously published umagillid sequences (NC_050392, NC_050391). Although not identified in all assemblies, our intention is not to assert the absence of *atp8* in these respective species. Rather, we cautiously acknowledge the current insufficiency of data at our disposal to confidently annotate this gene, as discussed by Monnens et al. (2020). A potential connection with any particular lifestyle remains unclear; while *atp8* is lost in neodermatan (fully parasitic) platyhelminths (McManus et al. 2004; Hardman and Hardman 2006; Egger et al. 2017), putative copies have recently been detected in various turbellarian lineages, including free-living and obligate symbiotic species alike (Ross et al. 2016; Egger et al. 2017; Rosa et al. 2017; Monnens et al. 2020; Shimada et al. 2023). However, in

several cases, this putative gene has undergone such extensive diversification (Ross et al. 2016; Shimada et al. 2023) that its biological significance and functionality are questionable, and its annotation remains extremely challenging as a result. We propose that future experimental investigations be conducted to ascertain the presence of *atp8* in rhabdocoels, and turbellarians in general, and to elucidate any functional consequence associated with its high divergence (e.g., Eipel et al. 2011; Dautant et al. 2018). Overall, with the data available today, the absence of *atp8* does not appear to be a signature of endosymbiosis, as it has now been detected in several endosymbiotic rhabdocoels. Instead, the loss of *atp8* seems to be a synapomorphy of Neodermata, and it is not convergently linked with a symbiotic lifestyle.

Phylogenetic insights and host-symbiont dynamics in neodalyellid flatworms

The topology inferred in this work is largely consistent with previously published neodalyellid phylogenies based on ribosomal DNA (Van Steenkiste et al. 2013; Cavaleiro et al. 2017; Monnens et al. 2017; Cavaleiro et al. 2018; Stephenson et al. 2018; Hutson 2019). The monophyly of all three endosymbiotic lineages had been previously demonstrated, and these results are now substantiated using a more comprehensive dataset. We also find strong support for most neodalyellid genera, including *Baicallelia*, *Orostylis*, *Pogaina*, *Provortex*, *Pterastericola*, and *Vejdovskya*. For most other genera, only a single species could be included in this analysis, or the interrelationships between species were either not fully resolved or lacked sufficient statistical support. Several deep polytomies remain in the inferred topology, hindering our full understanding of neodalyellid evolution. Resolving these ambiguities will likely require an extensive nucleogenomic dataset, which we suggest as an avenue for future research.

Such a deep polytomy also exists at the base of Umagillidae, complicating speculation on the earliest divergent (extant) host lineage of this family. The current

analysis suggests that the hosts of the earliest diverging lineage may be either holothuroids or sipunculids. A distance-based cophylogenetic analysis shows significant congruence between host and symbiont phylogenies. Event-based cophylogenetic analyses predict that both cospeciation and lineage sorting events have contributed somewhat equally to umagillid evolutionary history. These analyses also predict a host switch within *Anoplodium* between distantly related holothuroids (from *Holothuria* to *Isostichopus*), and an unidentified umagillid closely related to the echinoid-infecting genus *Syndesmis*, which has come to infect *Isostichopus fuscus* (Ludwig, 1875). Molecular data on crinoid-inhabiting umagillids are lacking to date, and the inclusion of several enigmatic genera such as *Bicladus*, *Desmote*, and *Fallacohospes* is imperative to fully infer the speciation mechanisms leading to the current host range of umagillids.

Implications of a transition towards endosymbiosis

Neodalyellid mitogenomes suggest a potential link between endosymbiosis and elevated AT%, with more pronounced strand asymmetry

A non-significant trend suggesting increased AT% content in species with an endosymbiotic lifestyle is observed. Moreover, a significant difference in GC (but not AT) skew was inferred between lifestyles. Strand asymmetry has been well documented in free-living and parasitic flatworms alike (Le et al. 2004; Solà et al. 2015), and considerable variation is known between species (Le et al. 2004). Variations in nucleotide composition and skew can be attributed to mutation bias, selection mechanisms, or a combination of both factors (Long et al. 2018). Mutation bias may result from DNA replication and repair bias or, in the case of skew, the asymmetric nature of the mitochondrial replication process. Natural selection may favour specific nucleotides, leading to composition bias, or vary between leading and lagging strands, resulting in differing skew levels.

To date, nothing is known on replication or repair mechanisms in rhabdocoel mt genomes. Most mt sequences included in this work display a codon usage bias towards A or T in degenerative positions (results not shown), compliant with the hypothesis of mutation bias. Our results concur with those of a recent genome-wide study of parasitic platyhelminths, where mutational and selection mechanisms were inferred to be at interplay (Lamolle et al. 2019). Assessing whether the observed trend in AT increase in endosymbiotic rhabdocoels is mirrored in their nuclear genomes would allow for greater comparability with these results.

Several factors might contribute to selection pressure on nucleotide composition. The lower energetic cost of AT production compared to G/C may result in differing energy demands between free-living and endosymbiotic lineages, potentially favouring AT-rich genomes (Rocha and Danchin 2002). Environmental factors, such as a limited supply of nitrogen or other elements, could also impact GC content (Foerstner et al. 2005). Considering that some neodalyellids reportedly feed on host tissues (Jennings 1981; Shinn 1981; Jennings and Cannon 1985; Jennings 1988), the nucleotide composition of the host may influence that of their parasites (Dennis et al. 2020). Indeed, all assembled mt genomes of host specimens were found to be AT-rich (data not shown). Nevertheless, confirming this finding from a genome-wide dataset may be more informative. Voucher samples have been collected for all hosts included in this study, allowing for direct quantification of nucleotide contents in these specimens and correlation with the values obtained for their respective symbionts.

Gene order within Neodalyellida is marked by a multitude of genetic rearrangements

Demonstrating 24 unique gene orders in Neodalyellida, our findings support the notion that mt gene order is a highly variable trait in this taxon (Monnens et al. 2020), unlike the previously held view that gene order is highly conserved in platyhelminths

(Wey-Fabrizius et al. 2013). Genetic rearrangements are prevalent across all neodalyellid lineages for which mitogenomic data were obtained (**Fig. 5**). The diversity in gene order differences is attributed to a combination of transpositions, TDRLs, reversals, and reverse transpositions, occurring in both free-living and endosymbiotic lineages. Variations in gene orders were identified even at the intrageneric level, primarily due to rearrangements involving one or more tRNA genes. The existence of such divergence at the lowest taxonomic levels emphasizes the need for thorough species sampling when using mitochondrial gene order as a phylogenetic marker (Le et al. 2000).

No unambiguous signal of increased mt genomic rearrangements is apparent in endosymbiotic groups. Nevertheless, the potential patterns that may emerge with increased sampling and sequencing efforts, particularly in groups like graffillids and free-living neodalyellids, remain unknown.

Do varying branch lengths across neodalyellid taxa indicate a potential link with endosymbiosis?

In the absence of a dated topology for Rhabdocoela, branch lengths were used as a proxy to estimate and compare relative evolutionary rates between endosymbiotic and free-living neodalyellids. **Fig. 7** illustrates discernible variation in cumulative substitutions along branches in the mitochondrial topology. Deep branches leading to Graffillidae and Pterastericolidae all exhibit a comparatively high number of substitutions. Similarly, many substitutions have accumulated on branches leading to several endosymbiotic genera, including *Graffilla, Pterastericola, Syndesmis*, and *Umagilla*.

Intrageneric branches are generally short, with the exception of *Pterastericola:* both P. *astropectinis* and *P. pellucida* were acquired from *Astropecten irregularis* (Pennant, 1777) off the west coast of Sweden. This contrasts with the shorter branches

observed in its free-living sister genus, *Baicalellia*, whose representatives originate from opposite sides of the Atlantic. The terminal branch leading to *Syndesmis kurakaikina* is notably lengthy. However, given that this species was collected in New Zealand, whereas its congeners were sampled from European waters, this branch could be cut shorter with a more geographically balanced representation of the genus. Several longer branches toward free-living taxa from distant localities also occur in the tree, such as those leading to *Orostylis caecus* (Cuba) and *Tamanawas kalipis* (Canada). Notably, there are also relatively long branches in free-living species originating from proximate localities, where the influence of this bias can be expected to be minimal. This is exemplified in both inter- and intrageneric branches of Provorticidae.

The hints of elevated substitution rates among endosymbionts remain speculative, considering the still non-exhaustive taxonomic coverage. To address these limitations in future investigations, we recommend augmenting the current dataset by incorporating additional (free-living) graffillids and umagillids infecting sipunculids and/or crinoids. Simultaneously, efforts should be directed towards establishing a more geographically balanced dataset to alleviate potential biases associated with this approach.

Five mitochondrial PCGs in endosymbiotic neodalyellid taxa display significant relaxation of selection pressure

In endosymbiotic taxa, five mitochondrial PCGs exhibited a significant relaxation of selection pressure. This finding aligns with the observed selection pressure on three rhabdocoel mt genes in Monnens et al. (2020), but extends these findings to include additional PCGs. Relaxation of natural selection can occur through various mechanisms, such as a reduction in selection efficiency (e.g., greater influence of genetic drift) or the removal of functional constraints (e.g., adaptation to a new ecological niche). For animals transitioning to endosymbiosis and thereby shifting to

a new environment, a combination of both factors may contribute to the observed relaxation of selection pressure.

Mitochondrial PCGs universally encode enzyme subunits imperative for cell respiration (Pfanner et al. 2019). Our observations prompt speculation on whether the potential impairment in the functionality of these genes, resulting from relaxed natural selection, would impact the fitness of these animals. For example, aerobic cellular respiration may not be as vital as traditionally presumed. Early studies have demonstrated that umagillids and graffillids primarily accommodate a glycogen-based metabolism. This metabolic preference has been associated with environments characterised by fluctuating oxygen tension, high fecundity levels, and a consistent supply of nutrition from the host (Jennings and Mettrick 1968; Calow and Jennings 1974; Jennings and Calow 1975; Jennings 1980, 1981, 1988). While these animals may emphasise an anaerobic metabolism, similar to some neodermatans (Komuniecki and Tielens 2003; Müller et al. 2012), the presence of haemoglobin in all three endosymbiotic lineages suggests that oxygen availability remains crucial, making a shift towards a fully anaerobic metabolism unlikely (Cuénot 1892; Nicol 1960; Jennings and Phillips 1978; Jennings 1981; Jennings and Cannon 1985, 1987; Jennings 1988; Jennings and Hick 1990; Jennings 1997). Additionally, pterastericolids, with their reliance on lipids instead of glycogen (Jennings and Cannon 1985; Jennings 1988), contradict this notion, further undermining the idea of a complete transition to anaerobic metabolism.

The nuclear genome may also compensate mitochondrial gene products (Brandvain and Wade 2009), but this cannot be evaluated within the scope of the available data. In addition, evidence from nucleogenomic datasets highlights that parasitic metazoans can (partially) lose metabolic pathways, either because they are no longer necessary in the host-provided environment or can be compensated for by lateral gene transfer (Berriman et al. 2009; Zhou et al. 2009; Wang et al. 2011; Tsai et al. 2013; Zheng et al. 2013; Zarowiecki and Berriman 2014; Mauer et al. 2020). However, umagillids are able to survive outside their host for at least several days (e.g.,

884 experimental set-up in Monnens et al. 2019). This indicates that these animals cannot
885 be entirely dependent on their host for metabolites involved in cell respiration.
886
887 Finally, it is worth noting that with the current (non-exhaustive) dataset available, we
888 were compelled to designate all endosymbiotic neodalyellids as test branches and all
889 free-living neodalyellids as the reference set in RELAX tests. Availability of mt
890 genomes—or even complete genomes—for all established outgroups would allow
891 this comparative testing to be repeated for each endosymbiotic family. In combination
892 with a more comprehensive sampling of ingroups and outgroups, a more powerful
893 statistical analysis could be conducted, featuring three truly independent hypothesis
894 tests instead of a single, generalised one.

Conclusions

We expand on the molecular sequence database of a highly understudied metazoan lineage, Neodalyellida, with a specific focus on endosymbiotic lineages. Publicly available mitochondrial genomes are multiplied over tenfold and nuclear ribosomal data are more than doubled. Phylogenomic work from this combined nuclear-mitogenomic dataset yielded a robust topology for neodalyellids, demonstrating strong congruence between hosts and symbionts for the largest family of endosymbiotic rhabdocoels (Umagillidae), as well as two host switches within this group. Mitogenomic nucleotide diversity was assessed and promising regions were revealed for future work on these animals' diversity, phylogeny, and population genetics. A high variation in neodalyellid gene order was inferred, with rearrangements occurring as low as the intrageneric level. While substitution rates vary throughout the topology, a more geographically balanced dataset is required to thoroughly test the hypothesized correlation between the switch to endosymbiosis and elevated substitution rates. A trend toward an increased AT% in endosymbionts was observed and GC skew is positively correlated with endosymbiosis: these findings are hypothesised to originate from a combination of mutation bias and selection pressure. Finally, a relaxed selection pressure was found for five of twelve mitochondrial protein-coding genes in endosymbiotic neodalyellids, yet the biological significance of this demands further characterisation of the metabolism of these animals. Overall, this study has shed new light onto the molecular landscape of neodalyellid flatworms, offering new insights into their evolutionary history, allowing us to test several hypotheses in relation to the acquisition of a symbiotic lifestyle, and positioning these animals as an exciting model system for understanding the evolution and genomic signatures of endosymbiosis.

Acknowledgements

This research was supported by the Research Foundation Flanders (FWO, grant number 1141817N to M.M.), the Special Research Fund of Hasselt University (BOF20TT06 to M.P.M.V.), the Tula Foundation's Hakai Institute (grant to B.S.L.), the National Sciences and Engineering Research Council of Canada (NSERC 2019-03986 to B.S.L.), and the Flanders Marine Institute (VLIZ, BMRI grant 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. The research was carried out with infrastructure funded by the European Marine Biological Resource Centre (EMBRC Belgium - FWO project GOH3817N). This research was partially supported by national funds through FCT - Foundation for Science and Technology (Portugal) within the scope of UIDB/04423/2020 and UIDP/04423/2020. Sampling campaigns to Sweden and Italy were financed by travel grants for international mobility awarded to M.M. by the Doctoral Schools of Hasselt University. Collecting in Cuba was supported by BOF-Hasselt University (BOF15BL09 granted to Y.L.D.). Dr Ulf Jondelius and Dr Marco Curini-Galletti are thanked for hosting us during our respective sampling campaigns to Sweden and Sardinia. We are grateful to the staff of the marine station of the University of Plymouth, the Station Biologique de Roscoff, and the Sven Lovén Centre for Marine Sciences in Kristineberg and Tjärnö. Dredging in the Gullmar fjord (Kristineberg) was financially supported by Dr Andreas Hejnol. Dr Francisca Cavaleiro from Porto University (Portugal) is thanked for her help with sampling. Dr Christopher Laumer is kindly acknowledged for his input on our wet lab methodology and for providing the salting-out extraction protocol. Dr Geert Jan Bex is cordially thanked for his helpful advice on use of the VSC cluster. The authors are grateful to Ria Vanderspikken for administrative support and Natascha Steffanie for her invaluable help in the lab. We also wish to thank Dr Mare Geraerts, Dr Sofie Thijs, and Dr Nikol Kmentová for sharing their thoughts and suggestions on the molecular part of this study, Jan-Pieter Ploem for his help with early DNA extractions, Lisa Huijgen for her assistance during field expeditions to

951 Tjärnö and Sardinia, and Jeff Vandeweyer for his aid in figure editing and
952 troubleshooting bioinformatic work.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403-410.

Bachmann L, Fromm B, Patella de Azambuja L, Boeger WA. 2016. The mitochondrial genome of the egg-laying flatworm *Aglaiogyrodactylus forficulatus* (Platyhelminthes: Monogenoidea). Parasites & Vectors 9:285.

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin V, Nikolenko S, Pham S, Prjibelski A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computional Biology 19:455-477.

Baeza JA. 2015. Crustaceans as symbionts: an overview of their diversity, host use, and lifestyles. In: Thiel M, Watling L, editors. The natural history of the Crustacea. Life styles and feeding biology. New York, USA: Oxford University Press.

Bengtsson-Palme J, Veldre V, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, Bertrand Y, De Wit P, Sanchez M, et al. 2013. 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 in Ecology and Evolution 4:914-919.

Benson DA, Karsch-Mizrachi, Clark K, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. Nucleic Acids Research 40:D48-D53.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27:573-580.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch J, Pütz M, Middendorf M, Stadler PF. 2013. MITOS: improved *de novo* metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69:313-319.

979 Bernt M, Merkle D, Ramch K, Fritzch G, Perseke M, Bernhard D, Schlegel M, Stadler
980 P, Middendorf M. 2007. CREx: inferring genomic rearrangements based on common
981 intervals. Bioinformatics 23:2957-2958.

982 Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama
983 ST, Al-Lazikani B, Andrade LF, Ashton PD, et al. 2009. The genome of the blood fluke
984 *Schistosoma mansoni*. Nature 352:352-358.

985 Botero-Castro F, Tilak M-K, Justy F, Catzeflis F, Delsuc F, Douzery EJP. 2018. In cold
986 blood: compositional bias and positive selection drive the high evolutionary rate of
987 vampire bats mitochondrial genomes. Genome Biology and Evolution 10:2218-2239.

988 Brandvain Y, Wade MJ. 2009. The functional transfer of genes from the mitochondria
989 to the nucleus: The effects of selection, mutation, population size and rate of self-
990 fertilization. Genetics 182:1129-1139.

991 Calow P, Jennings JB. 1974. Calorific values in the phylum Platyhelminthes: the
992 relationship between potential energy, mode of life and the evolution of
993 entoparasitism. Biological Bulletin 147:81-94.

994 Castresana J. 2000. Selection of conserved blocks from multiple alignments for their
995 use in phylogenetic analysis. Molecular Biology and Evolution 17:540-552.

996 Castro LR, Austin AD, Dowton M. 2002. Contrasting rates of mitochondrial molecular
997 evolution in parasitic Diptera and Hymenoptera. Molecular Biology and Evolution
998 19:1100-1113.

999 Cavaleiro FI, Frade DG, Rangel LF, Santos MJ. 2017. *Syndesmis aethopharynx*
1000 Westervelt & Kozloff, 1990 (Rhabdocoela: Umagillidae): a revisitation supported by
1001 scanning electron microscopy and molecular analyses. Systematic Parasitology
1002 94:1007-1017.

1003 Cavaleiro FI, Frade DG, Rangel LF, Santos MJ. 2018. *Syndesmis* François, 1886
1004 (Rhabdocoela: Umagillidae): a revisitation, with a synopsis and an identification key

1005 to species, and new molecular evidence for ascertaining the phylogeny of the group.
1006 Systematic Parasitology 95:147-171.

1007 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ
1008 preprocessor. Bioinformatics 34:i884–i890.

1009 Chen W-J, Bu Y, Carapelli A, Dallai R, Li S, Yin W-Y, Luan Y-X. 2011. The
1010 mitochondrial genome of *Sinentomon erythranum* (Arthropoda: Hexapoda: Protura):
1011 An example of highly divergent evolution. BMC Evolutionary Biology 246.

1012 Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for
1013 phylogenomic inference from supermatrices. Systematic Biology 65:997-1008.

1014 Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals
1015 and additional sequence information. Computer Science and Biology: Proceedings of
1016 the German Conference on Bioinformatics (GCB) 99:45-56.

1017 Cock PJA, Grüning BA, Paszkiewicz K, Pritchard L. 2013. Galaxy tools and workflows
1018 for sequence analysis with applications in molecular plant pathology. PeerJ 1:e167.

1019 Colson P, Tamalet C, Raoult D. 2006. SVARAP and aSVARAP: simple tools for
1020 quantitative analysis of nucleotide and amino acid variability and primer selection for
1021 clinical microbiology. BMC Microbiology 6:21.

1022 Cuénot L. 1892. Commensaux et parasites des Echinodermes. Revue Biologique du
1023 nord de la France 1:1-23.

1024 Dautant A, Meier T, Hahn A, Tribouillard-Tanvier D, di Rago J-P, Kucharczyk R. 2018.
1025 ATP synthase diseases of mitochondrial genetic origin. Frontiers in Physiology 9:329.

1026 Dennis AB, Ballesteros GI, Robin S, Schrader L, Bast J, Berghöfer J, Beukeboom LW,
1027 Belghazi M, Bretaudeau A, Buellesbach J, et al. 2020. Functional insights from the
1028 GC-poor genomes of two aphid parasitoids, *Aphidius ervi* and *Lysiphlebus fabarum*.
1029 BMC Genomics 21:376.

1030 Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of
1031 organelle genomes from whole genome data. Nucleic Acids Research 45:e18.

1032 Doignon G, Artois T. 2006. Annotated checklist of the umagillid turbellarians infesting
1033 echinoids (Echinodermata). Belgian Journal of Zoology 136:101-106.

1034 Dowton M, Austin AD. 1999. Evolutionary dynamics of a mitochondrial rearrangment
1035 'hotspot' in the Hymenoptera. Molecular Biology and Evolution 16:298-309.

1036 Dowton M, Austin AD. 1995. Increased genetic diversity in mitochondrial genes is
1037 correlated with the evolution of parasitism in the Hymenoptera. Journal of Molecular
1038 Evolution 41:958-965.

1039 Edgar RC. 2004a. MUSCLE: a multiple sequence alignment method with reduced time
1040 and space complexity. BMC Bioinformatics 5:113.

1041 Edgar RC. 2004b. MUSCLE: multiple sequence alignment with high accuracy and
1042 high throughput. Nucleic Acids Research 32:1792-1797.

1043 Egger B, Bachmann L, Fromm B. 2017. *Atp8* is in the ground pattern of flatworm
1044 mitochondrial genomes. BMC Genomics 18:1-10.

1045 Eipel C, Hildebrandt A, Scholz B, Schyschka L, Minor T, Kreikemeyer B, Ibrahim SM,
1046 Vollmar B. 2011. Mutation of mitochondrial ATP8 gene improves hepatic energy
1047 status in a murine model of acute endotoxemic liver failure. Life Sciences 88:343-349.

1048 Foerstner KU, von Mering C, Hooper SD, Bork P. 2005. Environments shape the
1049 nucleotide composition of genomes. EMBO Reports 6:1208-1213.

1050 Gillard GB, Garama DJ, Brown CM. 2014. The transcriptome of the NZ endemic sea
1051 urchin Kina (*Evechinus chloroticus*). BMC Genomics 15:45.

1052 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Fascuel O. 2010. New
1053 algorithms and methods to estimate maximum-likelihood phylogenies: assessing the
1054 performance of PhyML 3.0 Systematic Biology 59:307-321.

1055 Hadfield JD. 2010. MCMC methods for multi-response generalized linear mixed
1056 models: the MCMCglmm R package. Journal of Statistical Software 33:1–22.

1057 Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA.
1058 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites.
1059 Science 265:1087-1090.

1060 Hardman M, Hardman LM. 2006. Comparison of the phylogenetic performance of
1061 neodermatan mitochondrial protein-coding genes. Zoologica Scripta 35:655-665.

1062 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2017. UFBoot2:
1063 Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution
1064 35:518-522.

1065 Houben A. 2013. Diversity and phylogeny of the limnoterrestrial Rhabdocoela
1066 (Platyhelminthes), an enigmatic group of minute metazoans. PhD thesis.

1067 Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic
1068 trees. Bioinformatics 17:754-755.

1069 Hutson DC. 2019. *Collastoma esotericum* (Neodalyellida: Umagillidae), a new species
1070 of sipunculan-inhabiting rhabdocoel from Queensland, Australia. Zootaxa 4701:563-
1071 573.

1072 Jackson AP. 2014. Preface: the evolution of parasite genomes and the origins of
1073 parasitism. Parasitology 142:S1-S5.

1074 Jennings JB. 1988. Nutrition and respiration in symbiotic Turbellaria. Fortschritte der
1075 Zoologie/Progress in Zoology 36:3-13.

1076 Jennings JB. 1980. Nutrition in symbiotic Turbellaria. In: Smith DC, Tiffon Y, editors.
1077 Nutrition in the Lower Metazoa. Oxford, United Kingdom: Pergamon Press Ltd.

1078 Jennings JB. 1997. Nutritional and respiratory pathways to parasitism exemplified in
1079 the Turbellaria. International Journal for Parasitology 27:679-691.

1080 Jennings JB. 1971. Parasitism and commensalism in the Turbellaria. Advances in
1081 Parasitology 9:1-32.

1082 Jennings JB. 1981. Physiological adaptations to entosymbiosis in three species of
1083 graffillid rhabdocoels. Hydrobiologia 84:147-153.

1084 Jennings JB, Calow P. 1975. The relationship between high fecundity and the
1085 evolution of endoparasitism. Oecologia 21:109-115.

1086 Jennings JB, Cannon LRG. 1985. Observations on the occurence, nutritional
1087 physiology and respiratory pigment of three species of flatworms (Rhabdocoela:
1088 Pterastericolidae) entosymbymbiotic in starfish from temperate and tropical waters.
1089 Ophelia 24:199-215.

1090 Jennings JB, Cannon LRG. 1987. The occurence, spectral properties and probable
1091 role of haemoglobins in four species of entosymbiotic turbellarians (Rhabdocoela:
1092 Umagillidae). Ophelia 27:143-154.

1093 Jennings JB, Hick AJ. 1990. Differences in the distribution mitochondrial content and
1094 probable roles of haemoglobin-containing of entosymbiotic turbellarians
1095 (Rhabdocoela: umagillidae and pterastericolidae). Ophellia 31:163-175.

1096 Jennings JB, Mettrick DF. 1968. Observations on the ecology, morphology and
1097 nutrition of the rhabdocoel turbellarian *Syndesmis franciscana* (Lehman, 1946) in
1098 Jamaica. Caribbean Journal of Science 8:57-69.

1099 Jennings JB, Phillips JI. 1978. Feeding and digestion in three entosymbiotic graffillid
1100 rhabdocoels from bivalve and gastropod molluscs. Biological Bulletin 155:542-562.

1101 Justine J-L, Leblanc P, Keller F, Lester RJG. 2009. Turbellarian black spot disease in
1102 bluespine unicornfish *Naso unicornis* in New Caledonia, caused by the parasitic
1103 turbellarian *Piscinquilinus* sp. Disease of Aquatic Organisms 85:245-249.

1104 Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
1105 ModelFinder: fast model selection for accurate phylogenetic estimates. Nature
1106 Methods 14:587-589.

1107 Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence
1108 alignment, interactive sequence choice and visualization. Briefings in Bioinformatics
1109 20:1160–1166.

1110 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
1111 7: improvements in performance and usability. Molecular Biology and Evolution
1112 30:772-780.

1113 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S,
1114 Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and
1115 extendable dekstop software platform for the organization and analysis of sequence
1116 data. Bioinformatics 28:1647-1649.

1117 Kenny NJ, Noreña C, Damborenea C, Grande C. 2019. Probing recalcitrant problems
1118 in polyclad evolution and systematics with novel mitochondrial genome resources.
1119 Genomics 111:343-355.

1120 Khamis A, Colson P, Raoult D, Scola BL. 2003. Usefulness of *rpoB* gene sequencing
1121 for identification of *Afipia* and *Bosea* species, including a strategy for choosing
1122 discriminative partial sequences. Applied and Environmental Microbiology 69:6740–
1123 6749.

1124 Koblmüller S, Sturmbauer C, Verheyen E, Meyer A, Salzburger W. 2006. Mitochondrial
1125 phylogeny and phylogeography of East African squeaker catfishes (Siluriformes:
1126 *Synodontis*). BMC Evolutionary Biology 6:1471-2148.

1127 Komuniecki R, Tielens AGM. 2003. Chapter 14 - Carbohydrate and energy
1128 metabolism in parasitic helminths. In: Marr JJ, Nilsen TW, Komuniecki RW, editors.
1129 Molecular Medical Parasitology. London: Academic Press. p. 339-358.

1130 Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007.
1131 RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids
1132 Research 35:3100-3108.

1133 Lamolle G, Fontenla S, Rijo G, Tort JF, Smircich P. 2019. Compositional analysis of
1134 flatworm genomes shows strong codon usage biases across all classes. Frontiers in
1135 Genetics 10.

1136 Laslett D, Canbäck B. 2008. ARWEN, a program to detect tRNA genes in metazoan
1137 mitochondrial nucleotide sequences. Bioinformatics 24:172-175.

1138 Le TH, Blair D, Agatsuma T, Humair P-F, Campbell NJH, Iwagami M, Littlewood DTJ,
1139 Peacock B, Johnston DA, Bartley J, et al. 2000. Phylogenies inferred from
1140 mitochondrial gene orders - A cautionary tale from the parasitic flatworms. Molecular
1141 Biology and Evolution 17:1123-1125.

1142 Le TH, McManus DP, Blair D. 2004. Codon usage and bias in mitochondrial genomes
1143 of parasitic platyhelminthes. The Korean Journal of Parasitology 42:159-167.

1144 Leigh JW, Susko E, Baumgartner M, Roger AJ. 2008. Assessing congruence in
1145 phylogenomic data. Systematic Biology 57:104-115.

1146 Letunic I, Bork P. 2017. 20 years of the SMART protein domain annotation resource.
1147 Nucleic Acids Research 46:D493–D496.

1148 Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for
1149 phylogenetic tree display and annotation. Nucleic Acids Research 49:W293-W296.

1150 Letunic I, Doerks T, Bork P. 2015. SMART: recent updates, new developments and
1151 status in 2015. Nucleic Acids Research 43:D257-D260.

1152 Li X-y, Yan L-p, Pape T, Gao Y-y, Zhang D. 2020. Evolutionary insights into bot flies
1153 (Insecta: Diptera: Oestridae) from comparative analysis of the mitochondrial genomes.
1154 International Journal of Biological Macromolecules 149:371-380.

Long H, Sung W, Kucukyildirim S, Williams E, Miller SF, Guo W, Patterson C, Gregory C, Strauss C, Stone C, et al. 2018. Evolutionary determinants of genome-wide nucleotide composition. Nature Ecology & Evolution 2:237-240.

Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. Algorithms for Molecular Biology 6:26.

Lowe TM, Chen Pp. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research 44:W54-W57.

Mauer K, Hellmann SL, Groth M, Fröbius AC, Zischler H, Hankeln T, Herlyn H. 2020. The genome, transcriptome, and proteome of the fish parasite *Pomphorhynchus laevis* (Acanthocephala). PLOS One 15:e0232973.

McManus DP, Le TH, Blair D. 2004. Genomics of parasitic flatworms. International Journal for Parasitology 34:153-158.

Medvedev SG. 2017. Adaptations of fleas (Siphonaptera) to parasitism. Entomological Review 97:1023-1030.

Merkle M, Middendorf M, Wieseke N. 2010. A parameter-adaptive dynamic programming approach for inferring cophylogenies. BMC Bioinformatics 11:S60.

Mikheev VN, Pasternak AF, Valtonen ET. 2015. Behavioural adaptations of argulid parasites (Crustacea: Branchiura) to major challenges in their life cycle. Parasites & Vectors 8:10.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37:1530-1534.

Monnens M, Artois T, Vanhove MPM. 2017. *Syndesmis aethopharynx* (Umagillidae, Rhabdocoela, Platyhelminthes) from the sea urchin *Paracentrotus lividus*: first record from the Eastern Mediterranean, phylogenetic position and intraspecific morphological variation. Parasitology International 66:848-858.

Monnens M, Frost EJ, Clark M, Sewell MA, Vanhove MPM, Artois T. 2019. 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. International Journal for Parasitology: Parasites and Wildlife 10:71-82.

Monnens M, Thijs S, Briscoe AG, Clark M, Frost EJ, Littlewood DTJ, Sewell M, Smeets K, Artois T, Vanhove MPM. 2020. The first mitochondrial genomes of endosymbiotic rhabdocoels illustrate evolutionary relaxation of *atp*8 and genome plasticity in flatworms. International Journal of Biological Macromolecules 162:454-469.

Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu RY, van der Giezen M, Tielens AG, Martin WF. 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiology and Molecular Biology Reviews 76:444-495.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32:268-274.

Nguyen THT. 2020. Adaptation of anemonefish to their host anemones: From Genetics to Physiology. [[Bergen, Norway]: University of Bergen.

Nicol JAC. 1960. The Biology of Marine Animals. London, United Kingdom: Pitman.

Noe L, Kucherov G. 2005. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Research 33:W540-W543.

Okonechikov K, Golosova O, Fursov M, team tU. 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166–1167.

Perna NT, Kocher TD. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution 41:353-358.

1207 Pfanner N, Warscheid B, Wiedemann N. 2019. Mitochondrial proteins: from
1208 biogenesis to functional networks. Nature Reviews Molecular Cell Biology 20:267-
1209 284.

1210 Poulin R. 2011. Chapter 1 - The many roads to parasitism: a tale of convergence.
1211 Advances in parasitology 74: 1–40.

1212 Poulin R, Randhawa HS. 2015. Evolution of parasitism along convergent lines: from
1213 ecology to genomics. Parasitology 142:S6-S15.

1214 R: A language and environment for statistical computing [Internet]. Vienna, Austria: R
1215 Foundation for Statistical Computing; 2022. Available from: http://www.R-project.org/

1216 Rambaut A. 2006–2021. FigTree: Tree figure drawing tool. Version 1.4.4.

1217 Rocha EP, Danchin A. 2002. Base composition bias might result from competition for
1218 metabolic resources. TRENDS in Genetics 18:291-294.

1219 Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference
1220 under mixed models. Bioinformatics 19:1572-1574.

1221 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu
1222 L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2 Efficient Bayesian phylogenetic
1223 inference and model choice across a large model space. Systematic Biology 61:539-
1224 542.

1225 Rosa MT, Oliveira DS, Loreto ELS. 2017. Characterization of the first mitochondrial
1226 genome of a catenulid flatworm: *Stenostomum leucops* (Platyhelminthes). Journal of
1227 Zoological Systematics and Evolutionary Research 55:98-105.

1228 Ross E, Blair D, Guerrero-Hernández, Alvarado AS. 2016. Comparative and
1229 transcriptome analyses uncover key aspects of coding- and long noncoding RNAs in
1230 flatworm mitochondrial genomes. G3 (Bethesda) 6:1191-1200.

RStudio: Integrated Development for R [Internet]. Boston, Massachussets, USA: RStudio, Inc.; 2022. Available from: http://www.rstudio.com/

Schockaert ER. 1996. Turbellarians: the importance of turbellarians in ecosystems. In: Hall G, editor. Methods for the examination of organismal diversity in soils and sediments. Wallingford: CAB International in association with UNESCO and the IUBS. p. 211–225.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proceedings of the National Academy of Sciences of the United States of America 95:5857-5864.

Shao R, Campbell NJH, Schmidt ER, Barker SC. 2001. Increased rate of gene rearrangment in the mitochondrial genomes of three orders of hemipteroid insects. Molecular Biology and Evolution 18:1828-1832.

Shimada D, Hiruta SF, Takahoshi K, Kajihara H. 2023. Does *atp8* exist in the mitochondrial genome of Proseriata (Metazoa: Platyhelminthes)? Animal Gene 30:200161.

Shinn GL. 1981. The diet of three species of umagillid neorhabdocoel turbellarians inhabiting the intestine of echinoids. Hydrobiologia 84:155-162.

Silvestre D, Arias MC. 2006. Mitochondrial tRNA gene translocations in highly eusocial bees. Genetics and Molecular Biology 29:572-575.

Solà E, Álvarez-Presas M, Frías-López C, Littlewood DTJ, Rozas J, Riutort M. 2015. Evolutionary analysis of mitogenomes from parasitic and free-living flatworms. PLOS One 10:e0120081.

Sorenson MD, Payne RB. 2001. A single ancient origin of brood parasitism in African finches: implications for host-parasite coevolution. Evolution 55:2550-2567.

1255 Stephenson I, Van Steenkiste NWL, Leander BS. 2018. Molecular phylogeny of
1256 neodalyellid flatworms (Rhabdocoela), including three new species from British
1257 Columbia. Journal of Zoological Systematics and Evolutionary Research 57:41-56.

1258 Stöver BC, Müller KF. 2010. TreeGraph 2: Combining and visualizing evidence from
1259 different phylogenetic analyses. BMC Bioinformatics 11.

1260 Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent
1261 and ambiguously aligned blocks from protein sequence alignments. Systematic
1262 Biology 56:564-577.

1263 Tessens B, Monnens M, Backeljau T, Jordaens K, Van Steenkiste N, Breman FC,
1264 Smeets K, Artois T. 2021. Is 'everything everywhere'? Unprecedented cryptic diversity
1265 in the cosmopolitan flatworm *Gyratrix hermaphroditus*. Zoologica Scripta:1-15.

1266 Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast
1267 online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research
1268 44:W232-235.

1269 Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL,
1270 Tracey A, Bobes RJ, Fragoso G, Sciutto E, et al. 2013. The genomes of four tapeworm
1271 species reveal adaptations to parasitism. Nature 496:57-63.

1272 Van Steenkiste N, Tessens B, Willems W, Backeljau T, Jondelius U, Artois T. 2013. A
1273 comprehensive molecular phylogeny of Dalytyphloplanida (Platyhelminthes:
1274 Rhabdocoela) reveals multiple escapes from the marine environment and origins of
1275 symbiotic relationships. PLOS One 8:1-13.

1276 Vanhove MPM, Briscoe AG, Jorissen MWP, Littlewood DTJ, Huyse T. 2018. The first
1277 next-generation sequencing approach to the mitochondrial phylogeny of African
1278 monogenean parasites (Platyhelminthes: Gyrodactylidae and Dactylogyridae). BMC
1279 Genomics 19.

1280 Vanhove MPM, Tessens B, Schoelinck C, Jondelius U, Littlewood DTJ, Artois T,
1281 Huyse T. 2013. Problematic barcoding in flatworms: A case-study on monogeneans
1282 and rhabdocoels (Platyhelminthes). Zookeys 265:355-379.

1283 Wang X, Chen W, Huang Y, Sun J, Men J, Liu H, Luo F, Guo L, Lv X, Deng C, et al.
1284 2011. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*.
1285 Genome Biology 12:R107.

1286 Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX:
1287 detecting relaxed selection in a phylogenetic framework. Molecular Biology and
1288 Evolution 32:820-832.

1289 Wey-Fabrizius AR, Podsiadlowski L, Herlyn H, Hankeln T. 2013. Platyzoan
1290 mitochondrial genomes. Molecular Phylogenetics and Evolution 69:365-375.

1291 Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de*
1292 *novo* genome assemblies. Bioinformatics 31:3350-3352.

1293 Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution.
1294 International Review of Cytology 141:173-216.

1295 WoRMS. 2024. Rhabdocoela. Accessed 2022-03-07. Available from:
1296 www.marinespecies.org/aphia.php?p=taxdetails&id=16236.

1297 Xiao J-H, Jia J-G, Murphy RW, Huang D-W. 2011. Rapid evolution of the
1298 mitochondrial genome in chalcidoid wasps (Hymenoptera: Chalcidoidea) driven by
1299 parasitic lifestyles. PLOS One 6:e26645.

1300 Ye F, King SD, Cone DK, You P. 2014. The mitochondrial genome of *Paragyrodactylus*
1301 *variegatus* (Platyhelminthes: Monogenea): differences in major non-coding region and
1302 gene order compared to *Gyrodactylus*. Parasites & Vectors 7:377.

1303 Zarowiecki M, Berriman M. 2014. What helminth genomes have taught us about
1304 parasite evolution. Parasitology 142:S85-S97.

Zhang B, Havird JC, Wang E, Lv J, Xu X. 2021. Massive gene rearrangement in mitogenomes of phytoseiid mites. International Journal of Biological Macromolecules 186:33-39.

Zhang D. 2019. Mitochondrial genomes of two *Thaparocleidus* species (Platyhelminthes: Monogenea) reveal the first rRNA gene rearrangement among the Neodermata. International Journal of Molecular Sciences 20:4214.

Zhang D, Zou H, Wu SG, Li M, Jakovlić I, Zhang J, Chen R, Li WX, Wang GT. 2018. Three new Diplozoidae mitogenomes expose unusual compositional biases within the Monogenea class: implications for phylogenetic studies. BMC Evolutionary Biology 18:133-133.

Zheng H, Zhang W, Zhang L, Zhang Z, Li J, Lu G, Zhu Y, Wang Y, Huang Y, Liu J, et al. 2013. The genome of the hydatid tapeworm *Echinococcus granulosus*. Nature Genetics 45:1168-1175.

Zhou Y, Zheng H, Chen Y, Zhang L, Wang K, Guo J, Huang Z, Zhang B, Huang W, Jin K, et al. 2009. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. Nature 460:345-351.

Data Accessibility and Benefit-Sharing

Newly obtained raw sequence reads are deposited in the Sequence Read Archive on NCBI (BioProject numbers PRJNA606139 and PRJNA692190), along with all sample metadata. All annotated assemblies are made publicly available on GenBank. Related metadata, including localities, time of collection, geocoordinates, and, where applicable, symbionts' host and position in the host, are provided in the supplementary tables of this manuscript. DNA extracts of symbionts and hosts are stored in the collections of Hasselt University and linked to all existing photographic material, documentation, and, when applicable, host vouchers of the extracted specimens.

This research involved a large international sampling effort across several countries in Europe and the Americas. The benefits generated by this work are non-monetary in nature and concern the dissemination of research findings. The sampling conducted for this study was partially carried out in collaboration with local marine stations (Sven Lovén Centre for Marine Sciences in Kristineberg and Tjärnö, Station Biologique de Roscoff), which were reimbursed through bench fees and did not require additional compensation or benefit-sharing. In other countries (Canada, Cuba, Italy, Portugal), collection of resources was supported by newly established and existing scientific collaborations with local partners, with co-authorships offered in exchange for their efforts. A similar approach was taken for sampling in the United Kingdom, yet it is worth noting that, although the UK is a Party to the Nagoya Protocol, it has chosen not to introduce legislation regarding access to genetic resources. Sampling in Finland and the Netherlands was conducted before these countries ratified the Nagoya Protocol.

Author Contributions

M. Monnens: Conceptualization, Formal analysis, Investigation, Methodology, Writing - Original Draft, Visualization, Funding acquisition

T. Artois: Conceptualization, Methodology, Supervision, Funding acquisition, Project administration

A. Briscoe: Investigation, Writing - Review & Editing

Y. L. Diez: Resources, Writing - Review & Editing

K. P. P. Fraser: Resources, Writing - Review & Editing

B. S. Leander: Resources, Writing - Review & Editing

D. T. J. Littlewood: Investigation, Writing - Review & Editing

M. J. Santos: Resources, Writing - Review & Editing

K. Smeets: Writing - Review & Editing

N. W. L. Van Steenkiste: Resources, Writing - Review & Editing

M. P. M. Vanhove: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition