# scientific reports

OPEN

# The invasive land flatworm *Arthurdendyus triangulatus* has repeated sequences in the mitogenome, extra-long cox2 gene and paralogous nuclear rRNA clusters

Romain Gastineau[1]✉, Claude Lemieux[2], Monique Turmel[2], Christian Otis[3], Brian Boyle[3], Mathieu Coulis[4,5], Clément Gouraud[6], Brian Boag[7], Archie K. Murchie[8], Leigh Winsor[9] & Jean-Lou Justine[10]

Using a combination of short- and long-reads sequencing, we were able to sequence the complete mitochondrial genome of the invasive 'New Zealand flatworm' *Arthurdendyus triangulatus* (Geoplanidae, Rhynchodeminae, Caenoplanini) and its two complete paralogous nuclear rRNA gene clusters. The mitogenome has a total length of 20,309 bp and contains repetitions that includes two types of tandem-repeats that could not be solved by short-reads sequencing. We also sequenced for the first time the mitogenomes of four species of *Caenoplana* (Caenoplanini). A maximum likelihood phylogeny associated *A. triangulatus* with the other Caenoplanini but *Parakontikia ventrolineata* and *Australopacifica atrata* were rejected from the Caenoplanini and associated instead with the Rhynchodemini, with *Platydemus manokwari*. It was found that the mitogenomes of all species of the subfamily Rhynchodeminae share several unusual structural features, including a very long *cox2* gene. This is the first time that the complete paralogous rRNA clusters, which differ in length, sequence and seemingly number of copies, were obtained for a Geoplanidae.

*Arthurdendyus triangulatus* (Dendy, 1894) is commonly referred to as the 'New Zealand flatworm', indicating its origin from the Southern Hemisphere (Fig. 1). This species of terrestrial flatworm (Geoplanidae) has earned a poor reputation as an invasive species and predator of native earthworms in north-western Europe[1]. Whether the dispersal of this species resulted from a single or several introductions is still debated[2,3]. Nevertheless, *A. triangulatus* is now well established in Great Britain and Ireland and has been recorded from as far as the remote Faroe Islands[4–9]. Because it develops well under temperate climates[10], it has the potential to disperse among several other European countries[11,12].

[1]Institute of Marine and Environmental Sciences, University of Szczecin, Szczecin, Poland. [2]Département de Biochimie, de Microbiologie et de Bio-Informatique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec, QC, Canada. [3]Plateforme d'Analyse Génomique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec, QC, Canada. [4]CIRAD, UPR GECO, 97285 Le Lamentin, Martinique, France. [5]GECO, CIRAD, University Montpellier, Montpellier, France. [6]UMR CNRS 6553 Ecobio, Université de Rennes, 263 Avenue du Gal Leclerc, CS 74205, CEDEX, 35042 Rennes, France. [7]The James Hutton Institute, Invergowrie DD2 5DA, Scotland. [8]Sustainable Agri-Food Sciences Division, Agri-Food and Biosciences Institute, Belfast BT9 5PX, Northern Ireland. [9]College of Science and Engineering, James Cook University of North Queensland, Townsville, QLD, Australia. [10]ISYEB, Institut de Systématique, Évolution, Biodiversité (UMR7205 CNRS, EPHE, MNHN, UPMC, Université des Antilles), Muséum National d'Histoire Naturelle, CP 51, 55 Rue Buffon, 75231 Paris Cedex 05, France. ✉email: romain.gastineau@usz.edu.pl

**Figure 1.** A live specimen of the invasive land flatworm *Arthurdendyus triangulatus* in a garden in Northern Ireland. Note the presence of a cocoon near the head. Unscaled—specimens typically measure 7–10 cm in length. Photograph by A. K. Murchie.

*Arthurdendyus triangulatus* is known for its predatory activity on lumbricid earthworms[9,13–16]. Given all the environmental consequences that this might have[17–20], *A. triangulatus* has been included in the European list of Invasive Alien Species of Union concern. Transport or release of live specimens of *A. triangulatus* has thus been banned in the European Union to help prevent further dispersal.

*Arthurdendyus triangulatus* is not the only species of terrestrial flatworm that has become invasive in Europe and beyond[21–38]. Invading species of terrestrial flatworms are represented by several subfamilies of the Geoplanidae, among which some belong to the Rhynchodeminae, the subfamily that includes *A. triangula tus*[25,26,30,39] within the tribe Caenoplanini. A similarly case concerns the genus *Caenoplana* Moseley, 1877 that now has species present in Europe, with probable unsuspected and underestimated biodiversity[21,40–44]. In Table 1, a summary of the currently accepted classification is provided.

Since the pioneering work by Solà et al.[48], the first to include the description of the complete mitogenome of a geoplanid, i.e., that of *Obama nungara* (Carbayo, Álvarez-Presas, Jones & Riutort, 2016), several other species have been similarly investigated phylogenomically[31,49–54]. Amongst these results, one peculiar feature that was noticed was the presence of an unusually long *cox2* gene among the three species of Rhynchodeminae studied[50–52], namely *Platydemus manokwari* de Beauchamp, 1963, *Parakontikia ventrolineata* (Dendy, 1892) and *Australopacifica atrata* (Steel, 1897). To our knowledge, this feature has not been observed in other Metazoa. The extra-length in the *cox2* sequence has no known role, and beyond the three species already studied before this work, its distribution among Rhynchodeminae was unknown[50–52].

Another peculiarity not restricted to Rhynchodeminae, but also observed in two families of the superfamily Geoplanoidea (Table 1), is the presence of two paralogous clusters encoding the nuclear rRNA genes[40,54–57]. Aside from representing a biological trait that deserves deeper studies, the existence of these divergent nuclear rRNA gene clusters may be problematic for molecular barcoding and phylogenetic analyses based on nuclear rRNA genes.

In the current study, we assembled the complete mitochondrial genome of *A. triangulatus* by using a combination of short- and long-reads sequencing technologies. Our data also enabled us to obtain for the first time the complete sequences of the two paralogous rRNA gene clusters for a geoplanid. The *A. triangulatus* mitochondrial genes were used to produce a molecular phylogeny that included four distinct species of *Caenoplana*, namely *Caenoplana variegata* (Fig. 2a), *Caenoplana coerulea* (Fig. 2b), *Caenoplana decolorata* (Fig. 2c) and *Caenoplana* sp. "brown' (Fig. 2d), for which we also sequenced mitogenomes, although the completion of these genomes will be discussed later. In addition, the mitochondrial data were used for a broad comparison of the extra sequence present in the *cox2* gene.

| Superfamily GEOPLANOIDEA |
| --- |
| Family DUGESIIDAE |
| Genus *Dugesia*—including *Dugesia constrictiva, D. japonica, D. ryukyuensis* |
| Genus *Girardia*—including *Girardia* sp. |
| Genus *Schmidtea*—including *Schmidtea mediterranea* |
| Family GEOPLANIDAE |
| Subfamily BIPALIINAE |
| Genus *Bipalium*—including *Bipalium kewense, B. vagum, B. adventitium, B. admarginatum* |
| Genus *Diversibipalium*—including *Diversibipalium multilineatum, D. mayottensis* |
| Genus *Vermiviatum*—including *Vermiviatum covidum* |
| Subfamily MICROPLANINAE |
| Genus *Microplana* |
| Subfamily RHYNCHODEMINAE |
| Tribe Rhynchodemini |
| Genus *Rhynchodemus* |
| Genus *Platydemus*—including *Platydemus manokwari* |
| Tribe Caenoplanini |
| Genus *Caenoplana*—including *Caenoplana variegata, C. coerulea, C. decolorata, Caenoplana* sp. "brown" |
| Genus *Arthurdendyus*—including *Arthurdendyus triangulatus* |
| Genus *Australopacifica*—including *Australopacifica atrata** |
| Genus *Parakontikia*—including *Parakontikia ventrolineata** |
| Tribe Anzoplanini |
| Genus *Anzoplana* |
| Tribe Eudoxiatopoplanini |
| Genus *Eudoxiatopoplana* |
| Tribe Pelmatoplanini |
| Genus *Pelmatoplana* |
| Subfamily GEOPLANINAE |
| Genus *Amaga*—including *Amaga expatria* |
| Genus *Obama*—including *Obama nungara* |
| Subfamily TIMYMINAE |
| Genus *Timyma* |

**Table 1.** A simplified outline of the classification of the superfamily Geoplanoidea. Only some genera are listed, and only species with detailed data on the mitogenome are shown; however, all subfamilies and tribes are listed. Based on the phylogeny of Sluys et al.[45] with addition of the newly described subfamily Timyminae[46]. The species formerly known as *Humbertium covidum* is mentioned here as *Vermiviatum covidum* following the recent reclassification of Bipaliinae (2023)[47]. All the species indicated are included in the multigene phylogeny presented below. *Our results contradict the inclusion of *Australopacifica* and *Parakontikia* within the Caenoplanini; both are close to *Platydemus* (Rhynchodemini).

## Results
### Assembling mitogenomes using short reads
For each of the five species examined, a large linear contig with all conserved mitochondrial genes was found following short-reads assemblies. In the case of *Caenoplana* spp., overlapping sequences were found at the ends of these contigs after their assembly or following their treatment with Consed, sometimes displaying polymorphisms or single nucleotide misalignments. Two interpretations are possible: (a) the mitogenomes of *Caenoplana* spp. are complete or (b) they might also contain several repeats at one or both of their ends, making their real sizes uncertain. In the case of *A. triangulatus*, the retrieved 15,716 bp contig after assembly showed no overlapping sequences at its ends; however, the use of the addSolexaReads.pl function of Consed in conjunction with data-mining of the contigs file led to the discovery of six small contigs that could be merged into a circular mitogenome of 18,059 bp. However, coverages of these contigs varied extensively, ranging from 66 to 282X. This suggested the presence of repetitions that cannot be resolved using short reads. As indicated in the Material and Methods, the 15,716 bp contig was later used as a database for filtering long reads.

### Processing the long-reads sequencing data
The basic statistics of the long-reads that were obtained before and after selection of the sequences specific to the mitogenome and nuclear rRNA gene clusters of *A. triangulatus* are indicated in Supplementary Table 1.

The assembly of the reads selected using the mitochondrial reference resulted in two contigs. The first one corresponded to the mitogenome; it was 20,281 bp long with a coverage of 40X and was detected by Flye as a

**Figure 2.** Live specimens of the four species of *Caenoplana* sequenced here. A, *Caenoplana variegata*, specimen MNHN JL144, hologenophore; photograph by Jean-Lou Justine. B, *Caenoplana coerulea*, specimen from the same population (same garden) as specimen MNHN JL194; photograph by Damien Michalski. C, *Caenoplana decolorata*, specimen PT426 illustrated in the original description of the species[42]; photograph by Eduardo Mateos. D, E, *Caenoplana* sp. "brown", specimens from Martinique. D, specimen MNHN JL413, E, specimen MNHN JL399; photographs by Mathieu Coulis. All photographs, unscaled; live specimens measure 5–10 cm in length.

sequence that can be circularised. The second contig was 752 bp long with a coverage of 10X and could not be identified. After the three iterations of Pilon and subsequent corrections, the final size of the mitogenome was 20,309 bp.

The assembly of the reads selected using the nuclear rRNA gene reference resulted in three contigs. The longest was 39,450 bp long with a coverage of 781X, followed by a 21,307 bp contig with a coverage of 141X. As reported below, our analysis indicated that they represent polymers of two different versions of the rDNA cluster. The last contig was 14,571 bp long with a coverage of 38X. Because Megablast queries showed that it belonged to an earthworm, probably from the genus *Eisenia* Malm, 1877, it was considered prey DNA.

### Characteristics of the *A. triangulatus* and Caenoplana spp. mitogenomes

The five mitogenomes are all colinear with sizes ranging from 16,557 to 20,309 bp in size (Table 2), but as exemplified by our analyses of the *A. triangulatus* mitogenome, the genome sizes estimated for the *Caenoplana* spp. might have been underestimated. As suggested for *O. nungara*[48] and illustrated here for *A. triangulatus*, geoplanids might display repetitions in their mitogenome that may not be resolved by short-reads sequencing. The cumulated length of all coding sequences in *A. triangulatus* is 14,336 bp, meaning that more than a quarter

| Species | Accession number | Size of the mitogenome (bp) | Size of the *cox2* encoded protein | Size of the elongation fragment in the *cox2* encoded protein |
|---|---|---|---|---|
| *Arthurdendyus triangulatus* | OR835203 | 20,309 | 446 aa | 146 aa |
| *Caenoplana variegata* JL144 | OR835205 | 16,557 | 456 aa | 148 aa |
| *Caenoplana decolorata* JL150 | OR835204 | 17,722 | 458 aa | 150 aa |
| *Caenoplana* sp. 'brown' JL410 | OR835206 | 17,236 | 448 aa | 147 aa |
| *Caenoplana coerulea* JL194 | OR835207 | 18,621 | 505 aa | 148 aa |
| *Platydemus manokwari* | MT081580 | 19,959 | 452 aa | 147 aa |
| *Parakontikia ventrolineata* | MT081960 | 17,210 | 433 aa | 142 aa |
| *Australopacifica atrata* | OM456243 | 16,513 | 434 aa | 146 aa |
| *Bipalium kewense* | MK455837 | 15,666 | 225 aa | 0 aa |
| *Obama nungara* | KP208777 | 14,909 | 259 aa | 0 aa |

**Table 2.** Accession number, size of the mitogenomes, sizes of the *cox2*-encoded protein and of their central elongation fragment for all species of Rhynchodeminae available. The last two lines show *O. nungara* (Geoplaninae) and *B. kewense* (Bipaliinae) for comparison.

of the mitogenome is constituted by non-coding DNA. Using tandem repeat finder, we identified in the large non-coding part two conserved patterns with a noticeable number of repetitions. One has a consensus size of 67 bp with 98% match and was found 9 times. The second is longer and more conserved, being 182 bp long with 99% match and also present in 9 copies (Fig. 3).

The specific features previously reported for the mitogenomes of Rhynchodeminae[52] were all found among the mitogenomes of *A. triangulatus* and *Caenoplana* spp. They all display a 32-bp overlap between *ND4L* and *ND4*, the *ND5* gene is terminated by the presence of *tRNA-Ser*, and the *cox2* gene is of unusual length. As already observed for some other Geoplanidae[31,50–52], no *tRNA-Thr* gene could be identified in the mitogenomes of the species studied here with the exception of *C. coerulea*. For this species, a D-Loop missing tRNA was found at a position congruent with other species in which this tRNA was found. As in Soo et al.[54], it was possible to find the completely conserved **TGT** anticodon of a putative *tRNA-Thr* between the 16S rRNA gene and cytochrome b genes in the other species. However such tRNA would once again have a poorly conserved structure (no cloverleaf shape and missing D- and T-loops) and therefore was not annotated as such for any of these species. It is noteworthy that for *A. triangulatus*, it was possible to find two putative *tRNA-Phe* with a cloverleaf shape. One was found in a place congruent with all other geoplanids, which is between *ND4* and *cox1*. The second was found between *tRNA-Leu* and *tRNA-Asn*, which is where *tRNA-Thr* has been found among some other species. Pending further information, this second *tRNA-Phe* was not annotated.

### The extra sequence present in cox2

The amino acid alignment of the cytochrome c oxidase subunit II proteins is presented as LOGO (Fig. 4) The alignment stops and the expansion fragment starts after a conserved 6 amino acid pattern (surrounded by a red box in Fig. 4). The less-conserved region that follows, between positions 136 and 286, is due to substantial discrepancies in lengths and sequences in the Rhynchodeminae. The alignment resumes just before the C-terminal domain of the protein (highlighted by a green box on the figure), which contains, among others, the CuA binding site. It starts with a very conserved aspartate-serine dipeptide among all Geoplanidae, with a tyrosine residue mostly conserved thereafter.

The alignment was trimmed to include only the amino-acid residues comprised between the hepta- and dipeptide described above (Fig. 5). The length of this region was calculated for each species plus *O. nungara* and *B. kewense* and the resulting values were found to be highly similar among the rhynchodemins, ranging from 142 to 150 amino acids (Table 2). Only 11 residues are conserved among all the eight species of Rhynchodeminae examined. From the N-terminal to C-terminal portion, they consist of two cysteine residues separated by two non-conserved amino-acids, a phenylalanine, an alanine, a lysine, an asparagine, a proline, a glycine, a leucine-tyrosine dipeptide and finally a lysine. It should be noted that the extra sequence in the middle region of *cox2* is not the only factor accounting for the greater length of the protein among rhynchodemins. The protein is also longer at the C-terminal part. Although this applies to all rhynchodemins, it is especially true for *C. coerulea*. There was no sign of a premature termination because of the presence of a tRNA, as opposed to the *ND5* gene for example. We are not ruling out a mistake of assembly that would have altered the canonical stop codon, but based on our software and sequencing data, we could not find evidence of this. Sequencing more specimens of *C. coerulea* should help to answer this question.

### The mitochondrial protein phylogeny

The model of evolution returned by ModelTest-NG was the MTZOA + I + G4 + F for the llikelihood maximum (ML) phylogeny. The inferred phylogenetic tree revealed very high support at most of its nodes (Fig. 6). It unambiguously associates *A. triangulatus* with *Caenoplana* spp., but clearly distinguishes this clade from the other group of rhynchodemins represented by *Pl. manokwari*, *Pa. ventrolineata* and *Au. atrata*.
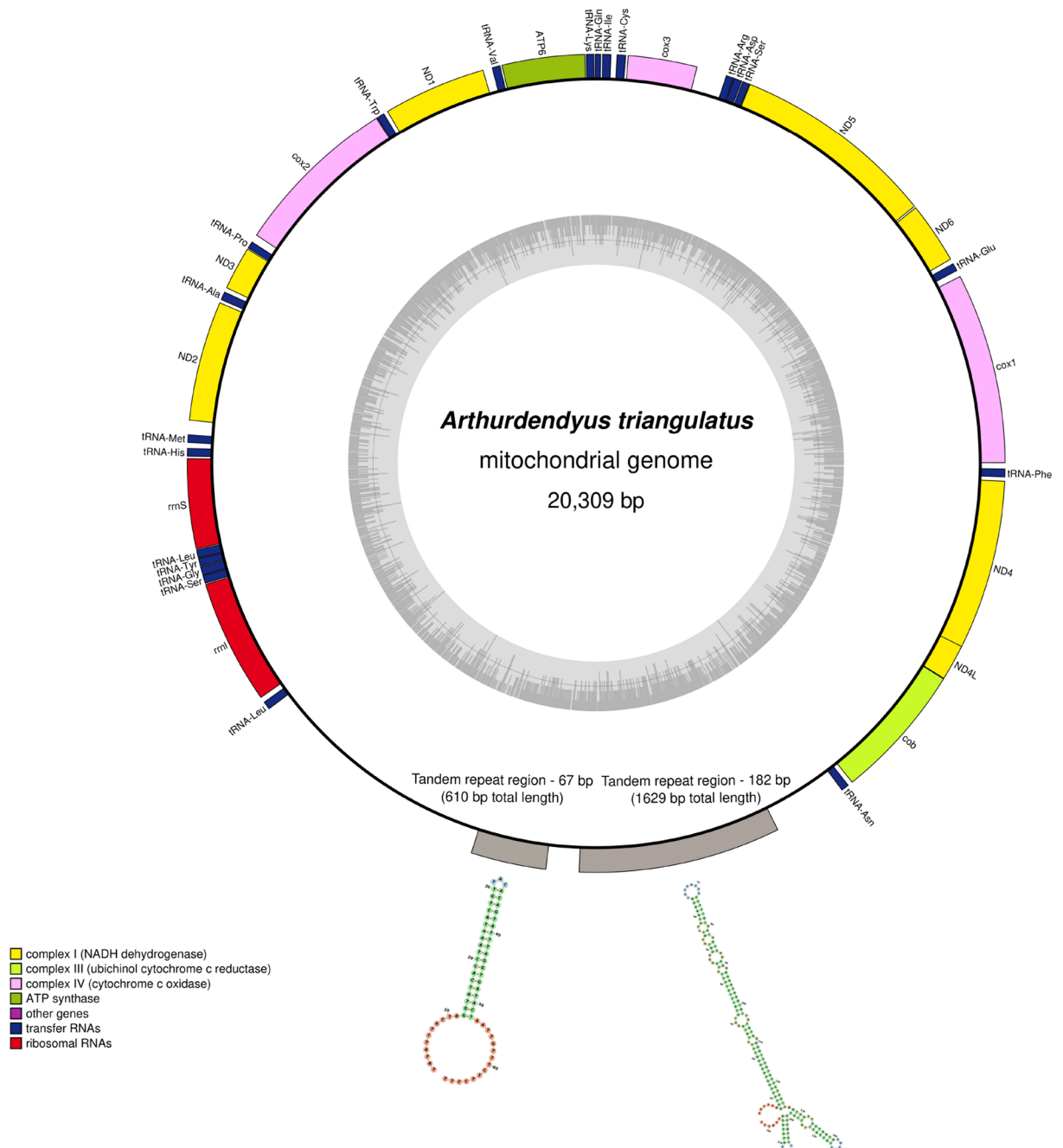
**Figure 3.** The mitogenome of *Arthurdendyus triangulatus* MNHN JL513. The colour boxes represent the different types of genes. The grey circle indicates the GC content. The grey boxes represent the extent of the two types of tandem repeats contained by the mitogenome. The position, length and secondary structure of the tandem repeats are indicated on the figure. The green colour of the dots indicates bases that could pair.

### Nuclear rRNA gene clusters

Using long-reads sequencing, it was possible to obtain the complete sequences of two paralogous nuclear rRNA gene clusters for *A. triangulatus*. Similarly to what was already suspected for *Bipalium admarginatum* de Beauchamp, 1933[54], these clusters showed different coverages after assembly (Table 2), suggesting that their numbers of copies in the nuclear genome are noticeably different. When submitted to a Megablast analysis, the 18S rRNA gene version from the 'high coverage' cluster (HCC, OR797297) was found to correspond to type II (99.77% identity with AF033044), while the 'low coverage' cluster (LCC, OR797296) corresponded to type I (99.44% identity with AF033038). The sequence identity between the HCC and LCC versions of the 18S rRNA gene was 93.68%. For comparison, an alignment between the partial 18S genes of two different species of Geoplaninae, *Obama burmeisteri* (Schultze & Muller, 1857) Carbayo et al., 2013 (DQ666004) and *Obama anthropophila* Amaral, Leal-Zanchet & Carbayo, 2015 (KP962341)[58,59] returned 96.13% identity, illustrating that the differences between two species might be lower than between the two clusters of the same species.
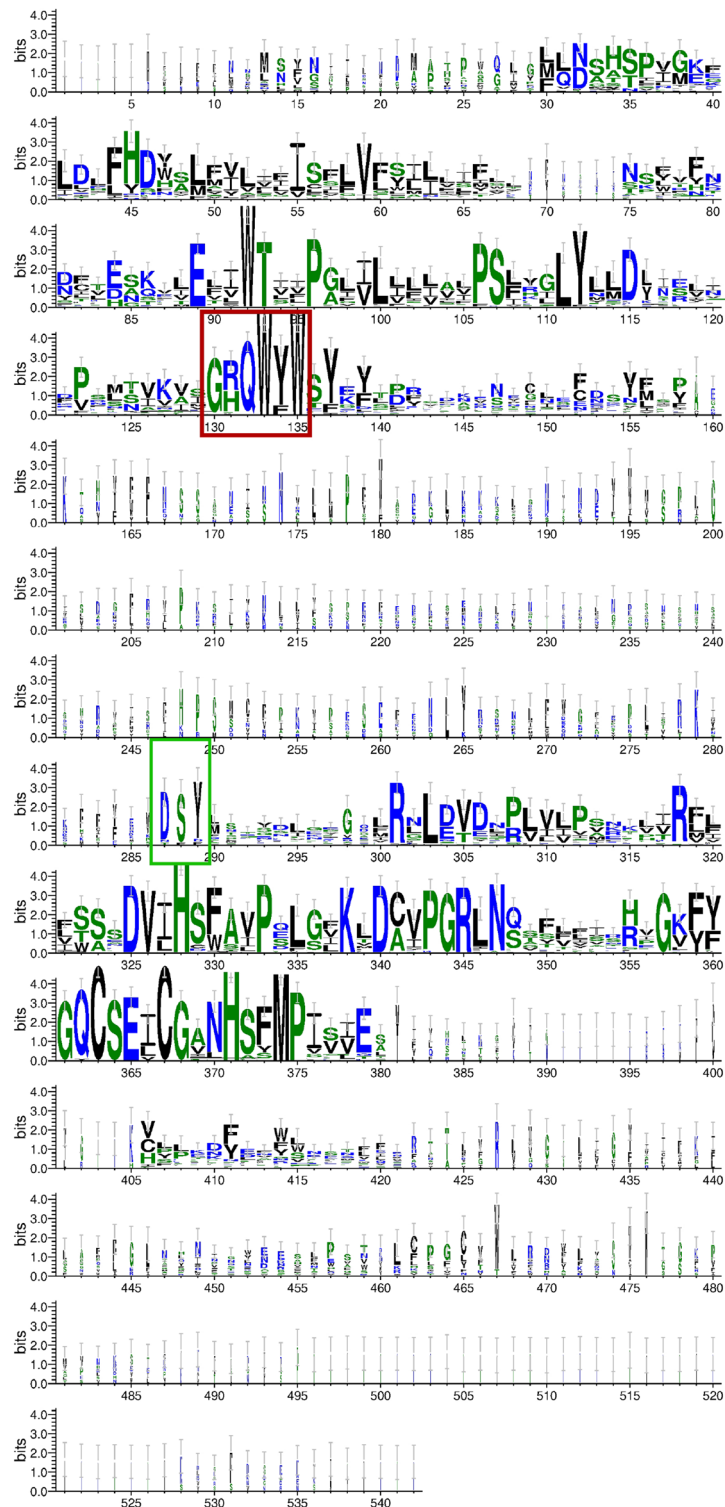
**Figure 4.** LOGO representation of the alignment of the Cox2 proteins obtained from several geoplanids and reference sequences downloaded from the Conserved Domains Database (all listed in Supplementary Table 2). The two coloured boxes indicate the conserved domains before (red box) and after (green box) the alignment breaks because of the extra-length displayed by rhynchodemins.

The two paralogous rRNA gene clusters of *A. triangulatus* also display distinct versions of the internal transcribed spacers 1 and 2 (ITS1 and ITS2), as well as distinct versions of the 5.8S rRNA gene. The sequence

**Figure 5.** The extra-length in the Cox2 protein of rhynchodemins. The conserved amino-acid found among all taxa are indicated by rectangles. Screen capture of the alignment.



**Figure 6.** Maximum likelihood phylogenetic tree obtained from concatenated alignments of mitochondrial proteins using the MTZOA + I + G4 + F model of evolution and after 1,000 ultrafast bootstrap replicates. The names of subfamilies and tribes are indicated. *Arthurdendyus triangulatus* appears in a strongly supported clade with all species of *Caenoplana*. *Australopacifica atrata* and *Parakontikia ventrolineata* are rejected from this clade and instead associated with *Platydemus manokwari*.

divergence between the ITS versions is substantial, with identities below 60%. It is also worth noting that the ITS1 size differs greatly between the two clusters (335 bp and 1207 bp for the LCC and HCC clusters, respectively). A difference in ITS length has also been observed for *Schmidtea mediterranea* Benazzi, Baguña, Ballester & del Papa, 1975[57], the only species of the superfamily Geoplanoidea for which both versions of the ITS sequences were available prior to our study. All existing ITS references for *A. triangulatus*[2,3] aligned with the HCC version of the ITS1-5.8S-ITS2 sequence. Note that the haplotypes detected in *A. triangulatus* by Roberts et al.[3] should not be mistaken for HCC and LCC, as rather, they represent inter-individual variabilities.

In addition, there are two distinct versions of the 28S gene in *A. triangulatus* (Table 3). Their sequences differ at their 3' ends. The two 28S gene sequences can align to a certain point, which corresponds to the 3' end of the version found in the LCC cluster. Beyond this point, they diverge; however, when aligned against the reference sequence from *Mus musculus* (see below in Material & Methods), the 28S gene sequence present in the HCC

|  | Accession | Coverage |  | 18S | ITS1 | 5.8S | ITS2 | 28S |
|---|---|---|---|---|---|---|---|---|
| HCC | OR797297 | 781X | Size | 1791 bp | 335 bp | 151 bp | 426 bp | 3476 bp |
| LCC | OR797296 | 141X |  | 1791 bp | 1207 bp | 153 bp | 510 bp | 3151 bp |
|  |  |  | Identity | 93.68% | 57.06% | 94.70% | 55.58% | 89.29% |

**Table 3.** Accession number and coverage obtained after assembly of the two paralogous clusters of nuclear rRNA genes in *A. triangulatus*. The size of its different components and the percentage of identity between paralogous clusters are also indicated.

cluster correctly aligns over a longer length (336 bp). To estimate the sequence identity between the two 28S genes, the gene sequence present in the HCC cluster was trimmed so that its 3' end coincides with that found in the LCC cluster. The D2 variable region is especially poorly conserved, showing only 60.43% identity.

## Discussion

With long-reads sequencing technologies becoming more widely available, it is expected that an increasing number of complex structures present in mitogenomes will be resolved, whether among vertebrates[60–62] or invertebrates[63–69].

In the current case, long-reads allowed us to resolve a nearly 5 kb long region between *rrnL* and *cob* containing two types of tandem repeats. Of particular interest for our study are the reports from Kinkar et al.[64–66] and Oey et al.[67,68]. Using long-read sequencing methods, these authors found that Platyhelminthes such as the trematode *Schistosoma haematobium* (Bilharz, 1852), *Schistosoma bovis* Sonsino, 1876, *Clonorchis sinensis* (Cobbold, 1875) and *Paragonimus westermani* (Kerbert 1878) or the cestode *Echinococcus granulosus* Batsch, 1786 all display repeated regions in their mitogenome, with the most complex and longest repeated structure found in some strains of *S. haematobium.* In the latter, the length of the mitogenome ranges from 22.6 to 33.4 kb depending on the specimen[66]. There are, however, noticeable differences when compared with *A. triangulatus.* First, the position of the repeated region is different. It is located between *ND1* and *cox1* in *S. haematobium* and *S. bovis* and between *ND5* and *cox3* in *C. sinensis.* The number of tandem-repeat regions may also differ, as is exemplified by *E. granulosus*, in which two regions were found, one large between *ND5* and *cox3* as in *C. sinensis*, and one shorter between *ND5* and *ND6.* Secondly, the motifs of tandem-repeats might alternate as in *P. westermani* and *S. bovis.* Finally, in some cases, tRNA could be found between repeated motifs (e.g. *S. bovis*), which is not the case of *A. triangulatus.*

However, the above-mentioned organisms were not the first Platyhelminthes in which extra-long mitogenomes were investigated. Similar to our study, the work of Ross et al.[69] on the 27,133 bp mitogenome of *S. mediterranea* uncovered the presence of a long non-coding region, but no mention of tandem-repeats was made in this publication. In this case also, the non-coding region was not resolved by long reads but rather by PCR amplification and Sanger sequencing. It is noteworthy that a 10-kb difference in the size of this non-coding region was observed between a sexual and an asexual specimen, the sexual specimen being the one displaying this extra length, which also contains a *tRNA-Ser* (missing from the asexual type). This order of magnitude compares with the observations done on the different specimen of *S. haematobium.* The non-coding region is located between *rrnS* and *ND2* in *S. mediterranea*, which also differs from *A. triangulatus.*

Notably, the position of the *A. triangulatus* mitochondrial non-coding region that is associated with repeated structures, corresponds to a portion of the mitogenome of the geoplanid *Diversibipalium multilineatum* (Makino & Shirasawa, 1983) which we previously failed to circularise after assembly. It was indeed impossible to find overlapping sequences at its endings[31]. The presence of complex repeated structures in *D. multilineatum* could thus not be ruled out. Generally speaking, the presence, structure, and distribution of repeated sequences in mitogenomes among Geoplanidae is an open field for investigation. At present, it remains unknown as to which taxa contain mitochondrial repeats and whether these are conserved within a species or between closely related species. In the future, we may need to investigate some of these species again with long-reads technologies to verify our previous findings.

Thanks to the expanded alignment that we performed in the course of this study (Fig. 5), we could identify which conserved residues are present in the extension segment of the *cox2*-encoded protein of Rhynchodeminae. Introns in the mitogenomes of Metazoa are rather scarce, especially among bilaterians[70], and we could not detect introns in the *cox2* gene. One of the explanations for the presence of this extra segment was that it could corresponds to an intein, i.e. a self-splicing element in the protein[71]. We therefore searched whether the conserved residues identified in the extra segment of Cox2 could represent conserved features of inteins as they are explicated in the InBase tool (https://inbase.ligsciss.com/iwai/InBase/tools.neb.com/inbase/index.html) based on the works of Perler[71,72] and Pietrokovski[73]. The only conserved residues that could be found is a cysteine in what would be the predicted N-terminal splicing region and a serine in what would be the C-extein part. Nohistidine-asparagine or histidine-glutamine dipeptide was identified at the putative C-terminal splicing site. Inteins are rather scarce among Eukaryota and to our knowledge, none has been found in mitochondrial-encoded proteins[74]. In the absence of more convincing clues regarding the identity of the extra segment in the *cox2*-encoded protein of rhynchodemins, this segment should be considered an intrinsic component of the functional protein. Deeper investigations would require a proteomic approach to look for the presence of this extension in the mature protein, which is beyond the scope of the present work. Importantly, several reports indicate that there might be unusual initiation codons in the mitogenome of geoplanids, including rhynchodemins[48,53,75]. All these peculiarities advocate for more efforts in sequencing that should use long-reads sequencing as much as possible. It should also be noted that several tribes of Rhynchodeminae have not yet been sampled, namely Eudoxiatopoplanini Winsor, 2009, Anzoplanini Winsor, 2006 and Pelmatoplanini Ogren & Kawakatsu, 1991 (Table 1). An exploratory study on these organisms to look at the conservation of the *cox2* gene extra-length would thus be of interest.

Within the framework of this study, we obtained for the first time the two complete clusters of rRNA of a geoplanid. Several questions remain unanswered, and in some cases, routine protocols might be re-evaluated. The origin of these duplications remains unknown, and it is difficult to understand why both variants have been conserved. It is also unclear whether or not both clusters would be expressed and transcribed into functional rRNA. In their first publication on the topic, Carranza et al.[55] had positive results only for the expression of the type I rRNA in *S. mediterranea.* However, in their second article on *Schmidtea polychroa* (Schmidt, 1861)[57], they saw that both types might be expressed, although at very low levels for type II.

As noted above, the LCC rDNA cluster of *A. triangulatus* corresponds to type I while the HCC cluster corresponds to type II when comparing with results previously obtained on *S. polychroa* by Carranza et al.[56]. If the results of Carranza et al.[57] on the expression of these two types were extrapolated to *A. triangulatus,* this would mean that the type associated with the highest coverage (thus, the highest number of copies) would be the least expressed, which is rather counter-intuitive. Technologies like RNAseq could be used to compare the coverage of both rDNA clusters in the genome with the coverages of the RNA they encode.

There are direct consequences of our new findings regarding the use of nuclear rRNA genes for barcoding and phylogenetic inference among Geoplanoidea. One can predict that the HCC/type II cluster has statistically more chance to be amplified and sequenced than the LCC/type I cluster. This would mean that the least expressed and possibly non-functional type would likely be the one amplified.

In case both variants were independently amplified on two specimens of the same species for which no reference is available, this would definitely be an issue in terms of molecular barcoding. This would be the case especially for the D2 region of the *28S* gene, for which there is substantial literature on a wide range of highly diverse Eukaryota[76–81]. Using this marker poses potential problems with the Geoplanidae as exemplified by the very low 60.43% identity between both variants in *A. triangulatus*. It could introduce a strong bias in any inferred phylogeny or lead to inaccurate taxonomic assessment when using molecular barcoding.

Another important issue raised by these paralogous clusters would be the use of the *18S* gene in the early detection of invasive flatworms in soil by the means of environmental DNA and metabarcoding. Such methods are often conducted on other Eukaryota by amplifying the V4 and V9 variable parts of the *18S* gene. With geoplanids, the protocol would preferably be adapted, or different barcodes (eg. the *cox1* gene) used.

The increasing availability of long-reads DNA sequencing technologies will make it possible to study paralogous rRNA gene clusters on more species of Geoplanidae and with larger sample sizes. With more complete sequences of the two types of nuclear rRNA gene clusters, their rate of evolution could be analysed. In addition, long-reads DNA sequencing of additional Geoplanidae would advance our knowledge about the distribution of repeats in the mitogenome. We hope to be able to go further in this direction in the near future.

The protocol used for phylogeny (concatenated amino-acid sequences) once again returned robust results. Based on the tree presented here, the inclusion of *Pa. ventrolineata* and *Au. atrata* in the Caenoplanini is not supported by this phylogenetic analysis, corroborating the results previously obtained by other teams[21] from partial *cox1* and 28S genes. Instead, both species are associated with maximum support to the Rhynchodemini (*Pl. manokwari*). As with previous phylogenies[31], Geoplaninae and Bipaliinae appear as distinct, highly supported clusters. As already stated, many tribes of Rhynchodeminae remain to be sampled (Table 1), and in some cases (e.g. the genus *Anzoplana* Winsor, 2006), there is currently not a single sequence available in GenBank. How the results of such an upgraded phylogeny would articulate with the morphological classification is an exciting question we hope to address soon.

## Material and methods
### Biological material
The origins of the specimens used in the course of this study are reported below. All specimens were registered in the collections of the Muséum National d'Histoire Naturelle in Paris, France. All were killed by immersion in hot water or 95% ethanol.

*Arthurdendyus triangulatus*: five specimens collected on July 12, 2022, by Brian Boag; Birch Brae, Knapp, Inchture, Perthshire, PH14 9RN, Scotland; coordinates: N 56.47005205158123, W -3.1614500498816174. One specimen used for molecular analysis; four specimens deposited in MNHN under registration number MNHN JL513 (Fig. 1).

*Caenoplana variegata* (Fletcher & Hamilton, 1888): two specimens collected on May 6, 2014, by Dhyma Gomez; La-Plaine-Saint-Denis, Seine Saint Denis, Metropolitan France. Specimens kept in MNHN under registration number MNHN JL144, portion of body used for molecular analysis. The specimen in Fig. 2A is the hologenophore.

*Caenoplana coerulea* Moseley, 1877: one specimen collected on November 7, 2014, by Damien Michalski; Arles, Bouches-du-Rhône, Metropolitan France. Specimen deposited in MNHN under registration number MNHN JL194, portion of body used for molecular analysis. Note that Álvarez-Presas et al.[21] have emphasized that *C. coerulea* is a species complex, based on their molecular work and information from one of us (LW). The *cox1* gene of our specimen is 100% identical to several sequences in GenBank which were attributed to *C. coerulea* sensu lato morphotype Ca1[21]. The specimen photographed in Fig. 2B is from the same population (same garden) as the hologenophore.

*Caenoplana decolorata* Mateos, Jones, Riutort, & Álvarez-Presas, 2020: one specimen collected on May 2, 2014, by Clément Gouraud; Nantes, Loire, Metropolitan France. Specimen deposited in MNHN under registration number MNHN JL150, portion of body used for molecular analysis. The species identification was confirmed on the basis of the *cox1* gene (GenBank: MW203125)[82]. Specimen not photographed. The specimen in Fig. 2C is specimen PT426 illustrated in the original description of the species[42].

*Caenoplana* sp. "brown": three specimens collected February 26, 2019, by Mathieu Coulis; Le Lamentin, Martinique, French West Indies; coordinates N 1420001, W-6012222. Specimens deposited in MNHN under registration number MNHN JL410, portion of body of one specimen used for molecular analysis. This species is currently unnamed but has been called "Brown-striped flatworm" in the 2020 Reference of Wildlife of tropical North Queensland; there are records of its presence in Martinique, Florida (USA), and Australia (Queensland); the species is believed to be a native of New Caledonia or New Guinea. The specimens illustrated in Fig. 2D,E are from the same locality and are deposited in the MNHN as MNHN JL399 and JL413, respectively.

### Short-reads sequencing and assembly

*Arthurdendyus triangulatus* was sequenced by the Genomic Analysis Platform (PAG) of the Institute of Integrative Biology and Systems at Laval University (Quebec, Canada) (https://www.ibis.ulaval.ca/en/services-2/genomic-analysis-platform/). In order to minimize contaminations from the digestive tract, tissues from longitudinal regions were separated with a scalpel. After having been frozen in liquid nitrogen, tissues were first shredded using a vibro-grinding device MM 400 (Retsch) and cells were then transferred in an Eppendorf tube containing 1.0 mL of lysis buffer prepared with 50 mM Tris–HCl pH 8.0, 200 mM NaCl, 20 mM EDTA, 2.0% SDS and 20 mg/mL proteinase K. The latter mixture was incubated at 65 °C for 30 min. An equal volume of CTAB buffer containing 50 mM Tris–HCl pH 8.0, 1.4 M NaCl, 20 mM EDTA, 2.0% CTAB, 1.0% PVP 40,000 was added to the lysate and incubation was pursued for an additional 30 min at 65 °C. The suspension was cooled down for a few minutes before 5 µL of RNase A (100 mg/mL) were added; it was incubated at room temperature for 20 min and then split in two tubes, the contents of which were extracted twice with an equal volume of chloroform: isoamyl alcohol (24:1). Finally, DNA was precipitated with two volumes of EtOH, dried and dissolved in 100 µL of TE buffer (10 mM Tris–HCl pH 8.0, 0.1 mM EDTA). A total amount of 20.4 µg DNA was recovered. The distribution of the size of fragments in the DNA preparation was determined using the Femto Pulse from Agilent (Santa Clara, CA, USA). The library was produced using 500 ng of DNA broken with a Covaris M220 (Covaris, Woburn MA, USA) and the NEBNext Ultra II DNA Library Prep Kit Illumina from New England Biolab (Ipswich, MA, USA). A total amount of ca. 40 million clean 150 bp paired-end reads was obtained from the NovaSeq 6000 platform of Génome Québec (https://www.genomequebec.com/).

The four specimens of *Caenoplana* spp. were sequenced at the Beijing Genomics Institute (BGI) (Shenzhen, China) on a DNBSEQ-G400 platform. Tissues were sent in 95% ethanol and DNA was extracted at the BGI facilities following an internal protocol. For *C. coerulea* and *C. variegata*, 60 million clean 100 bp paired-end reads were obtained per specimen. For *C. decolorata* and *Caenoplana* sp. 'brown', 40 million clean 150 bp paired-end reads were obtained per specimen.

For all five species, short reads were assembled using SPAdes 3.15.5[83] and a k-mer of 85 for the 100 bp reads and a k-mer of 125 for the 150 bp reads. Consed[84] was used to verify the terminal endings of the linear contigs corresponding to the mitogenome by using its 'addSolexaReads.pl' script.

### Long-reads sequencing and assembly

Long-reads sequencing of the *A. triangulatus* DNA preparation was performed at the PAG using the Oxford Nanopore Technology. First, 3 µg of genomic DNA were treated with the PacBio Short Read Eliminator (SRE_XS) Kit (Circulomics/PacBio, Menlo Park, CA, USA). A DNA library was then prepared using the SQK-LSK-109 Kit from Oxford Nanopore Technology (Oxford Nanopore, Littlemore, UK) and a fraction containing 700 ng of DNA was loaded onto a R9.4 MinION cell that had 1438 active pores. After 24 h of sequencing, the cell was washed with a nuclease solution, loaded with the remaining ca. 500 ng of DNA library, and sequencing was resumed for a total time of 72 h. Basic statistics of long reads were obtained from NanoStat[85].

The reads presumably assigned to the mitogenome and to the nuclear rRNA gene clusters were selected using the mtblaster.py script (https://github.com/nidafra92/squirrel-project/blob/master/mtblaster.py). For the mitogenome, the reference sequence for this search was the contig containing all the conserved mitochondrial genes that was assembled from short reads. The Blast-based parameters were 90% identity and $1e^{-150}$ evalue, with maximum size of 35 kb. For the rRNA gene clusters, the reference consisted of the partial sequences of the 18S (AF033038) and 28S (DQ665953) rRNA genes of *A. triangulatus*, and the filtration parameters were also 90% identity and $1e^{-150}$ evalue, with a maximal size of 35 kb. The resulting sets of selected reads were assembled using Flye 2.9.1[86] with the—meta option and overlap parameter of 3000 for the mitogenome and 10,000 for the rRNA gene clusters. In the case of the mitogenome, the assembly was submitted to three iterations of Pilon 1.24[87] using the pool of short reads previously obtained.

### Annotation of mitogenomes

All mitochondrial genes were annotated with the help of MITOS[88] followed by manual curation, using the genetic code 9, except for the rRNA genes whose termini were mapped using alignments against published homologs. Positions of the tRNA genes were verified with Arwen v.1.2[89]. Repeats in the *A. triangulatus* mitogenome were analysed using Tandem repeats finder[90] and Microsatellite repeats finder (http://insilico.ehu.es/mini_tools/microsatellites/?info). Tandem repeats were drawn as explained in Kinkar et al.[65]. Mitogenome maps were drawn with OGDRAW[91].

### Annotation of nuclear rRNA gene clusters

Boundaries of the two rRNA gene clusters of *A. triangulatus* were determined using Rfam[92]. In the case of the 28S rRNA gene, alignments with the reference sequence of *Mus musculus* Linnaeus, 1758 (NR_003279)[93] were required.

### Alignment of Cox2 proteins

The amino-acid sequences of the predicted proteins encoded by the *cox2* genes of *A. triangulatus* and *Caenoplana* spp. were aligned with the corresponding sequences of other species of Geoplanidae, Platyhelminthes, and reference sequences from the Conserved Domains Database (https://www.ncbi.nlm.nih.gov/cdd). All accession numbers are listed in Supplementary Table 2. The alignment was performed with MEGA11[94]. LOGO alignment was done on the online LOGO website (https://weblogo.threeplusone.com/).

## Phylogenetic analysis

The dataset previously used to infer a phylogeny of the Geoplanidae[31,54] based on 21 mitogenome-encoded proteins was appended with the five new species examined here plus the recently described *Dugesia constrictiva* Chen & Dong, sp. nov.[95]. The amino-acid sequences of the individual proteins were first aligned using MAFFT 7[96] and trimmed with the -automated1 option of trimAl[97]; then, the different protein alignments were concatenated using Phyutility 2.7.1[98]. ModelTest-NG v0.1.7[99] was used to select the best model of evolution, with default option for maximum likelihood inference (ML). The ML phylogenetic analysis was performed using IQ-TREE 2.2.0[100] and 1000 ultrafast bootstrap replicates.

## Data availability

## References

1. Jones, H. D. A new genus and species of terrestrial planarian (Platyhelminthes; Tricladida; Terricola) from Scotland, and an emendation of the genus *Artioposthia. J. Nat. Hist.* **33**, 387–394. https://doi.org/10.1080/002229399300308 (1999).
2. Dynes, C., Fleming, C. C. & Murchie, A. K. Genetic variation in native and introduced populations of the 'New Zealand flatworm'. *Arthurdendyus triangulatus. Ann Appl Biol.* **139**, 165–174. https://doi.org/10.1111/j.1744-7348.2001.tb00393.x (2005).
3. Roberts, D. M. *et al.* Genetic variability of *Arthurdendyus triangulatus* (Dendy, 1894), a non-native invasive land planarian. *Zootaxa* https://doi.org/10.11646/zootaxa.4808.1.2 (2020).
4. Willis, R. J. & Edwards, A. R. The occurrence of the land planarian *Artioposthia triangulata* (Dendy) in Northern Ireland. *Ir. Nat. J.* **19**, 112–116 (1977).
5. Blackshaw, R. P. & Stewart, V. I. Artioposthia triangulata (Dendy 1894, a predatory terrestrial planarian and its potential impact on lumbricid earthworms. *Agric. Zool. Rev.* **5**, 201–219 (1992).
6. Moore, J. P., Dynes, C. & Murchie, A. K. Status and public perception of the 'New Zealand flatworm', *Artioposthia triangulata* (Dendy), in Northern Ireland. *Pedobiologia* **42**, 563–571 (1998).
7. Cannon, R. J. C., Baker, R. H. A., Taylor, M. C. & Moore, J. P. A review of the status of the New Zealand flatworm in the UK. *Ann. Appl. Biol.* **135**, 597–614. https://doi.org/10.1111/j.1744-7348.1999.tb00892.x (1999).
8. Bloch, D. A note on the occurrence of land planarians in the Faroe Islands. *Fródskaparrit* **38**, 63–68 (1992).
9. Mather, J. G. & Christensen, O. M. The exotic land planarian *Artioposthia triangulata* in the Faroe Islands: Colonisation and habitats. *Fródskaparrit* **40**, 49–60 (1992).
10. Christensen, O. M. & Mather, J. G. Long-term study of growth in the New Zealand flatworm *Arthurdendyus triangulatus* under laboratory conditions. *Pedobiologia* **45**, 535–549. https://doi.org/10.1078/0031-4056-00105 (2001).
11. Boag, B. *et al.* The potential spread of terrestrial planarians *Artioposthia triangulata* and *Australoplana sanguinea var. alba* to continental Europe. *Ann Appl Biol.* **127**, 385–390. https://doi.org/10.1111/j.1744-7348.1995.tb06682.x (1995).
12. Boag, B., Evans, K. A., Yeates, G. W., Johns, P. M. & Neilson, R. Assessment of the global potential distribution of the predatory land planarian *Artioposthia triangulata* (Dendy) (Tricladida, Terricola) from ecoclimatic data. *New Zeal. J. Zool.* **22**, 311–318. https://doi.org/10.1080/03014223.1995.9518046 (1995).
13. Blackshaw, R. P. Studies on *Artioposthia triangulata* (Dendy) (Tricladida, Terricola), a predator of earthworms. *Ann. Appl. Biol.* **116**, 169–176. https://doi.org/10.1111/j.1744-7348.1990.tb06596.x (1990).
14. Murchie, A. & Gordon, A. The impact of the 'New Zealand flatworm', *Arthurdendyus triangulatus*, on earthworm populations in the field. *Biol. Invasions* **15**, 569–586. https://doi.org/10.1007/s10530-012-0309-7 (2013).
15. Christensen, O. M. & Mather, J. G. Colonisation by the land planarian *Artioposthia triangulata* and impact on lumbricid earthworms at a horticultural site. *Pedobiologia* **39**, 144–154 (1995).
16. Fraser, P. M. & Boag, B. The distribution of lumbricid earthworm communities in relation to flatworms: A comparison between New Zealand and Europe. *Pedobiologia* **42**, 542–553 (1998).
17. Haria, A. H. Hydrological and environmental impact of earthworm depletion by the New Zealand flatworm (*Artioposthia triangulata*). *J. Hydrol.* **171**, 1–3. https://doi.org/10.1016/0022-1694(95)02734-7 (1995).
18. Haria, A. H., McGrath, S. P., Moore, J. P., Bell, J. P. & Blackshaw, R. P. Impact of the New Zealand flatworm (*Artioposthia triangulata*) on soil structure and hydrology in the UK. *Sci. Total Environ.* **215**, 259–265. https://doi.org/10.1016/s0048-9697(98)00126-0 (1998).
19. Alford, D. V. Potential problems posed by non-indigenous terrestrial flatworms in the United Kingdom. *Pedobiologia* **42**, 574–578 (1998).
20. Boag, B. The impact of the New Zealand flatworm on earthworms and moles in agricultural land in western Scotland. *Asp. Appl. Biol.* **62**, 79–84 (2000).
21. Álvarez-Presas, M., Mateos, E., Tudó, À., Jones, H. & Riutort, M. Diversity of introduced terrestrial flatworms in the Iberian Peninsula: A cautionary tale. *PeerJ* **2**, e430. https://doi.org/10.7717/peerj.430 (2014).
22. Carbayo, F., Álvarez-Presas, M., Jones, H. D. & Riutort, M. The true identity of *Obama* (Platyhelminthes: Geoplanidae) flatworm spreading across Europe. *Zool. J. Linn. Soc.-Lond.* **177**, 5–28. https://doi.org/10.1111/zoj.12358 (2016).
23. Fourcade, Y. Fine-tuning niche models matters in invasion ecology. A lesson from the land planarian Obama nungara. *Ecol. Model.* **457**, e109686. https://doi.org/10.1016/j.ecolmodel.2021.109686 (2021).
24. Fourcade, Y., Winsor, L. & Justine, J.-L. Hammerhead worms everywhere? Modelling the invasion of bipaliin flatworms in a changing climate. *Divers. Distrib.* **28**, 844–858. https://doi.org/10.1111/ddi.13489 (2022).
25. Jones, H. D. & Sluys, R. A new terrestrial planarian species of the genus *Marionfyfea* (Platyhelminthes: Tricladida) found in Europe. *J. Nat. Hist.* **50**, 2673–2690. https://doi.org/10.1080/00222933.2016.1208907 (2016).
26. Jones, H. D. Another alien terrestrial planarian in the United Kingdom: *Australopacifica atrata* (Steel, 1897) (Platyhelminthes: Tricladida: Continenticola). *Zootaxa* https://doi.org/10.11646/zootaxa.4604.3.12 (2019).
27. Justine, J. L. *et al.* The invasive land planarian *Platydemus manokwari* (Platyhelminthes, Geoplanidae): Records from six new localities, including the first in the USA. *PeerJ* **3**, e1037. https://doi.org/10.7717/peerj.1037 (2015).
28. Justine, J. L., Winsor, L., Gey, D., Gros, P. & Thévenot, J. Giant worms chez moi! Hammerhead flatworms (Platyhelminthes, Geoplanidae, *Bipalium* spp., *Diversibipalium* spp.) in metropolitan France and overseas French territories. *PeerJ* **6**, e4672. https://doi.org/10.7717/peerj.4672 (2018).

29. Justine, J. L., Winsor, L., Gey, D., Gros, P. & Thévenot, J. *Obama chez moi*! The invasion of metropolitan France by the land planarian *Obama nungara* (Platyhelminthes, Geoplanidae). *PeerJ* **8**, e8385. https://doi.org/10.7717/peerj.8385 (2020).

30. Justine, J. L. *et al.* Presence of the invasive land flatworm *Platydemus manokwari* (Platyhelminthes, Geoplanidae) in Guadeloupe, Martinique and Saint Martin (French West Indies). *Zootaxa* **4951**, 381–390. https://doi.org/10.11646/zootaxa.4951.2.11 (2021).

31. Justine, J. L. *et al.* Hammerhead flatworms (Platyhelminthes, Geoplanidae, Bipaliinae): Mitochondrial genomes and description of two new species from France, Italy, and Mayotte. *PeerJ* **10**, e12725. https://doi.org/10.7717/peerj.12725 (2022).

32. Justine, J. L., Marie, A. D., Gastineau, R., Fourcade, Y. & Winsor, L. The invasive land flatworm *Obama nungara* in La Reunion, a French island in the Indian Ocean, the first report of the species for Africa. *Zootaxa* **5154**, 469–476. https://doi.org/10.11646/zootaxa.5154.4.4 (2022).

33. Lago-Barcia, D. *et al.* Morphology and DNA barcodes reveal the presence of the non-native land planarian *Obama marmorata* (Platyhelminthes : Geoplanidae) in Europe. *Invertebr. Syst.* **29**, 12–22. https://doi.org/10.1071/IS14033 (2015).

34. Lago-Barcia, D. *et al.* Reconstructing routes of invasion of *Obama nungara* (Platyhelminthes: Tricladida) in the Iberian Peninsula. *Biol. Invasions* **21**, 289–302. https://doi.org/10.1007/s10530-018-1834-9 (2019).

35. Mazza, G. *et al.* First report of the land planarian *Diversibipalium multilineatum* (Makino & Shirasawa, 1983) (Platyhelminthes, Tricladida, Continenticola) in Europe. *Zootaxa* **4067**, 577–580. https://doi.org/10.11646/zootaxa.4067.5.4 (2016).

36. Mori, E. *et al.* Discovering the Pandora's box: The invasion of alien flatworms in Italy. *Biol. Invasions* **24**, 205–216. https://doi.org/10.1007/s10530-021-02638-w (2022).

37. Negrete, L., Lenguas Francavilla, M., Damborenea, C. & Brusa, F. Trying to take over the world: Potential distribution of *Obama nungara* (Platyhelminthes: Geoplanidae), the Neotropical land planarian that has reached Europe. *Glob. Change Biol.* **26**, 4907–4918. https://doi.org/10.1111/gcb.15208 (2020).

38. Roy, V. *et al.* Gut content metabarcoding and citizen science reveal the earthworm prey of the exotic terrestrial flatworm, *Obama nungara*. *Eur. J. Soil Biol.* **113**, 103449. https://doi.org/10.1016/j.ejsobi.2022.103449 (2022).

39. Hu, J., Yang, M., Ye, E. R., Ye, Y. & Niu, Y. First record of the New Guinea flatworm *Platydemus manokwari* (Platyhelminthes, Geoplanidae) as an alien species in Hong Kong Island, China. *ZooKeys* **873**, 1–7. https://doi.org/10.3897/zookeys.873.36458 (2019).

40. Breugelmans, K., Quintana Cardona, J., Artois, T., Jordaens, K. & Backeljau, T. First report of the exotic blue land planarian, *Caenoplana coerulea* (Platyhelminthes, Geoplanidae), on Menorca (Balearic Islands, Spain). *ZooKeys* **199**, 91–105. https://doi.org/10.3897/zookeys.199.3215 (2012).

41. Jones, H. D., Mateos, E., Riutort, M. & Álvarez-Presas, M. The identity of the invasive yellow-striped terrestrial planarian found recently in Europe: *Caenoplana variegata* (Fletcher amp; Hamilton, 1888) or *Caenoplana bicolor* (Graff, 1899)?. *Zootaxa* https://doi.org/10.11646/zootaxa.4731.2.2 (2020).

42. Mateos, E., Jones, H. D., Riutort, M. & Álvarez-Presas, M. A new species of alien terrestrial planarian in Spain: *Caenoplana decolorata*. *PeerJ* **8**, e10013. https://doi.org/10.7717/peerj.10013 (2020).

43. Suárez, D., Martín, S. & Naranjo, M. First report of the invasive alien species *Caenoplana coerulea* Moseley, 1877 (Platyhelminthes, Tricladida, Geoplanidae) in the subterranean environment of the Canary Islands. *Subterr. Biol.* **26**, 67–74. https://doi.org/10.3897/subtbiol.26.25921 (2018).

44. Suárez, D., Pedrianes, J. R. & Andjar, C. DNA barcoding reveals new records of invasive terrestrial flatworms (Platyhelminthes, Tricladida, Geoplanidae) in the Macaronesian region. *Zootaxa* **5129**, 447–450. https://doi.org/10.11646/zootaxa.5129.3.9 (2022).

45. Sluys, R., Kawakatsu, M., Riutort, M. & Baguñà, J. A new higher classification of planarian flatworms (Platyhelminthes, Tricladida). *J. Natl. Hist.* **43**, 1763–1777. https://doi.org/10.1080/00222930902741669 (2009).

46. Almeida, A. L., Francoy, T. M., Álvarez-Presas, M. & Carbayo, F. Convergent evolution: A new subfamily for bipaliin-like Chilean land planarians (platyhelminthes). *Zool. Scr.* **50**, 500–508. https://doi.org/10.1111/zsc.12479 (2021).

47. Solà, E., Sluys, R., Riutort, M. & Kawakatsu, M. Molecular phylogenetics facilitates the first historical biogeographic analysis of the hammerhead worms (Platyhelminthes: Tricladida: Bipaliinae), with the description of twelve new species and two new genera. *Zootaxa* **5335**, 1–77. https://doi.org/10.11646/zootaxa.5335.1.1 (2023).

48. Solà, E. *et al.* Evolutionary analysis of mitogenomes from parasitic and free-living flatworms. *PLoS ONE* **10**(3), e0120081. https://doi.org/10.1371/journal.pone.0120081 (2015).

49. Gastineau, R., Justine, J. L., Lemieux, C., Turmel, M. & Witkowski, A. Complete mitogenome of the giant invasive hammerhead flatworm *Bipalium kewense*. *Mitochond DNA Part B Resources* **4**, 1343–1344. https://doi.org/10.1080/23802359.2019.1596768 (2019).

50. Gastineau, R., Lemieux, C., Turmel, M. & Justine, J. L. Complete mitogenome of the invasive land flatworm *Platydemus manokwari*. *Mitochond DNA Part B Resources* **5**, 1689–1690. https://doi.org/10.1080/23802359.2020.1748532 (2020).

51. Gastineau, R., Winsor, L. & Justine, J. L. The complete mitogenome of the potentially invasive flatworm *Australopacifica atrata* (Platyhelminthes, Geoplanidae) displays unusual features common to other Rhynchodeminae. *ZooKeys* **1110**, 121–133. https://doi.org/10.3897/zookeys.1110.83228 (2022).

52. Gastineau, R. & Justine, J. L. Complete mitogenome of the invasive land flatworm *Parakontikia ventrolineata*, the second Geoplanidae (Platyhelminthes) to display an unusually long cox2 gene. *Mitochond DNA Part B Resources* **5**, 2115–2116. https://doi.org/10.1080/23802359.2020.1765709 (2020).

53. Justine, J. L., Gey, D., Thévenot, J., Gastineau, R. & Jones, H. D. The land flatworm *Amaga expatria* (Geoplanidae) in Guadeloupe and Martinique: New reports and molecular characterization including complete mitogenome. *PeerJ* **8**, e10098. https://doi.org/10.7717/peerj.10098 (2020).

54. Soo, O. Y. M., Gastineau, R., Verdon, G., Winsor, L. & Justine, J. L. Rediscovery of *Bipalium admarginatum* de Beauchamp, 1933 (Platyhelminthes, Tricladida, Geoplanidae) in Malaysia, with molecular characterisation including the mitogenome. *Zootaxa* **5277**, 585–599. https://doi.org/10.11646/zootaxa.5277.3.11 (2023).

55. Carranza, S., Giribet, G., Ribera, C., Baguñà, & Riutort, M. Evidence that two types of 18S rDNA coexist in the genome of *Dugesia* (*Schmidtea*) *mediterranea* (Platyhelminthes, Turbellaria, Tricladida). *Mol. Biol. Evol.* **13**, 824–832. https://doi.org/10.1093/oxfordjournals.molbev.a025643 (1996).

56. Carranza, S. *et al.* A robust molecular phylogeny of the Tricladida (Platyhelminthes: Seriata) with a discussion on morphological synapomorphies. *Proc. R. Soc. B-Biol. Sci.* **265**, 631–640. https://doi.org/10.1098/rspb.1998.0341 (1998).

57. Carranza, S., Baguñà, J. & Riutort, M. Origin and evolution of paralogous rRNA gene clusters within the flatworm family Dugesiidae (Platyhelminthes, Tricladida). *J. Mol. Evol.* **49**, 250–259. https://doi.org/10.1007/pl00006547 (1999).

58. Alvarez-Presas, M., Baguñà, J. & Riutort, M. Molecular phylogeny of land and freshwater planarians (Tricladida, Platyhelminthes): From freshwater to land and back. *Mol. Phylogenet. Evol.* **47**, 555–568. https://doi.org/10.1016/j.ympev.2008.01.032 (2008).

59. Álvarez-Presas, M., Amaral, S. V., Carbayo, F., Leal-Zanchet, A. M. & Riutort, M. Focus on the details: Morphological evidence supports new cryptic land flatworm (Platyhelminthes) species revealed with molecules. *Org. Divers. Evol.* **15**, 379–403. https://doi.org/10.1007/s13127-014-0197-z (2015).

60. Gastineau, R. *et al.* The mitochondrial genome of the bioluminescent fish *Malacosteus niger* Ayres, 1848 (Stomiidae, Actinopterygii) is large and complex, and contains an inverted-repeat structure. *ZooKeys* **1157**, 177–191. https://doi.org/10.3897/zookeys.1157.97921 (2023).

61. Macey, J. R. *et al.* Evidence of two deeply divergent co-existing mitochondrial genomes in the Tuatara reveals an extremely complex genomic organization. *Commun Biol.* **4**, e116. https://doi.org/10.1038/s42003-020-01639-0 (2021).

62. Pinninti, L. R. *et al.* The complete mitochondrial genome of the Atlantic spiny lumpsucker *Eumicrotremus spinosus* (Fabricius, 1776). *Mitochond DNA Part B Resources* **8**, 364–367. https://doi.org/10.1080/23802359.2023.2184649 (2023).

63. De Vivo, M. *et al.* Utilisation of Oxford Nanopore sequencing to generate six complete gastropod mitochondrial genomes as part of a biodiversity curriculum. *Sci Rep.* **12**, e9973. https://doi.org/10.1038/s41598-022-14121-0 (2022).

64. Kinkar, L. *et al.* Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1. *Parasite Vector.* **12**, 1–7. https://doi.org/10.1186/s13071-019-3492-x (2019).

65. Kinkar, L. *et al.* First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. *PLOS Neglect. Trop. Dis.* **14**, e0008552. https://doi.org/10.1371/journal.pntd.0008552 (2020).

66. Kinkar, L. *et al.* Nanopore sequencing resolves elusive long tandem-repeat regions in mitochondrial genomes. *Int. J. Mol. Sci.* **22**, 1811. https://doi.org/10.3390/ijms22041811 (2021).

67. Oey, H. *et al.* Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*. *GigaScience* **8**, giy146. https://doi.org/10.1093/gigascience/giy146 (2019).

68. Oey, H. *et al.* Whole-genome sequence of the bovine blood fluke Schistosoma bovis supports interspecific hybridization with *S. haematobium*. *PLoS Pathog.* **15**, e1007513. https://doi.org/10.1371/journal.ppat.1007513 (2019).

69. Ross, E., Blair, D., Guerrero-Hernández, C. & SánchezAlvarado, A. Comparative and transcriptome analyses uncover key aspects of coding- and long noncoding RNAs in flatworm mitochondrial genomes. *G3* **6**, 1191–1200. https://doi.org/10.1534/g3.116.028175 (2016).

70. Jenkins, H. L. *et al.* Unprecedented frequency of mitochondrial introns in colonial bilaterians. *Sci. Rep.* **12**, 10889. https://doi.org/10.1038/s41598-022-14477-3 (2022).

71. Perler, F. B. *et al.* Protein splicing elements: Inteins and exteins: A definition of terms and recommended nomenclature. *Nucleic Acids Res.* **22**, 1125–1127. https://doi.org/10.1093/nar/22.7.1125 (1994).

72. Perler, F. B., Olsen, G. J. & Adam, E. Compilation and analysis of intein sequences. *Nucleic Acids Res.* **25**, 1087–1093. https://doi.org/10.1093/nar/25.6.1087 (1997).

73. Pietrokovski, S. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci.* **3**, 2340–2350. https://doi.org/10.1002/pro.5560031218 (1994).

74. Green, C. M., Novikova, O. & Belfort, M. The dynamic intein landscape of eukaryotes. *Mobile DNA* **9**, 4. https://doi.org/10.1186/s13100-018-0111-x (2018).

75. Sakai, M. & Sakaizumi, M. The complete mitochondrial genome of *Dugesia japonica* (Platyhelminthes; order Tricladida). *Zool. sci.* **29**, 672–680. https://doi.org/10.2108/zsj.29.672 (2012).

76. Hamsher, S. E., Evans, K. M., Mann, D. G., Poulíčková, A. & Saunders, G. W. Barcoding diatoms: exploring alternatives to COI-5P. *Protist* **162**, 405–422. https://doi.org/10.1016/j.protis.2010.09.005 (2011).

77. Kurtzman, C. P. & Robnett, C. J. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *A. van Leeuw. J. Microb.* **73**, 331–371. https://doi.org/10.1023/A:1001761008817 (1998).

78. Pedro, P. M. *et al.* Culicidae-centric metabarcoding through targeted use of D2 ribosomal DNA primers. *PeerJ* **8**, e9057. https://doi.org/10.7717/peerj.9057 (2020).

79. Perkins, E. M., Donnellan, S. C., Bertozzi, T., Chisholm, L. A. & Whittington, I. D. Looks can deceive: molecular phylogeny of a family of flatworm ectoparasites (Monogenea: Capsalidae) does not reflect current morphological classification. *Mol. Phylogenet. Evol.* **52**, 705–714. https://doi.org/10.1016/j.ympev.2009.05.008 (2009).

80. Sonnenberg, R., Nolte, A. W. & Tautz, D. An evaluation of LSU rDNA D1–D2 sequences for their use in species identification. *Front. Zool.* **4**, 6. https://doi.org/10.1186/1742-9994-4-6 (2007).

81. Stoeck, T., Przybos, E. & Dunthorn, M. The D1–D2 region of the large subunit ribosomal DNA as barcode for ciliates. *Mol. Ecol. Resour.* **14**, 458–468. https://doi.org/10.1111/1755-0998.12195 (2014).

82. Justine, J., Gey, D., Thévenot, J., Gouraud, C. & Winsor, L. First report in France of *Caenoplana decolorata*, a recently described species of alien terrestrial flatworm (Platyhelminthes, Geoplanidae). BioRxiv 2020.11.06.371385 (2020). https://doi.org/10.1101/2020.11.06.371385

83. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477. https://doi.org/10.1089/cmb.2012.0021 (2012).

84. Gordon, D. & Green, P. Consed: A graphical editor for next-generation sequencing. *Bioinformatics* **29**, 2936–2937. https://doi.org/10.1093/bioinformatics/btt515 (2013).

85. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669. https://doi.org/10.1093/bioinformatics/bty149 (2018).

86. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546. https://doi.org/10.1038/s41587-019-0072-8 (2019).

87. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* **9**, e112963. https://doi.org/10.1371/journal.pone.0112963 (2014).

88. Bernt, M. *et al.* MITOS: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319. https://doi.org/10.1016/j.ympev.2012.08.023 (2013).

89. Laslett, D. & Canbäck, B. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175. https://doi.org/10.1093/bioinformatics/btm573 (2008).

90. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580. https://doi.org/10.1093/nar/27.2.573 (1999).

91. Lohse, M., Drechsel, O., Kahlau, S. & Bock, R. OrganellarGenomeDRAW: A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **41**, W575–W581. https://doi.org/10.1093/nar/gkt289 (2013).

92. Kalvari, I. *et al.* Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200. https://doi.org/10.1093/nar/gkaa1047 (2021).

93. Hassouna, N., Michot, B. & Bachellerie, J. P. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Res.* **12**, 3563–3583. https://doi.org/10.1093/nar/12.8.3563 (1984).

94. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).

95. Wang, L. *et al.* Integrative taxonomy unveils a new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the southern portion of the Taihang Mountains in northern China, with the description of its complete mitogenome and an exploratory analysis of mitochondrial gene order as a taxonomic character. *Integr. Zool.* **17**, 1193–1214. https://doi.org/10.1111/1749-4877.12605 (2022).

96. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. https://doi.org/10.1093/molbev/mst010 (2013).

97. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348 (2009).

98. Smith, S. A. & Dunn, C. W. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**, 715–716. https://doi.org/10.1093/bioinformatics/btm619 (2008).

99. Darriba, D. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Boil. Evol.* **37**, 291–294. https://doi.org/10.1093/molbev/msz189 (2020).

100. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534. https://doi.org/10.1093/molbev/msaa015 (2020).

### Acknowledgements

### Author contributions

Collection of the samples by M.C., C.G., B.B. and A.K.M. Taxonomic identifications by J.L.J., L.W., B.B. and A.K.M. Curation of the samples at the M.N.H.N. and registration of the vouchers by J.L.J. Short and long reads sequencing of A. triangulatus by C.O. and B.B.. Bioinformatic analyses by R.G., C.L. and M.T. First draft written by R.G. Draft edited by M.T., C.L. and J.L.J. All authors read and approved the final draft.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-58600-y.

**Correspondence** and requests for materials should be addressed to R.G.

**Reprints and permissions information** is available at www.nature.com/reprints.

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