

RESEARCH ARTICLE

Comparative Mitogenomics of Leeches (Annelida: Clitellata): Genome Conservation and *Placobdella*-Specific *trnD* Gene Duplication

Alejandro Ocegüera-Figueroa^{1,2*}, Alejandro Manzano-Marín³, Sebastian Kvist^{4,5}, Andrés Moya^{3,6}, Mark E. Siddall⁷, Amparo Latorre^{3,6}



1 Laboratorio de Helmintología, Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, Coyoacán, 04510, Mexico City, Mexico, **2** Research Collaborator, Department of Invertebrate Zoology, Smithsonian Institution, National Museum of Natural History, Washington D. C., United States of America, **3** Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Catedrático José Beltrán 2, 46008, Paterna, Valencia, Spain, **4** Department of Natural History, Royal Ontario Museum, 100 Queen's Park, Toronto, ON, M5S 2C6, Canada, **5** Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks Street, Toronto, ON, M5S 3B2, Canada, **6** Área de Genómica y Salud de la Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana (FISABIO), Avenida de Catalunya 21, 46020, Valencia, Spain, **7** Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West at 79th Street, New York, NY, 10024, United States of America

OPEN ACCESS

Citation: Ocegüera-Figueroa A, Manzano-Marín A, Kvist S, Moya A, Siddall ME, Latorre A (2016) Comparative Mitogenomics of Leeches (Annelida: Clitellata): Genome Conservation and *Placobdella*-Specific *trnD* Gene Duplication. PLoS ONE 11(5): e0155441. doi:10.1371/journal.pone.0155441

Editor: Marc Robinson-Rechavi, University of Lausanne, SWITZERLAND

Received: October 29, 2015

Accepted: April 28, 2016

Published: May 13, 2016

Copyright: © 2016 Ocegüera-Figueroa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. All 3 files are available with the newly generated mitochondrial genomes in public repositories (European Nucleotide Archive): *Haementeria officinalis* (ENA accn. number: LT159848), *Placobdella lamothei* (ENA accn. number: LT159849) and *Placobdella parasitica* (ENA accn. number: LT159850).

Funding: This work was supported by Consejo Nacional de Ciencia y Tecnología - Mexico, Posdoctoral grant (<http://www.conacyt.gob.mx/>) and

These authors contributed equally to this work.
* aocegüera@ib.unam.mx

Abstract

Mitochondrial DNA sequences, often in combination with nuclear markers and morphological data, are frequently used to unravel the phylogenetic relationships, population dynamics and biogeographic histories of a plethora of organisms. The information provided by examining complete mitochondrial genomes also enables investigation of other evolutionary events such as gene rearrangements, gene duplication and gene loss. Despite efforts to generate information to represent most of the currently recognized groups, some taxa are underrepresented in mitochondrial genomic databases. One such group is leeches (Annelida: Hirudinea: Clitellata). Herein, we expand our knowledge concerning leech mitochondrial makeup including gene arrangement, gene duplication and the evolution of mitochondrial genomes by adding newly sequenced mitochondrial genomes for three bloodfeeding species: *Haementeria officinalis*, *Placobdella lamothei* and *Placobdella parasitica*. With the inclusion of three new mitochondrial genomes of leeches, a better understanding of evolution for this organelle within the group is emerging. We found that gene order and genomic arrangement in the three new mitochondrial genomes is identical to previously sequenced members of Clitellata. Interestingly, within *Placobdella*, we recovered a genus-specific duplication of the *trnD* gene located between *cox2* and *atp8*. We performed phylogenetic analyses using 12 protein-coding genes and expanded our taxon sampling by including GenBank sequences for 39 taxa; the analyses confirm the monophyletic status of

Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica - Universidad Nacional Autónoma de México Project IA204114 (<http://dgapa.unam.mx/html/papiit/papit.html>) to AOF; European Commission (Marie Curie FP7-PEOPLE-2010-ITN SYMBIOMICS 264774) Ph. D. grant (<https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/fp7/calls/fp7-people-2013-ief.html>) and Consejo Nacional de Ciencia y Tecnología – Mexico, Ph. D. grant CONACYT 327211/381508 (<http://www.conacyt.gob.mx/>) to AMM; Royal Ontario Museum Schad Gallery of Biodiversity Research Fund, Olle Engkvist Byggmästare Foundation and NSERC Discovery Grant (RGPIN-2016-06125) (<http://www.nserc-crsng.gc.ca>) to SK; Grants BFU2012-39816-C02-01 (co-financed by FEDER funds and Ministerio de Economía y Competitividad, Spain) (www.mineco.gob.es/) to AL; and Grant PrometeoII/2014/065 (Conselleria d'Educació, Generalitat Valenciana, Spain) (<http://www.cece.gva.es/es/>) to AM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Clitellata, yet disagree in several respects with other phylogenetic hypotheses based on morphology and analyses of non-mitochondrial data.

Introduction

Mitochondrial DNA sequences, alone or in combination with nuclear markers and morphological data, have been widely used to investigate the phylogenetic relationships of an extensive array of organisms (e.g. [1–5]), as well as in efforts to estimate biological diversity and identify specimens through the Barcode of Life initiative [6], which relies on a short fragment of the mitochondrial genome (*cox1*) in order to identify unknown specimens. With increased availability of complete mitochondrial genomes, in part due to the advancement of high-throughput sequencing, comparative studies targeting the differences in rates of evolution between genes, as well as gene rearrangements, gene duplication and gene loss are progressively feasible [7–9]. Whereas the timing and mode of these evolutionary events are quite well understood in mammals [10–12] and other model organisms [13,14], our knowledge concerning the mitochondrial genomic makeup of most invertebrate taxa is still limited. One such group is Annelida, a rather large phylum with more than 17,000 currently recognized species [15]. Once considered to be an easily diagnosable group given the serial homology of the segmented body plan (see [16]), modern phylogenetic hypotheses for Annelida have revealed that the phylum includes several groups without noticeable segmentation, such as Siboglonidae (= Pogonophora), Urechidae (Echiura) and Sipunculidae (Sipuncula) [17–23]. Whereas some other groups of organisms, such as arthropods and molluscs, display a relatively high number of rearrangements within mitochondrial genomes, gene order within Annelida has been hypothesized to be relatively well conserved [24–26]. Somewhat contrary to this hypothesis, recent investigations into the mitogenomic makeup of several annelids have revealed substantial disparity between taxa, especially when including early diverging lineages (see e.g. [27–30]). As a result of the bulk of taxonomic diversity within Annelida being contained within the paraphyletic class Polychaeta, most mitogenomic studies concern polychaete taxa. That is to say that there is notable paucity of comparative data for the remaining class Clitellata (including hirudineans, branchiobdellids, acanthobdellids and oligochaetes), which often forms a well-supported monophyletic group in contemporary phylogenetic analyses (but see [17]); in addition, representatives possess a clear morphological synapomorphy, the clitellum. This lack of data is most conspicuous when regarding the sequenced mitogenomes of Hirudinea (including more than 680 species, which exhibit a variety of morphological traits, feeding preferences and ecological functions [31,32]). In fact, only two complete (*Whitmania pigra* [Whitman 1886] and *Whitmania laevis* [Baird, 1869]) and one partial mitochondrial genome (*Helobdella robusta* [Shankland et al. 1992]) have so far been generated [33,34]. Note here that leech mitochondrial genomes available in public databases under the taxonomic labels *Hirudo nipponia* Whitman 1886 (GenBank accn. number: NC_023776 and KC667144), *Erpobdella octoculata* (Linnaeus 1758) (GenBank KC688270), *Whitmania acranulata* (Whitman 1886) (GenBank KC688271) and *Hirudinaria manillensis* Lesson 1842 (GenBank KC688268.1) were excluded from the present study. This is because a comparison between *cox1* sequences from these complete genomes and isolated *cox1* sequences from the same taxa (data not shown), confirm the notion that these mitogenomes are either derived from misidentified taxa or are a result of sequencing contaminations, as has previously been pointed out [35]. Most likely, the aforementioned mitogenomes belong to an unidentified *Whitmania* species. Additionally, after further investigation of these genomes, we concluded that they all share the same gene order as the *W. pigra* mitogenome, such that they would add no further data regarding (e.g.) genome rearrangements.

In the present study, we provide a detailed comparative investigation into the composition of mitochondrial genomes for three bloodfeeding glossiphoniid leeches: *Placobdella lamothei* Ocegüera-Figueroa and Siddall 2008, *Placobdella parasitica* (Say 1824) and *Haementeria officinalis* De Filippi 1849. In addition, we place our results in a phylogenetic context using the 12 protein coding genes from 42 complete or nearly complete annelid mitochondrial genomes, using other representatives of Lophotrochozoa (Bryozoa and Brachiopoda) as outgroups.

Materials and Methods

Collection of specimens, sequencing, assembly and annotation of mitochondrial genomes

Specimens of *Haementeria officinalis* and *Placobdella lamothei* were collected in the Mexican states of Guanajuato and Estado de Mexico, respectively, under the collection permit SEMARNAT 12099/14 to AO-F. The two leech species are not listed as endangered or protected by Mexican government and are not CITES listed. Studying these species does not require approval of ethics committee. Voucher specimens were deposited in the Colección Nacional de Helmintos, at the Instituto de Biología, Universidad Nacional Autónoma de México (CNHE Catalogue numbers 8621, 5679).

Reads from the mitochondrial genomes of *H. officinalis* (454 flx+) and *P. lamothei* (454 flx+ and HiSeq2000) were generated as a by-product of a sequencing effort focused on the non-cultivable bacterial endosymbionts of these organisms [36]. In short, 454 reads were filtered using PyroCleaner v1.3 [37] and the remaining reads were taxonomically assigned using PhymmBL v4.0 [38] with custom-added genomes of representatives from the class Insecta (*Atta cephalotes*, *Acyrtosiphon pisum*, *Drosophila melanogaster* and *Tribolium castaneum*), *Homo sapiens* GRCh37.p5, and the leech *Helobdella robusta*, in addition to their corresponding mitochondrial genomes. For *P. parasitica*, 454 flx+ titanium reads were recovered from the Short Read Archive (SRA) from NCBI (accession SRX101489) (see details in [39]). These data resulted from a previous study aimed at characterizing an endosymbiotic leech bacteria. The 454 reads were cleaned using PyroCleaner in order to remove duplicates, reads shorter than 100 bp, reads longer than 1000 bp and reads that had quality PHRED scores lower than 35. HiSeq2000 reads were cleaned from artifacts, right-tail clipped and scrutinized for length (minimum length of 50 bp) using the FASTX-toolkit v0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). The standalone version of PRINSEQ v0.20.3 [40] was used to remove reads containing undefined nucleotides and perform de-duplication. For the reconstruction of *H. officinalis* (Genbank:JN850907) and *P. parasitica* (Genbank:AF003261) mitochondrial genomes, partial *cox1* sequences were retrieved from the nucleotide database from NCBI. For *P. lamothei*, a partial *cox1* sequence from a previous study [41] was used. A process of iterative assembly (mapping and extension of contigs) starting with the partial *cox1* nucleotide sequences using MIRA v3.4.1 [42] was performed for each of the genomes, missing only part of the long non-coding regions. For *P. lamothei*, HiSeq2000 reads were used to perform gap-filling with GapFiller v1.9 [43] and correction of nucleotides using Polisher v2.0.8 (available for academic use from the JGI <http://www.jgi.doe.gov/software/>), although no errors were detected. Annotations were performed using MITOS [44] and were originally used to identify tRNAs, rRNA regions and putative coding gene regions. From there, we identified putative start codons on the basis of them immediately following the 3' ends of other coding sequences, tRNAs or rRNAs, with subsequent comparison against other Clitellata genomes. Stop codons were determined as follows: 1) canonical stop codons, which did not overlap with tRNAs or 2) partial stop codons, where tRNA-adjacent T or TA bases have been transcriptionally modified to add alanine residues to the 3' of the mRNA to complete the stop codon. If none of the above conditions were met, the

gene was annotated as overlapping the adjacent feature in its 3' end. All annotated sequences are deposited in the International Nucleotide Sequence Database Collaboration [45] (INSDC; <http://www.insdc.org>) under the previously mentioned accession numbers.

Mitochondrial gene order

A set of 34 previously sequenced mitochondrial genomes of annelids, with almost complete protein-coding gene sequences, was retrieved from the NCBI nucleotide database (keywords: "annelida" and "mitochondrion"). Note that the genome of *Whitmania laevis*, which was recently published [35] has been shown to match the gene order of *Whitmania pigra* (see above) and, therefore, this taxon was not included in the present study. Visual display of genes was performed in R [46] using the package genoPlotR v0.8.2 [47] and edited in Inkscape v0.91 (<https://inkscape.org/en/>).

Phylogenetic analysis

The previously described set of annelid mitochondrial genomes, plus an additional 6 complete mitochondrial genomes from representatives of Bryozoa and Brachipoda (outgroups) were used for the phylogenetic analyses. Accession numbers and taxonomic classification for all mitochondrial genomes retrieved are available in S1 Table. All the retrieved sequences were then re-annotated using MITOS with subsequent manual curation as explained above. From the NCBI-formatted files, amino acid sequences in FASTA format were extracted and aligned using MAFFT v7.220 [48] with the '—localpair' option. Then, Gblocks v0.91b [49] was used with the options '-t = p -b2 = 21 -b3 = 10 -b4 = 5 -b5 = h' to permit smaller blocks. The rationale behind the use of amino acids instead of nucleotides resides in the difficulty of accurately modeling saturations, which is particularly exacerbating when working with distantly related taxa and/or fragmentary taxon sampling. The use of Gblocks reduced the final alignment (an average of 31% for each gene), resulting in a matrix with a smaller number of gaps. The resulting amino acid sequence alignments were then fed into ProtTest v3.4 [50] to search for the best-fit model of evolution for each gene. ProtTest suggested the following models: MtArt+I+G+F for *atp6*, *cox2*, *nad1*, *nad2*, *nad4*, *nad5*; MtArt+G for *cytb*, *cox3*, *nad3*, *nad6*; LG+I+G+F for *cox1*; MtREV+I+G+F for *nad4L*. The MtArt was implemented in MrBayes v3.2.4 [51] through the use of the option prset aarevmatpr for the replacement matrix and prset statefreqpr for the amino acid frequencies. For the Bayesian analyses, two independent runs, each with four chains (three "heated", one "cold") were run for 3,000,000 generations discarding the first 25% as burn-in, even though convergence was reached before 30,000 generations. The tree was then loaded into FigTree v1.4.1 (<http://tree.bio.ed.ac.uk/software/figtree/>) and edited with Inkscape v0.91 (<https://inkscape.org/en/>). For the parsimony analyses, a new technology search was performed in TNT [52] treating gaps as a fifth state and using 1000 initial addition sequences, five rounds of ratcheting [53] and three rounds of tree fusing [54] after the initial Wagner tree builds, and requiring that the minimum length tree be found a total of ten times. The resulting trees were then returned to TNT and used as starting trees for a new TBR branch swapping search using the command 'bbreak'. Bootstrap support values were calculated from 1000 pseudoreplicates with the same settings as mentioned above.

Results & Discussion

Mitochondrial genomes of *H. officinalis*, *P. lamothei* and *P. parasitica*

Genomic organization. Mitochondrial genomes of *H. officinalis*, *P. lamothei* and *P. parasitica* were assembled into single contigs of 14,849bp (linear) (ENA accn. number: LT159848),

14,909bp (linear) (ENA accn. number: LT159849) and 15,190bp (circular; closed) (ENA accn. number: LT159850), respectively. Although the first two genomes were not closed, they contain all 37 mitochondrial genes and are only lacking a very short part of the putative control region (pCR), when compared to the remaining taxa (although its length when complete is still unknown). All 37 genes are encoded on the same strand in all three mitochondrial genomes, which all contain a large non-coding region between the genes *trnR* and *trnH*—this region corresponds to the pCR. The position of the pCR is variable across the diversity of annelid genomes but fully conserved among representatives of Clitellata, with *Whitmania pigra* showing the shortest length (80bp) and highest A+T content (78.75%) (S2 Table).

Protein-coding and RNA genes. Protein-coding genes for the three leech genomes sequenced in this study were initially identified using MITOS, an integrated platform for mitochondrial genome annotation. The order of all 13 protein-coding genes is identical to that of previously sequenced mitochondrial genomes of clitellates (Fig 1). Several instances of tRNA gene rearrangements and duplications of tRNA genes were detected across the tree. Importantly, a *trnD* duplication was detected in the two *Placobdella* species included in the analysis (see below). Subsequently, we investigated the nucleotide compositions of all start and stop codons for the protein-coding genes (S3 Table). Most of these genes present putative incomplete stop codons (T or less frequently TA) but these are likely completed by post-transcriptional addition of 3'-end alanine residues [55–57]. These putative incomplete stop codons are flanked at the 3'-end by the 5'-end of tRNA genes or, less commonly, by another protein-coding gene. Much like other members of Clitellata, the 3'-end of the *nad4L* gene overlaps the 5'-end of the *nad4* gene. This feature is conserved in all Annelida mitochondrial genomes sequenced so far, except for that of *Clymenella torquata* (Leidy 1855)—this is probably a product of a misannotation in RefSeq. Regarding the start codons for the newly sequenced mitogenomes, all protein-coding genes share the canonical ATG start codon, except for that of *cox3*, where we observed a Glossiphoniidae-specific shift to the alternative start codon TTG. In *P. lamothei*, we also observed a shift from ATG to GTG in the gene *nad1*. This event seems to be something that has arisen independently in this particular species since its relative *P. parasitica* presents the canonical ATG.

In all three newly sequenced mitogenomes, as well as those of previously sequenced clitellates, rRNA genes are located between the *trnM* and the *trnL1* genes, and are only separated by the mid-rRNA placement of the *trnV* gene. The tRNA genes are relatively conserved in sequence order and structure in all three genomes, except for the remarkable case of the *trnR* gene in *P. parasitica* (S1 Fig). In this organism there is a 20 bp non-coding region between the *atp6* and *trnR* genes. This tRNA still preserves a typical cloverleaf structure, pointing towards the validity of the annotation of this apparent *P. parasitica*-specific change. Finally, putative overlap between tRNA genes with adjacent tRNA or protein-coding genes seems to be perfectly conserved in the three genomes sequenced here (Fig 1).

***Placobdella*-specific *trnD* duplication.** Surprisingly, our results indicate a *Placobdella*-specific duplication of the *trnD* gene (Figs 1 and 2). We were unable to find this duplication in any other completely sequenced annelid mitochondrial genome; we propose that this duplication is restricted to the *Placobdella* clade. Even more surprising is the fact that these two *trnD* genes seem to be potentially functional, inasmuch as they both preserve a cloverleaf structure and strong sequence similarity when compared to the remaining copy in the same species (Fig 2). Importantly, whereas there is no overlap or intergenic region between these two tRNA genes in *P. parasitica*, we found a 128 bp non-coding region (the longest such region except for the pCR) in *P. lamothei*. Using the mfold [58] webserver, we were able to find two small stem-loop structures in this region, the functions of which (if any) are unknown. It is also noteworthy that, in the case of *P. lamothei*, we found a change in the anticodon of the second *trnD* gene

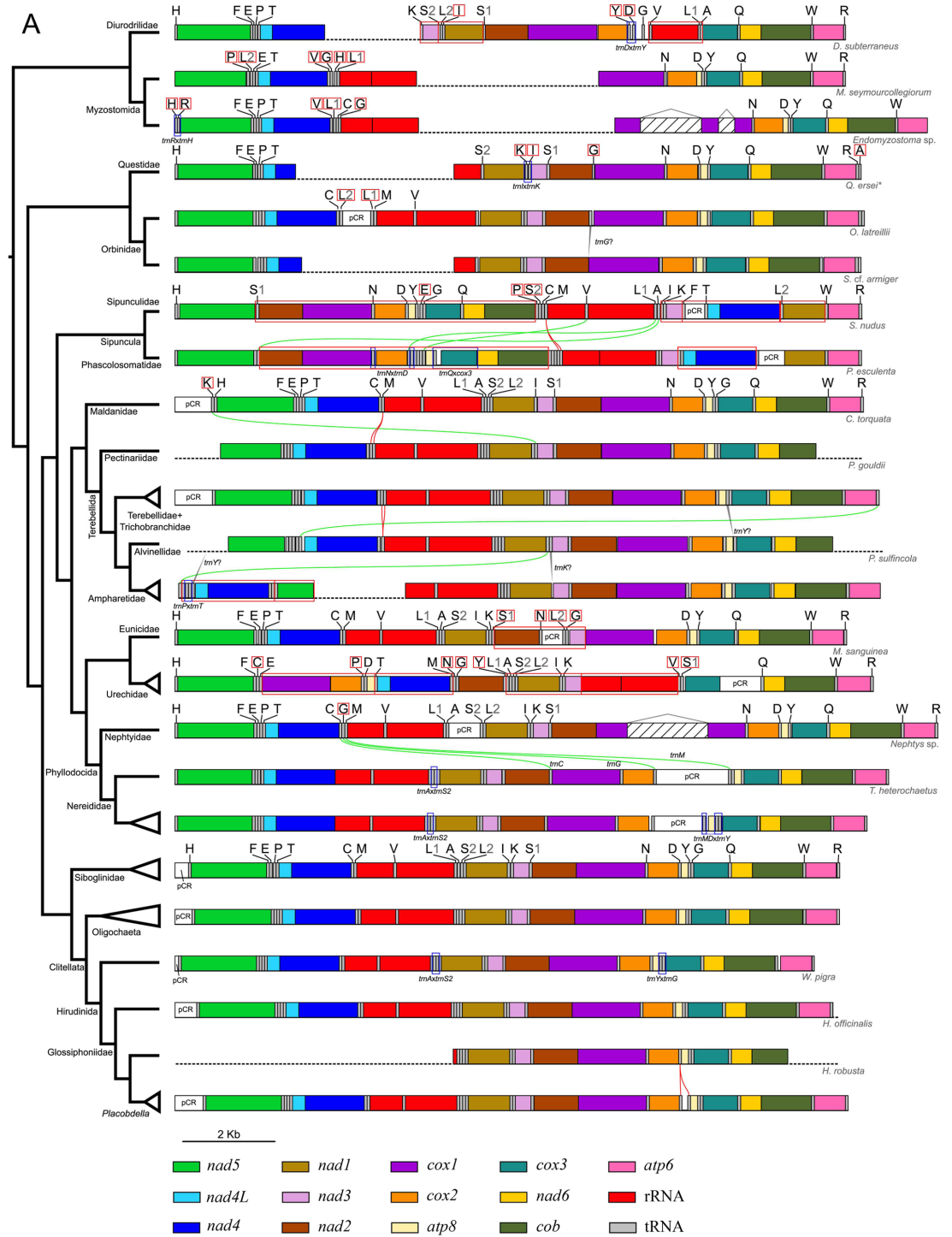


Fig 1. Gene order from available mitochondrial genomes of Annelida. The phylogeny on the left was inferred by Bayesian methodology. Colors indicate each of the 13 different protein-coding genes present in the mitochondrial genomes. Genes are scaled to real length. Green lines connecting one horizontal bar to another track the position of tRNA gene rearrangements. Red lines are used to indicate duplication of tRNA genes. Red boxes around tRNA gene names indicate putative rearrangements of tRNA genes relative to the mutual gene order for Clitellata. Red boxes around blocks of genes highlight syntenic regions, which may have been rearranged as a single unit. Blue boxes around blocks of genes denote inversions and, finally, incomplete or missing data are denoted by dashed lines. Names of RNA genes are denoted only when there has been a change in sequence, if RNA genes are not denoted, their sequence is equal to the closest sequence above for which genes are denoted. In some cases, however, minor

rearrangements are present and these are then denoted with the names of the RNA genes involved and separated by "x". For example, *trnAtrnS2* indicates that the *trnA* and *trnS2* genes have switched places. Gapped insertions inside *cox1* in *Endomyzostoma* sp. and *Nephtys* sp. indicate group II introns.

doi:10.1371/journal.pone.0155441.g001

from GUC to AUC (codons GAC and GAT, respectively). Given this change, we explored the amino acid composition of the proteins from Clitellata mitochondrial genomes, finding no statistical difference in the amino acid frequency for any proteome (Kruskal-Wallis paired test: chi-squared = 7.2333, df = 7, p-value = 0.405). We also analyzed, using the same set of proteins, the codon usage for the two alternative codons for aspartic acid (GAC and GAT) (Table 1). Using Fisher's exact test, we determined that there was indeed a difference in codon usage among the different mitochondrial genomes for Clitellata (p-value = 1.455E⁻¹⁵). By removing one or more terminals from all possible combinations, progressively, until reaching subgroups with no significant statistical differences among their members, we found two groups with similar codon usage: (i) Oligochaeta (Fisher's exact test: p-value = 0.1508), (ii) *W. pigra*, *H. officinalis* and *P. parasitica* (Fisher's exact test: p-value = 0.3999). The inclusion of *P. lamothei* into either one of these groups gave statistically significant differences (with i, Fisher's exact test: p-value = 0.01662; with ii, Fisher's exact test: p-value = 0.006462). This difference in codon usage in *P. lamothei* could be evidence for the actual functionality of the two *trnD* genes, as the shift in codon usage may allow a more balanced utilization in the anticodons.

Nevertheless, while all genomes of Hirudinea have a bias towards using the GAT codon, they still use the GUC anticodon to code for *trnD* genes (with the exception of *P. lamothei*). This evinces the fact that the specific tRNA anticodon is not directly influencing the codon usage for aspartic acid in *W. pigra*, *H. officinalis* and *P. parasitica*. It is worth noting that, within Annelida, there are at least two other cases of apparent functional tRNA gene

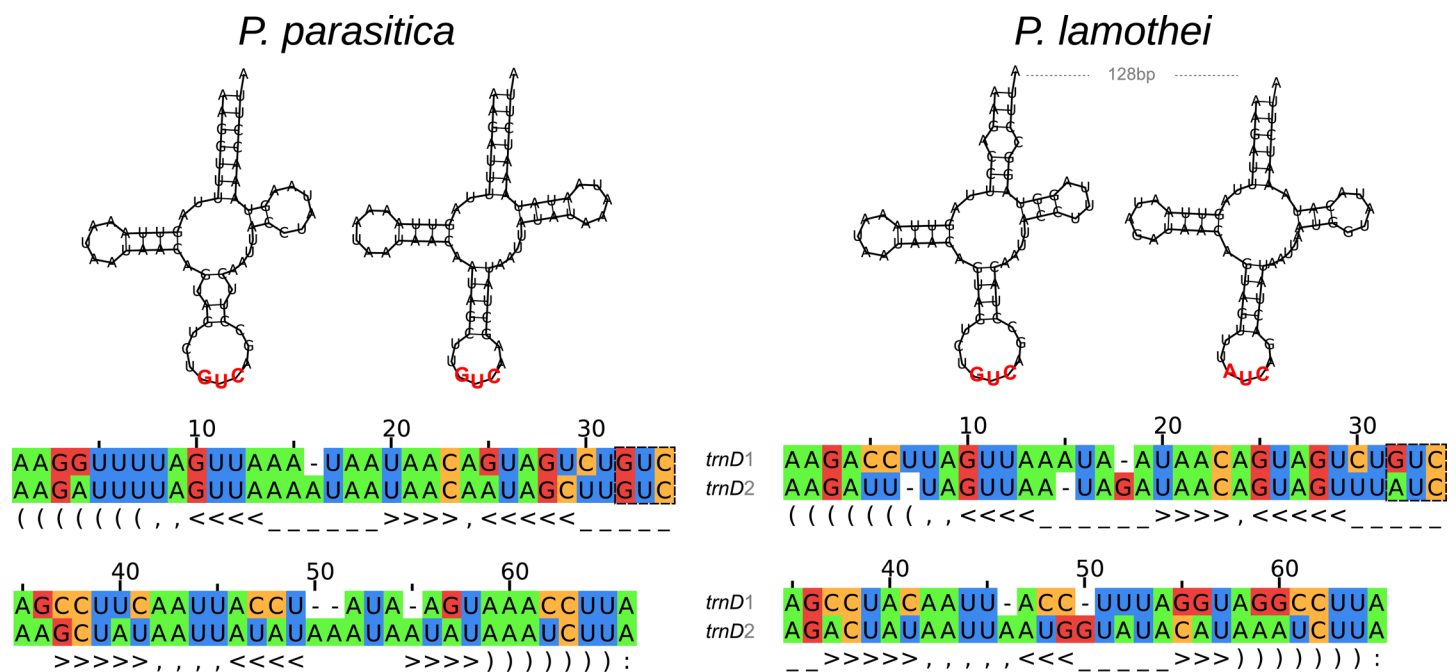


Fig 2. *trnD* duplication in *Placobdella*. Top: Secondary structures of *Placobdella trnD* genes as calculated by MITOS. Bottom: *trnD* sequences aligned against the 'Metazoa_D' model from MitFi v0.1 [77] using cmatch from Infernal v1.1rc4 [78]; sequence alignments took into account the secondary structure of the *trnD* genes. Note that gaps in the alignment occur in the loop structures or the *trnD*'s.

doi:10.1371/journal.pone.0155441.g002

Table 1. Codon usage for aspartic acid in Clitellata mitochondrial genomes. Number of codons present in proteins used to code aspartic acid in Clitellata mitochondrial genomes. Even though all Hirudinea, except for *P. lamothei*, have a strong bias for the use of the GAT codon, they exclusively code for the *trnD* gene with a GUC anticodon.

Organism	GAC	GAT	GAT	Ratio (GAC/GAT)
<i>Lumbricus terrestris</i>	34	40	64.42	0.85
<i>Perionyx excavatus</i>	42	32	75.00	1.31
<i>Tonoscolex birmanicus</i>	41	28	76.64	1.46
<i>Amynthes aspergillus</i>	43	28	71.68	1.54
<i>Metaphire vulgaris</i>	32	40	74.38	0.80
<i>Whitmania pigra</i>	12	78	78.75	0.15
<i>Haementeria officinalis</i>	8	58	~69.18	0.14
<i>Placobdella parasitica</i>	14	56	~70.10	0.25
<i>Placobdella lamothei</i>	27	45	75.63	0.60

doi:10.1371/journal.pone.0155441.t001

duplication—one within the sipunculids (*Phascolosoma esculenta* [Chen and Yeh 1958]) [59] and one within Terebellida (*Pectinaria gouldii* [Verrill 1873], *Pista cristata* [Müller 1776] and *Terebellides stroemi* [Sars 1835]) [26]. No functional explanation for these duplications has yet been provided.

It is worth noting that both *trnD* copies are located on the *cox2*—*atp8* junction, the very same position proposed as a hotspot for mitochondrial rearrangements in hymenopterans [60]; whether or not the occurrence of duplications at this junction has a biological meaning remains to be investigated in detail. This finding, apparently restricted to *Placobdella* species, represents a new case for the study of the mechanisms of “duplication/loss”, in which a fragment of the mitochondrial genome is duplicated by slipped-mispairing during replication, followed by the slow conversion of supernumerary genes to pseudogenes through random replication errors. If the duplication of the *trnD* occurred in the last common ancestor of the genus, estimated to exist around 250 mya. [61], each one of the more than 20 species of the genus represents an independent case for the study of disposal of supernumerary genes, as seems to be an explanation for mitogenomic stability. Otherwise, if independent duplications of *trnD* are inferred based on phylogenetic analyses of the genus, the hypothesis of the *cox2*—*atp8* junction as a hotspot should be explored in detail (see [62]).

Mitochondrial gene-order conservation across Annelida

Mitochondrial gene order within Annelida remains contentious; both a high level of stasis [25,59,63] and relatively high amount of rearrangements (e.g. [28,29,64]) has been reported for select taxa. In a comprehensive overview of the phylum, Weigert et al. [30] suggest that major mitogenomic rearrangements are confined to early diverging lineages, whereas representatives of Pleistoannelida, including Errantia and Sedentaria (*sensu* [65]) show high conservation of gene order. To shed further light on this issue, we utilized our phylogenetic inferences as guides and analyzed the rearrangements that have occurred in the different taxa (Fig 1). Gene order (only the 13 protein codifying genes) in Clitellata is identical to that present in Siboglinidae, Nereididae, Nephtyidae, Alvinellidae, Terebellidae + Trichobranchidae, Pectinariidae, Maldanidae, Orbinidae, Questidae and Myzostomida, and is also identical to the putative ground pattern of Pleistoannelida as described by Weigert et al. [30]. Changes in mitochondrial architecture are restricted to Eunicidae, Urechidae, Ampharetidae, Sipunculidae and Diurodrilidae and are discussed in comparison with the putative ground pattern for Pleistoannelida. In our analyses, Diurodrilidae by itself (Fig 3B) or as a clade together with Myzostomida (Figs 1

and 3A) places as the sister group to the remaining annelids and shows two modifications in its mitochondrial genome gene order, the first is a change in the order of *nad1* and *nad3* and the second is a new allocation of the rRNA. By contrast, the two representatives of Myzostomida both display the same protein coding gene order as the ground pattern for Pleistoannelida (Fig 1). Whereas Myzostomida seems to only show rearrangements of tRNA genes, Sipunculidae (also an early diverging lineage; Fig 3B) displays large rearrangements involving the translocation of various multigene syntenic blocks (*nad2*–*cob* switch location with *nad4L* and *nad4*). In addition, *nad1* also is reallocated. In corroborating the findings of Weigert et al. [30], we also observed identical gene order stasis within some distinct and well-defined groups. Within Terebellomorpha and Maldanidae, gene order is perfectly conserved, with the exception of a relocation of the *trnK* gene in *C. torquata* and a secondary loss of the *trnM* gene duplication in Ampharetidae; the latter taxon also portrays a seemingly more derived gene order, having rearranged the syntenic block *nad4L*–*nad4* for *nad5*. Within Phyllodocida, *Nephtys* sp. displays the peculiarity of having a group II intron inserted in the *cox1* gene [66]. Interestingly, this is also the case for the included species of *Endomyzostoma*. Nereididae shows a translocation of the *trnC*, *trnG* and *trnM* genes as well as an inversion of the *trnA* and *trnS2* genes when compared to *Nephtys* sp. Within the Questidae/Orbiniidae clade, *Questa ersei* Jamieson and Webb, 1984 shows an apparent lineage-specific inversion of the *trnI* and *trnK* genes—whether or not other species of Questidae share this is not known due to the lack of comparative data. In addition, *Scoloplos* cf. *armiger* shows a translocation of the *trnG* gene to an undetermined region between the *nad4* and *rrnL* gene. Lastly, the Siboglinidae/Clitellata clade recovered here, notwithstanding the phylogenetic method used or inclusion of Myzostomida, displays perfect synteny in mitochondrial gene-order, the same order as that observed for Questidae, Orbiniidae, Phyllodocida, Maldanidae, Terebellomorpha and matching the gene order recovered for most annelids [67]. The modified gene order within Urechidae (*cox1*–*atp8*; *nad1*–*nad2* and rRNA reallocations) and Eunicidae (*nad2* and *nad3* show switched positions) contradicts, to some extent, the finding by Weigert et al. [30] that major mitogenomic rearrangements are confined to early diverging lineages.

Phylogenetic analyses

Beyond placing the newly sequenced taxa in a molecular phylogenetic framework, based on the mitochondrial genomes, the present study also capitalized on the assembly of the main dataset for the addressing of phylogenetic relationships across Annelida. A dataset consisting of 30 complete and 12 partially sequenced annelid mitochondrial genomes was assembled, including representatives of Myzostomida, Urechidae and Sipunculidae [18,19,21,22,27,63,68]. From these genomes, we generated a concatenated alignment of amino acids that was then used both for a partitioned Bayesian phylogenetic reconstruction and parsimony analysis. Because the inclusion of Myzostomida has been shown to possibly create discrepancies due to long-branch attraction [67], we decided to run two independent phylogenetic analyses, including and excluding myzostomids—the datasets are here referred to as the 42-taxa dataset (including Myzostomida) and the 40-taxa dataset (excluding Myzostomida). Both parsimony analyses resulted in single most parsimonious trees; 21,874 steps were recorded for the 42-taxa dataset with a retention index (RI) of 0.47 and consistency index (CI) of 0.46, and for the 40-taxa dataset, the single most parsimonious tree was 21,594 steps with an RI of 0.48 and a CI of 0.47. Two minor discrepancies exist between the Bayesian and the parsimony trees using the 42-taxa dataset. In the Bayesian tree, the sipunculids place as sister to a clade that includes most of the polychaetes (Bayesian posterior probability [PP] = 1.00), whereas in the parsimony tree, sipunculids is sister to *Clymenella torquata* (bootstrap support value [BS] = <50%). Moreover,

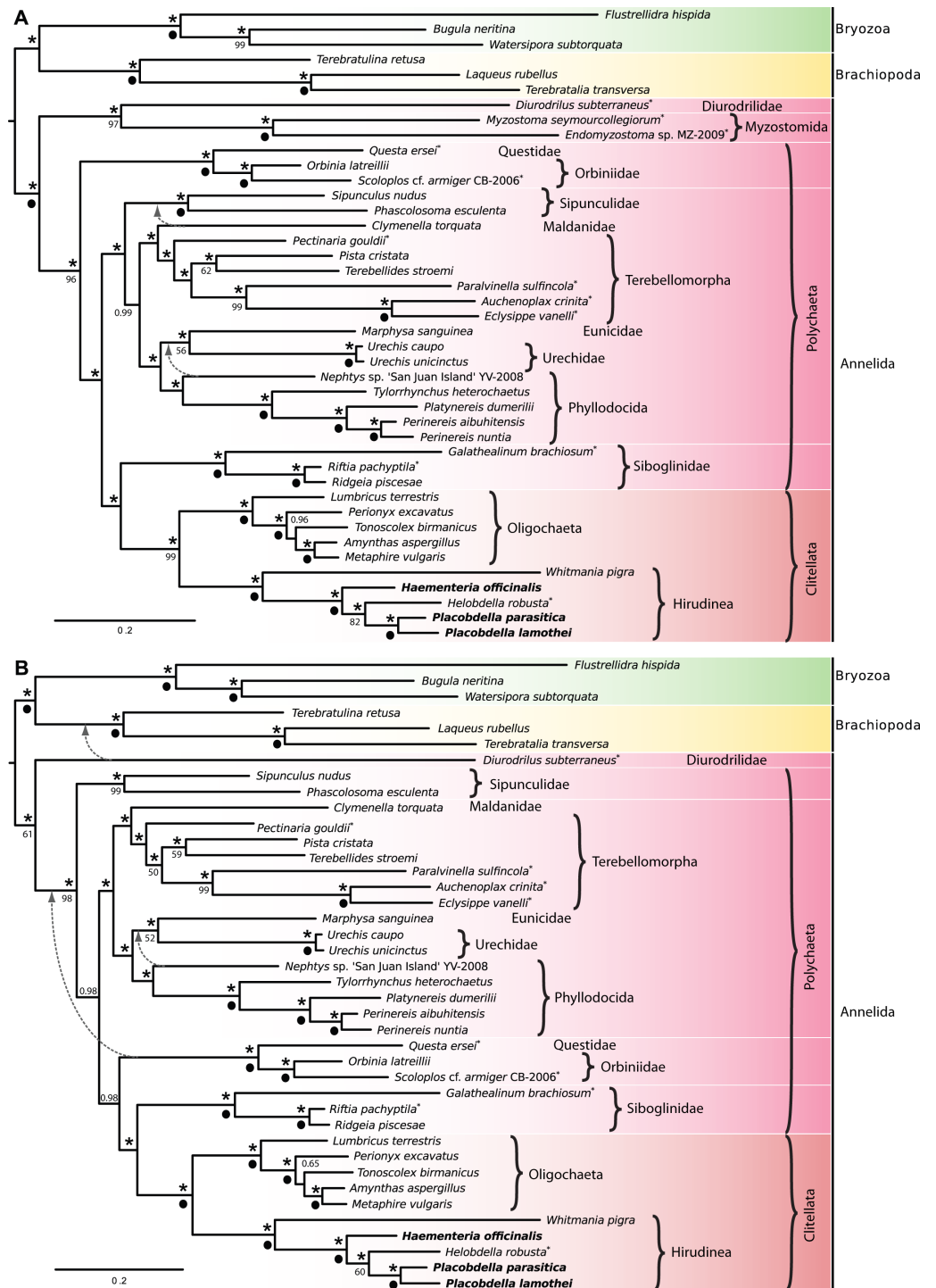


Fig 3. Phylogenetic relationships of Annelida. Bayesian phylogenetic reconstruction as inferred by MrBayes with Myzostomida included, A, and excluded, B. Values above nodes are Bayesian posterior probabilities, and those below the nodes are parsimony bootstrap support values above 50%; asterisks denote maximum posterior probabilities and black circles denote maximum parsimony bootstrap support. Taxa in bold were sequenced for the present study and broken lines indicate the alternative positioning of the respective taxa in the single most parsimonious trees.

doi:10.1371/journal.pone.0155441.g003

Nephtys sp. places as the sister group of the remaining phyllodocids (PP = 1.00) in the Bayesian tree (Urechidae + Eunicidae places as the sister to this group), whereas the taxon places as sister to Urechidae+Eunicidae (BS = <50%) in the parsimony tree (Fig 3A). For the 40-taxa dataset, some major discrepancies are recovered when comparing the trees from the different analyses (Fig 3B). Most importantly, *Diurodrilus subterraneus* Remane 1934 is recovered as the sister group to Brachiopoda in the parsimony tree (BS = 61%), whereas its placement as the sister group to the remaining annelids is maintained in the Bayesian reconstruction (PP = 1.00). Oddly enough, the placement of Questidae+Orbiniidae changes rather dramatically in the Bayesian analysis when excluding the myzostomids as it is recovered as the sister to Siboglinidae+Clitellata (PP = 1.00). To shed light on this issue, future studies should focus on increasing the sampling of annelid taxa, as it has been shown that the inclusion of potential sister-taxa can minimize the effects of long-branch attraction in phylogenetic reconstructions [69,70].

Regarding the three glossiphoniid leeches sequenced for the present study, these form a monophyletic group with the partially sequenced *Helobdella robusta* (PP = 1.00 and BS = 100% for both datasets) and this group, in turn, places as the sister taxon to Arhynchobdellida (PP = 1.00 and BS = 100% for both datasets), which is here represented by the *W. pigra*. Interestingly, *H. officinalis* is recovered as the sister group to an *H. robusta*+*Placobdella* clade (PP = 1.00 for both datasets; BS = 82% and 84% for the 42-taxa dataset and 40-taxa dataset, respectively), which contradicts previous findings in which *Haementeria*+*Helobdella* are part of a sister radiation of *Placobdella* [71–73]. However, our results are most likely affected by shallow taxon sampling; the inclusion of more fully sequenced mitochondrial genomes for leeches may revert this result. It is also important to note that the included mitochondrial genome of *H. robusta* is incomplete and a substantial amount of missing data was present for this taxon, which may or may not alter the results. The phylogenetic positioning of *H. robusta* (a non-blood feeder) would further support the idea that *H. robusta* has secondarily lost the capacity for feeding on blood, along with the loss of its bacteriome and related endosymbiotic bacteria, traits that are present in species of both *Haementeria* and *Placobdella* [61]. Despite the paucity of available mitochondrial genome data for leeches, the current study represents the first phylogenetic analysis of full genomes that includes representatives of both leech orders (Rhynchobdellida and Arhynchobdellida). Considering that the gene order among the leeches included here is fully conserved, we hypothesize that no major rearrangements in the gene order and composition will be found after inclusion of more species.

The phylogenetic hypotheses presented here disagree to a large extent with some contemporary hypotheses. For example, contrary to Golombek et al. [67] and Weigert et al. [30], *Diurodrilus subterraneus* was here recovered as an early diverging lineage in all trees but in some of them with low support (PP = 1.00 for both datasets; BS = <50% and 61% for the 42-taxa dataset and the 40-taxa dataset, respectively) (Fig 3A and 3B). Furthermore, by virtue of the urechids nesting within Errantia, and Errantia nesting within Sedentaria (both placements with maximum posterior probabilities, but with much lower parsimony bootstrap support, 52% and <50%, respectively), Errantia and Sedentaria were never recovered as monophyletic. Beyond major inconsistencies regarding Errantia and Sedentaria, there are numerous minor differences in the branching patterns (cf. Fig 3 and Weigert et al. [30]: Fig 3) and seeing as the better part of these datasets overlap (with the exception of our use of Gblocks), these differences seem to be attributable to some specifics of the analyses and/or slightly different terminals included. For example, only *Helobdella robusta* and *Hirudo nipponia* were included in the datasets analyzed by Weigert et al. [30], and only one of these was included in each of the analyses performed by Weigert et al. [30]. Although the taxon sampling for Hirudinea was improved for the present analyses, we lack the inclusion of more “basal branching” members, which might also impact the phylogenetic inference. Regardless, the varying topologies in contemporary

leech mitogenome analyses may simply indicate sensitivity to the elevated substitution rates for both amino acids and nucleotides, inherent in mitochondrial genomes [74]. Beyond this, recent investigations into the mitochondrial genomes of bilaterians suggest that mitochondrial data alone may not be fit for phylogenetic reconstructions inasmuch as the inversion of genes, changes in replication orders and compositional differences (e.g. AT-biases) all affect the resulting phylogeny [75,76]. In light of the discrepancies between our trees and previously published hypotheses, our results seem to substantiate this claim.

Conclusions

As shown in Fig 1, genome architecture is conserved among the currently sequenced mitochondrial genomes of clitellates. Among other things, conservation of gene order is perfect, with the exception of two inversions found in *W. pigra* involving the tRNA gene-pairs *trnA-trnS2* and *trnY-trnG*. We also discovered a *trnD* gene duplication, which seems to be restricted to species of *Placobdella* but, nevertheless, we were unable to identify an explanation as to why this duplication occurred and has been maintained—it could be argued that not enough time has elapsed since the duplication event for it to be selectively purged. Additionally, by analyzing the different genome rearrangements of the currently available annelid mitochondrial genomes, one can find distinct inversions or translocations, which are specific to the different taxa.

Supporting Information

S1 Fig. tRNA structures of Clitellata mitochondrial genomes. tRNA structures of Clitellata tRNA genes as predicted by MITOS and manually curated. At the top of each set of blocks, dendograms showing the phylogenetic relationship of the organisms as calculated in the bayesian phylogenetic reconstruction. At the left of each individual block, the tRNA gene name indicated in bold italic letters with the anticodon sequence between parenthesis. On the tRNA structures, red letters indicate the anticodon, grey parallelograms indicate base pairs that overlap another gene sequence. In a blue block, the *Placobdella*-specific *trnD2* gene is shown with its two possible anticodons. Missing data for *H. robusta* are due to the lack of comparative data.

(PDF)

S1 File. Concatenated alignment including myzostomids. Concatenated amino acid alignment used for phylogenetic reconstruction including myzostomids. The alignment was “cleaned” using Gblocks and is in nexus format.

(NEX)

S2 File. Concatenated alignment excluding myzostomids. Concatenated amino acid alignment used for phylogenetic reconstruction excluding myzostomids. The alignment was “cleaned” using Gblocks and is in nexus format.

(NEX)

S1 Table. Accession numbers and taxonomic classification for all mitochondrial genomes used in this study.

(XLSX)

S2 Table. General characteristics of putative control regions (pCRs) of mitochondrial genomes of Annelida.

(XLSX)

S3 Table. Start and stop codons for all 13 protein-coding genes from the mitochondrial genomes of Clitellata.
(XLSX)

Acknowledgments

We thank the two anonymous reviewers and Editor Marc Robinson-Rechavi for thoughtful comments that greatly improved an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: AOF AMM SK AM MS AL. Performed the experiments: AOF AMM SK. Analyzed the data: AOF AMM SK. Contributed reagents/materials/analysis tools: AOF AMM SK AM MS AL. Wrote the paper: AOF AMM SK AM MS AL.

References

1. Miya M, Nishida M. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Mol Phylogenet Evol.* 2000; 17: 437–455. doi: [10.1006/mpev.2000.0839](https://doi.org/10.1006/mpev.2000.0839) PMID: [11133198](https://pubmed.ncbi.nlm.nih.gov/11133198/)
2. Paton T, Haddrath O, Baker AJ. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. *Proc Roy Soc London B: Biol Sci.* 2002; 269: 839–846. doi: [10.1098/rspb.2002.1961](https://doi.org/10.1098/rspb.2002.1961)
3. Roos J, Aggarwal RK, Janke A. Extended mitogenomic phylogenetic analyses yield new insight into crocodylian evolution and their survival of the Cretaceous–Tertiary boundary. *Mol Phylogenet Evol.* 2007; 45: 663–673. doi: [10.1016/j.ympev.2007.06.018](https://doi.org/10.1016/j.ympev.2007.06.018) PMID: [17719245](https://pubmed.ncbi.nlm.nih.gov/17719245/)
4. Jia WZ, Yan HB, Guo AJ, Zhu XQ, Wang YC, Shi WG, et al. Complete mitochondrial genomes of *Taenia multiceps*, *T. hydatigena* and *T. pisiformis*: additional molecular markers for a tapeworm genus of human and animal health significance. *BMC Genomics.* 2010; 11: 447. doi: [10.1186/1471-2164-11-447](https://doi.org/10.1186/1471-2164-11-447) PMID: [20649981](https://pubmed.ncbi.nlm.nih.gov/20649981/)
5. Plazzi F, Ceregato A, Taviani M, Passamonti M. A molecular phylogeny of bivalve mollusks: ancient radiations and divergences as revealed by mitochondrial genes. *PLoS ONE.* 2011; 6: e27147. doi: [10.1371/journal.pone.0027147](https://doi.org/10.1371/journal.pone.0027147) PMID: [22069499](https://pubmed.ncbi.nlm.nih.gov/22069499/)
6. Hebert PDN, Cywinska A, Ball SL. Biological identifications through DNA barcodes. *Proc Roy Soc London B: Biol Sci.* 2003; 270: 313–321. doi: [10.1098/rspb.2002.2218](https://doi.org/10.1098/rspb.2002.2218)
7. Boore JL, Brown WM. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev.* 1998; 8: 668–674. doi: [10.1016/S0959-437X\(98\)80035-X](https://doi.org/10.1016/S0959-437X(98)80035-X) PMID: [9914213](https://pubmed.ncbi.nlm.nih.gov/9914213/)
8. Boore JL, Macey JR, Medina M. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol.* 2005; 395: 311–348. doi: [10.1016/S0076-6879\(05\)95019-2](https://doi.org/10.1016/S0076-6879(05)95019-2) PMID: [15865975](https://pubmed.ncbi.nlm.nih.gov/15865975/)
9. Lin FJ, Liu Y, Sha Z, Tsang LM, Chu KH, Chan TY, et al. Evolution and phylogeny of the mud shrimps (Crustacea: Decapoda) revealed from complete mitochondrial genomes. *BMC Genomics.* 2012; 13: 631. doi: [10.1186/1471-2164-13-631](https://doi.org/10.1186/1471-2164-13-631) PMID: [23153176](https://pubmed.ncbi.nlm.nih.gov/23153176/)
10. Anderson L. Identification of mitochondrial proteins and some of their precursors in two-dimensional electrophoretic maps of human cells. *Proc Natl Acad Sci.* 1981; 78: 2407–2411. PMID: [6941300](https://pubmed.ncbi.nlm.nih.gov/6941300/)
11. Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, et al. A comparison of the human and bovine mitochondrial genomes. In: Slonimski P, Borst P, Attardi G, editors. *Mitochondrial Genes.* Cold Spring Harbor Laboratories, Cold Spring Harbor, New York. 1982; pp. 5–43.
12. Desjardins P, Morais R. Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *J Mol Biol.* 1990; 212: 599–634. doi: [10.1016/0022-2836\(90\)90225-B](https://doi.org/10.1016/0022-2836(90)90225-B) PMID: [2329578](https://pubmed.ncbi.nlm.nih.gov/2329578/)
13. Lavrov DV, Brown WM. *Trichinella spiralis* mtDNA: a nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene arrangement relatable to those of coelomate metazoans. *Genetics.* 2001; 157: 621–637. PMID: [11156984](https://pubmed.ncbi.nlm.nih.gov/11156984/)
14. Broughton RE, Milam JE, Roe BA. The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res.* 2001; 11: 1958–1967. doi: [10.1101/gr.156801](https://doi.org/10.1101/gr.156801) PMID: [11691861](https://pubmed.ncbi.nlm.nih.gov/11691861/)

15. Zhang ZQ. Animal biodiversity: An introduction to higher-level classification and taxonomic richness. *Zootaxa*. 2011; 3148: 7–12.
16. Brusca RC, Brusca GJ. *Invertebrates* 2nd ed. Sunderland, Massachusetts: Sinauer Associates; 1990.
17. Rousset V, Pleijel F, Rouse GW, Erseus C, Siddall ME. A molecular phylogeny of annelids. *Cladistics*. 2007; 23: 41–63. doi: [10.1111/j.1096-0031.2006.00128.x](https://doi.org/10.1111/j.1096-0031.2006.00128.x)
18. Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D. et al. Annelid phylogeny and the status of Sipuncula and Echiura. *BMC Evol Biol*. 2007; 7: 57. doi: [10.1186/1471-2148-7-57](https://doi.org/10.1186/1471-2148-7-57) PMID: [17411434](https://pubmed.ncbi.nlm.nih.gov/17411434/)
19. Struck TH, Paul C, Hill N, Hartmann S, Hösel C, Kube M, et al. Phylogenomic analyses unravel annelid evolution. *Nature*. 2011; 471: 95–98. doi: [10.1038/nature09864](https://doi.org/10.1038/nature09864) PMID: [21368831](https://pubmed.ncbi.nlm.nih.gov/21368831/)
20. Zrzavý J, Říha P, Piálek L, Janouškovec J. Phylogeny of Annelida (Lophotrochozoa): total-evidence analysis of morphology and six genes. *BMC Evol Biol*. 2009; 9: 189. doi: [10.1186/1471-2148-9-189](https://doi.org/10.1186/1471-2148-9-189) PMID: [19660115](https://pubmed.ncbi.nlm.nih.gov/19660115/)
21. Kvist S, Siddall ME. Phylogenomics of Annelida revisited: a cladistic approach using genome-wide expressed sequence tag data mining and examining the effects of missing data. *Cladistics*. 2013; 29: 435–448. doi: [10.1111/cla.12015](https://doi.org/10.1111/cla.12015)
22. Weigert A, Helm C, Meyer M, Nickel B, Arendt D, Hausdorf B, et al. Illuminating the base of the annelid tree using transcriptomics. *Mol Biol Evol*. 2014; 31: 1391–401. doi: [10.1093/molbev/msu080](https://doi.org/10.1093/molbev/msu080) PMID: [24567512](https://pubmed.ncbi.nlm.nih.gov/24567512/)
23. Andrade S, Novo M, Kawauchi G, Worsaae K, Pleijel F, Giribet G, et al. Articulating “archannelids”: Phylogenomics and annelid relationships, with emphasis on meiofaunal taxa. *Mol Biol Evol*. 2015; In press. doi: [10.1093/molbev/msv157](https://doi.org/10.1093/molbev/msv157)
24. Jennings RM, Halanych KM. Mitochondrial genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): evidence for conserved gene order in Annelida. *Mol Biol Evol*. 2005; 22: 210–222. PMID: [15483328](https://pubmed.ncbi.nlm.nih.gov/15483328/)
25. Vallès Y, Boore JL. Lophotrochozoan mitochondrial genomes. *Integr Comp Biol*. 2006; 46: 544–57. doi: [10.1093/icb/icj056](https://doi.org/10.1093/icb/icj056) PMID: [21672765](https://pubmed.ncbi.nlm.nih.gov/21672765/)
26. Zhong M, Struck TH, Halanych KM. Phylogenetic information from three mitochondrial genomes of Terbelliformia (Annelida) worms and duplication of the methionine tRNA. *Gene*. 2008; 416: 11–21. doi: [10.1016/j.gene.2008.02.020](https://doi.org/10.1016/j.gene.2008.02.020) PMID: [18439769](https://pubmed.ncbi.nlm.nih.gov/18439769/)
27. Bleidorn C, Eeckhaut I, Podsiadlowski L, Schult N, McHigh D, Halanych KM, et al. Mitochondrial genome and nuclear sequence data support myzostomida as part of the annelid radiation. *Mol Biol Evol*. 2007; 24: 1690–701. doi: [10.1093/molbev/msm086](https://doi.org/10.1093/molbev/msm086) PMID: [17483114](https://pubmed.ncbi.nlm.nih.gov/17483114/)
28. Aguado MT, Glasby C, Schroeder PC, Weigert A, Bleidorn C. The making of a branching annelid: an analysis of complete mitochondrial genome and ribosomal data of *Ramisyllis multicaudata*. *Scientific Reports*. 2015; 5: 12072. doi: [10.1038/srep12072](https://doi.org/10.1038/srep12072) PMID: [26183383](https://pubmed.ncbi.nlm.nih.gov/26183383/)
29. Li Y, Kocot KM, Schander C, Santos SR, Thornhill DJ, Halanych KM. Mitogenomics reveals phylogeny and repeated motifs in control regions of the deep-sea family Siboglinidae (Annelida). *Mol Phylogenet Evol*. 2015; 85: 221–229. doi: [10.1016/j.ympev.2015.02.008](https://doi.org/10.1016/j.ympev.2015.02.008) PMID: [25721539](https://pubmed.ncbi.nlm.nih.gov/25721539/)
30. Weigert A, Golombek A, Gerth M, Schwarz F, Struck TH, Bleidorn C. Evolution of mitochondrial gene order in Annelida. *Mol Phylogenet Evol*. 2016; 94: 196–206. doi: [10.1016/j.ympev.2015.08.008](https://doi.org/10.1016/j.ympev.2015.08.008) PMID: [26299879](https://pubmed.ncbi.nlm.nih.gov/26299879/)
31. Apakupakul K, Siddall ME, Burreson EM. Higher level relationships of leeches (Annelida: Clitellata: Euhirudinea) based on morphology and gene sequences. *Mol Phylogenet Evol*. 1999; 12: 350–359. doi: [10.1006/mpev.1999.0639](https://doi.org/10.1006/mpev.1999.0639) PMID: [10413628](https://pubmed.ncbi.nlm.nih.gov/10413628/)
32. Sket B, Trontelj P. Global diversity of leeches (Hirudinea) in freshwater. *Hydrobiologia*. 2008; 595: 129–137. doi: [10.1007/s10750-007-9010-8](https://doi.org/10.1007/s10750-007-9010-8)
33. Boore JL, Brown WM. Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol Biol Evol*. 2000; 17: 87–106. PMID: [10666709](https://pubmed.ncbi.nlm.nih.gov/10666709/)
34. Shen X, Wu Z, Sun MA, Ren J, Liu B. The complete mitochondrial genome sequence of *Whitmania pigra* (Annelida, Hirudinea): the first representative from the class Hirudinea. *Comp Biochem Physiol Part D: Genomics and Proteomics*. 2011; 6: 133–138. doi: [10.1016/j.cbd.2010.12.001](https://doi.org/10.1016/j.cbd.2010.12.001) PMID: [21212033](https://pubmed.ncbi.nlm.nih.gov/21212033/)
35. Ye F, Liu T, Zhu W, You P. Complete mitochondrial genome of *Whitmania laevis* (Annelida, Hirudinea) and comparative analyses within *Whitmania* mitochondrial genomes. *Belg J Zool*. 2015; 145: 114–128.
36. Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jimenez-García LF, Moya A. Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech *Haementeria officinalis*. *Genome Biol Evol*. 2015; In press. doi: [10.1093/gbe/evv188](https://doi.org/10.1093/gbe/evv188)

37. Jérôme M, Noirot C, Klopp C. Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool. *BMC Res Notes*. 2011; 4: 149. doi: [10.1186/1756-0500-4-149](https://doi.org/10.1186/1756-0500-4-149) PMID: [21615897](https://pubmed.ncbi.nlm.nih.gov/21615897/)
38. Brady A, Salzberg SL. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods*. 2009; 6: 673–676. doi: [10.1038/nmeth.1358](https://doi.org/10.1038/nmeth.1358) PMID: [19648916](https://pubmed.ncbi.nlm.nih.gov/19648916/)
39. Kvist S, Narechania A, Oceguera-Figueroa A, Fuks B, Siddall ME. Phylogenomics of *Reichenowia parasitica*, an alphaproteobacterial endosymbiont of the freshwater leech *Placobdella parasitica*. *PLoS ONE*. 2011; 6: e28192. doi: [10.1371/journal.pone.0028192](https://doi.org/10.1371/journal.pone.0028192) PMID: [22132238](https://pubmed.ncbi.nlm.nih.gov/22132238/)
40. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011; 27: 863–4. doi: [10.1093/bioinformatics/btr026](https://doi.org/10.1093/bioinformatics/btr026) PMID: [21278185](https://pubmed.ncbi.nlm.nih.gov/21278185/)
41. Oceguera-Figueroa, A. Biodiversity, taxonomy and systematics of the New World freshwater leeches (Annelida: Hirudinea) with particular emphasis on Glossiphoniid leeches and their bacterial endosymbionts. (Ph.D. thesis) The City University of New York. New York, USA. 2011; 257 pp.
42. Chevreux B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. In: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*. Heidelberg, Germany 1999; pp. 45–46. <http://www.bioinfo.de/isb/gcb99/talks/chevreux/>.
43. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol*. 2012; 13: R56. doi: [10.1186/gb-2012-13-6-r56](https://doi.org/10.1186/gb-2012-13-6-r56) PMID: [22731987](https://pubmed.ncbi.nlm.nih.gov/22731987/)
44. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013; 69: 313–319. doi: [10.1016/j.ympev.2012.08.023](https://doi.org/10.1016/j.ympev.2012.08.023) PMID: [22982435](https://pubmed.ncbi.nlm.nih.gov/22982435/)
45. Nakamura Y, Cochrane G, Karsch-Mizrachi I. The international nucleotide sequence database. *Nucl Acids Res*. 2013; 41 (D1): D21–D24.
46. R Core Team. R: A Language and Environment for Statistical Computing. 2015; <http://www.r-project.org/>.
47. Guy L, Kultima JR, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics*. 2010; 26: 2334–2335. doi: [10.1093/bioinformatics/btq413](https://doi.org/10.1093/bioinformatics/btq413) PMID: [20624783](https://pubmed.ncbi.nlm.nih.gov/20624783/)
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. doi: [10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
49. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007; 56: 564–577. doi: [10.1080/10635150701472164](https://doi.org/10.1080/10635150701472164) PMID: [17654362](https://pubmed.ncbi.nlm.nih.gov/17654362/)
50. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27: 1164–1165. doi: [10.1093/bioinformatics/btr088](https://doi.org/10.1093/bioinformatics/btr088) PMID: [21335321](https://pubmed.ncbi.nlm.nih.gov/21335321/)
51. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012; 61: 539–542. doi: [10.1093/sysbio/sys029](https://doi.org/10.1093/sysbio/sys029) PMID: [22357727](https://pubmed.ncbi.nlm.nih.gov/22357727/)
52. Goloboff PA, Farris JS, Nixon KC. TNT, a free program for phylogenetic analysis. *Cladistics*. 2008; 24: 774–786. doi: [10.1111/j.1096-0031.2008.00217.x](https://doi.org/10.1111/j.1096-0031.2008.00217.x)
53. Nixon K. The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*. 1999; 15: 407–414.
54. Goloboff PA. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics*. 1999; 15: 415–428.
55. Ojala D, Merkel C, Gelfand R, Attardi G. The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell*. 1980; 22:393–403. doi: [10.1016/0092-8674\(80\)90350-5](https://doi.org/10.1016/0092-8674(80)90350-5) PMID: [7448867](https://pubmed.ncbi.nlm.nih.gov/7448867/)
56. Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. *Nature*. 1981; 290: 470–474. doi: [10.1038/290470a0](https://doi.org/10.1038/290470a0) PMID: [7219536](https://pubmed.ncbi.nlm.nih.gov/7219536/)
57. Stewart JB, Beckenbach AT. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. *Gene*. 2009; 445: 49–57. doi: [10.1016/j.gene.2009.06.006](https://doi.org/10.1016/j.gene.2009.06.006) PMID: [19540318](https://pubmed.ncbi.nlm.nih.gov/19540318/)
58. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003; 31: 3406–3415. doi: [10.1093/nar/gkg595](https://doi.org/10.1093/nar/gkg595) PMID: [12824337](https://pubmed.ncbi.nlm.nih.gov/12824337/)
59. Shen X, Ma X, Ren J, Zhao F. A close phylogenetic relationship between Sipuncula and Annelida evidenced from the complete mitochondrial genome sequence of *Phascolosoma esculenta*. *BMC Genomics*. 2009; 10: 136. doi: [10.1186/1471-2164-10-136](https://doi.org/10.1186/1471-2164-10-136) PMID: [19327168](https://pubmed.ncbi.nlm.nih.gov/19327168/)

60. Downton M, Austin AD. Evolutionary dynamics of a mitochondrial rearrangement" hot spot" in the Hymenoptera. *Mol Biol Evol.* 1999; 16: 298–309. PMID: [10028295](#)
61. Perkins SL, Budinoff R, Siddall ME. New Gammaproteobacteria associated with blood-feeding leeches and a broad phylogenetic analysis of leech endosymbionts. *Appl Environ Microbiol.* 2005; 71: 5219–5224. doi: [10.1128/AEM.71.9.5219](#) PMID: [16151107](#)
62. Downton M, Campbell NJH. Intramitochondrial recombination—is it why some mitochondrial genes sleep around?. *Trends Ecol Evol.* 2001; 16: 269–271. doi: [10.1016/S0169-5347\(01\)02182-6](#) PMID: [11369092](#)
63. Wu Z, Shen X, Sun MA, Ren J, Wang Y, Huang Y, et al. Phylogenetic analyses of complete mitochondrial genome of *Urechis unicinctus* (Echiura) support that echiurans are derived annelids. *Mol Phylogenet Evol.* 2009; 52: 558–562. doi: [10.1016/j.ympev.2009.03.009](#) PMID: [19306935](#)
64. Bleidorn C, Podsiadlowski L, Bartolomaeus T. The complete mitochondrial genome of the orbiniid polychaete *Orbinia latrellii* (Annelida, Orbinidae)—A novel gene order for Annelida and implications for annelid phylogeny. *Gene.* 2006; 370: 96–103. doi: [10.1016/j.gene.2005.11.018](#) PMID: [16448787](#)
65. Struck TH. Direction of evolution within Annelida and the definition of Pleistoannelida. *J Zool Syst Evol Res.* 2011; 49: 340–345. doi: [10.1111/j.1439-0469.2011.00640.x](#)
66. Vallès Y, Halanych KM, Boore JL. Group II introns break new boundaries: presence in a bilaterian's genome. *PLoS ONE.* 2008; 3: e1488. doi: [10.1371/journal.pone.0001488](#) PMID: [18213396](#)
67. Golombek A, Tobergte S, Nesnidal MP, Purschke G, Struck TH. Mitochondrial genomes to the rescue —Diurodrilidae in the myzostomid trap. *Mol Phylogenet Evol.* 2013; 68: 312–326. doi: [10.1016/j.ympev.2013.03.026](#) PMID: [23563272](#)
68. Bleidorn C, Podsiadlowski L, Zhong M, Eeckhaut I, Hartmann S, Halanych KM, et al. On the phylogenetic position of Myzostomida: can 77 genes get it wrong? *BMC Evol Biol.* 2009; 9: 150. doi: [10.1186/1471-2148-9-150](#) PMID: [19570199](#)
69. Siddall ME, Whiting MF. Long-branch abstractions. *Cladistics.* 1999; 15: 9–24.
70. Bergsten J. A review of long-branch attraction. *Cladistics.* 2005; 21: 163–193. doi: [10.1111/j.1096-0031.2005.00059.x](#)
71. Light JE, Siddall ME. Phylogeny of the leech family Glossiphoniidae based on mitochondrial gene sequences and morphological data. *J Parasitol.* 1999; 85: 815–823. doi: [10.2307/3285816](#) PMID: [10577715](#)
72. Siddall ME, Budinoff RB, Borda E. Phylogenetic evaluation of systematics and biogeography of the leech family Glossiphoniidae. *Invert Syst.* 2005; 19: 105–112. doi: [10.1071/IS04034](#)
73. Ocegüera-Figueroa A. Molecular phylogeny of the New World bloodfeeding leeches of the genus *Hæmenteria* and reconsideration of the biannulate genus *Oligobdella*. *Mol Phylogenet Evol.* 2012; 62: 508–514. doi: [10.1016/j.ympev.2011.10.020](#) PMID: [22100824](#)
74. Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA.* 1979; 76: 1967–1971. PMID: [109836](#)
75. Bernt M, Bleidorn C, Braband A, Dambach J, Donath A, Fritsch G, et al. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Mol Phylogenet Evol.* 2013; 69: 352–364. doi: [10.1016/j.ympev.2013.05.002](#) PMID: [23684911](#)
76. Stöger I, Schrödl M. Mitogenomics does not resolve deep molluscan relationships (yet?). *Mol Phylogenet Evol.* 2013; 69: 376–392. doi: [10.1016/j.ympev.2012.11.017](#) PMID: [23228545](#)
77. Jühling F, Pütz J, Bernt M, Donath A, Middendorf M, Florentz C, et al. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Res.* 2012; 40: 2833–2845. doi: [10.1093/nar/gkr1131](#) PMID: [22139921](#)
78. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013; 29: 2933–2935. doi: [10.1093/bioinformatics/btt509](#) PMID: [24008419](#)