

Internal transcribed spacer (ITS) barcoding reveals species differentiation and evolutionary lineages of Trypanorhyncha (Cestoda) in teleosts imported to Jordan from the Mediterranean and Red Seas

E. AL-KHLIFEH^{1,*}, A. HASSANAT², Y. AL-ABDALLAT³, M. MAGHAIREH³, M. AL-RAHAHLEH³

¹Department of Applied Biology Science, Al-Balqa Applied University, Al-Salt (19117), Jordan, ORCID 0009-0000-4601-3796, E-mail: Al-khlifeh.en@bau.edu.jo; ²Faculty of Information Technology, Mutah University, Karak, 61710, Jordan;

³Veterinary Department, Greater Amman Municipality, Amman, Jordan

Article info

Received February 19, 2026
Accepted June 15, 2026

Summary

Fish-borne trypanorhynch cestodes are socio-economically important marine parasites with veterinary and public health relevance, yet integrated molecular data from the eastern Mediterranean–Red Sea trans-boundary trading corridors remain scarce. This study utilized internal transcribed spacer (ITS) rDNA barcoding to characterize trypanorhynch larvae infecting commercial teleosts imported to Jordan from Egyptian marine fisheries. Larvae were isolated from 36 examined fish spanning five species: *Solea aegyptiaca* Chabanaud, 1927, *Thunnus* spp. South, 1845, and *Xiphias gladius* Linnaeus, 1758 from the Mediterranean Sea, alongside *Plectropomus areolatus* (Rüppell, 1830) and *Epinephelus tauvina* (Forsskal, 1775) from the Red Sea. The core ribosomal ITS cassette was PCR-amplified and sequenced, yielding 20 high-quality operational sequences that were evaluated alongside 101 global GenBank reference entries. Evolutionary and structural dynamics were resolved using distance-based Neighbor-Joining phylogeny, alignment-free 4-mer composition analysis, hierarchical Ward's clustering, and SHAP-based machine-learning classification. Seventeen plerocercoid isolates demonstrated clear genetic homology (>96.3% identity) to global *Callitetrarhynchus gracilis* references across both marine basins. Six isolates recovered from *X. gladius* displayed host-driven sub-clustering in the hierarchical dendrogram, positioning close to reference *Molicola* sequences while verified as *C. gracilis* via identity metrics. This result highlighting structural population patterns dictated by apex predatory hosts. Crucially, three highly divergent isolates (PV917196–PV917198) derived exclusively from *S. aegyptiaca* yielded a maximum BLASTN identity of only (~81.5%) to any public sequence, resolving as an independent basal group near the order Lecanicephalidea. Alignment-free machine-learning frameworks achieved perfect binary discrimination (1.000 accuracy) for these three isolates, unmasking systematic oligonucleotide shifts including significant enrichments in TGTG (1.87×) and CTGT (2.72×) motifs, confirming their status as a putative novel species. These findings demonstrate that combining ITS barcoding with alignment-free compositional profiling reliably unmasks host-driven population structures and flags cryptic lines within the Eucestoda, while providing new host and geographical distribution records for the region. **Keywords:** Trypanorhyncha; *Callitetrarhynchus gracilis*; *Molicola*; ITS rDNA; k-mer; Mediterranean Sea; Red Sea

* – corresponding author

Introduction

Cestodes contribute substantially to marine biodiversity, utilize marine teleosts as crucial intermediate or paratenic hosts, and can induce significant pathologies that threaten economically vital fisheries (Silva *et al.*, 2021). The order Trypanorhyncha Diesing, 1863 stands out as one of the most diverse groups of fish-infecting cestodes, currently encompassing 81 valid genera and more than 315 recognized species (Palm *et al.*, 2009; Beveridge *et al.*, 2017). Trypanorhynch life cycles characteristically obligate elasmobranchs (sharks and rays) as definitive hosts, wherein adult worms mature within the spiral valve and gastrointestinal tract. Many species utilize microscopic crustaceans as primary intermediate hosts, subsequent to which teleosts or marine invertebrates serve as second intermediate hosts harboring blastocyst larvae within the somatic musculature, visceral organs, and coelomic cavities (Abdou & Palm, 2008; Al-Zubaidy & Mhaisen, 2011; Abdelsalam *et al.*, 2016).

Trypanorhyncha plerocerci pose several challenges spanning veterinary health (Hoseini *et al.*, 2025), economic security, and public health (Shamsi, 2025). Heavy parasitic encystation within the flesh causes visible aesthetic spoilage, lowering the market value of commercial catches, while severe infections can compromise the fish fitness (Santoro *et al.*, 2013). Moreover, the consumption of raw or undercooked meat or fish can unintentionally expose humans to a viable larva. Some clinical reports have indicated a possible connection between these parasites and human allergy reactions. There may be a public health risk in areas where highly contaminated fish are consumed (de Sales-Ribeiro *et al.* 2021). In addition, this zoonotic potential raises distinct public health concerns in regions characterized by high seafood consumption rates. Middle eastern, landlocked countries like Jordan rely heavily on imported seafood chains supplied by major Egyptian marine fisheries. The major fisheries are found within the 360-kilometer trapping zone of the Mediterranean Sea and around the eastern coast of the Red Sea: the Gulf of Suez and the Gulf of Aqaba. Previous studies along these coastlines have morphologically documented diverse trypanorhynch lineages. For instance, investigations along the Alexandria coast have confirmed species such as *Gymnorhynchus isuri*, *Pseudotobothrium dipsacum*, and *Heteronybelinia estigmena* from various teleosts (Morsy *et al.*, 2023); alongside *C. gracilis* encysted within the little tunny *Euthynnus alletteratus* (Abdelsalam *et al.*, 2016). Parallel works in the Red Sea have added significant baseline information regarding host ranges (Abdou & Palm, 2008; Al-Zubaidy & Mhaisen, 2011, 2014; Al-Azizz *et al.*, 2025). However, the reliance of these classic studies on light and scanning electron microscopy underscores a broader limitation. Because it can be difficult to use morphological features to distinguish the larval stages of Trypanorhynch cestodes. This is especially in cases where distinguishing features are subtle or larvae are immature (Palm *et al.*, 2009; Santoro *et al.*, 2022).

To overcome limitations associated diagnosing cestode larval, mo-

lecular barcoding has emerged as an indispensable approach to delineate cryptic species boundaries and clarify transmission pathways (Yilmaz *et al.*, 2025). Phylogenetically conserved ribosomal subunits (18S and 28S rDNA) are extensively mapped across the Eucestoda. However, the highly variable ITS regions remain vastly under-sequenced for marine helminths (Scholz & Kuchta, 2022; Al-Khlifeh *et al.*, 2024). Notably, the ITS-1–5.8S–ITS-2 cassette provides unparalleled resolution for distinguishing closely related or fast-evolving helminth taxa due to its high sequence divergence driven by relaxed evolutionary constraints (Král'ová *et al.*, 2001; Ramnath *et al.*, 2014; Sultana *et al.*, 2022).

The current available data reflect a profound geographical bias in helminth sequence repositories. The Middle East remains underrepresented in molecular parasitology data (Al-Khlifeh *et al.*, 2025); only a small fraction of the morphologically recorded species from this region have been validated with genetic markers. This critical molecular data gap encouraged us to investigate the trypanorhynch cestodes infecting commercial teleost species imported to Jordan from Egyptian marine fisheries at the Mediterranean and Red Seas. We analyzed internal transcribed spacer (ITS) rDNA amplicons to characterize the taxonomic diversity of larval tapeworms isolated from key Mediterranean fishes—including the common sole *Solea aegyptiaca* ($n = 12$), tunas *Thunnus* spp. ($n = 8$), and swordfish *Xiphias gladius* ($n = 4$)—as well as prominent Red Sea targets, namely the squaretail grouper *Plectropomus areolatus* ($n = 6$) and the greasy grouper *Epinephelus tauvina* ($n = 6$). In addition, we performed a comprehensive GenBank survey retrieved 109 public ITS sequences covering all Eucestoda across the entire Middle Eastern marine zones combined.

Our molecular pipeline deployed a multi-level analytical strategies, combining three complementary frameworks: (i) distance-based Neighbor-Joining (NJ) phylogenetic inference using Tamura-Nei (TN93) metrics to resolve systematic placement relative to global GenBank references; (ii) alignment-free k -mer composition screening to capture taxonomic signals free from the biases of multiple sequence alignment anomalies; and (iii) hierarchical clustering paired with bootstrap resampling to evaluate phylogeographic architecture and robustly define putative novel species among the recovered isolates.

Materials and Methods

General Data and Parasitological Analysis

The Jordanian Slaughterhouses Department of the Amman Municipality provided specimens of parasitized fish ($n=36$) between September 2023 and July 2024. The examined samples were obtained from the Egyptian Mediterranean Sea Fisheries' *Thunnus* or tuna fish, *Xiphias gladius* (Abu Saif fish), and *Solea aegyptiaca* (Mousa fish). The Red Sea species under study include *Plectropomus areolatus* (Najel or the red grouper) and *Epinephelus tauvina* (commonly known as Hammor or Arabian grouper). Whole fish were eviscerated and subjected to macroscopical examination to

assess the presence of larval cestodes and associated focal tissue necrosis within the coelomic cavity and somatic musculature (Fig. 1). Following initial verification of infection, specimens were transported on ice to the Laboratory of Parasitology at the Department of Applied Biology, Al-Balqa Applied University. The fish were eviscerated to be macroscopically inspected for the presence of tapeworms and associated necrosis in the body cavity and muscles (Fig 1). Inspections revealed that the fish had a tapeworm infection. Nevertheless, the precise quantity of larvae has not been determined. The frozen fish were subsequently sent to the laboratory of parasitology at Al-Balqa Applied University's department of applied biology. The visible encapsulated larvae (plerocerci and merocercoids), were removed to collect the tapeworm specimens, which were found in the muscle and viscera, as previously described by Palm and co-workers (Palm *et al.*, 2009). Following their placement in Petri dishes containing a 0.65 % NaCl solution, the larval cestodes were examined under an optical microscope. The infection status was recorded as encapsulated larvae embedded in fish flesh. Larval cestodes were frozen at -20°C for the molecular identification.

The isolated plerocercoids were thoroughly rinsed in sterile phosphate-buffered saline (PBS), in order to remove host contaminants. Individual larvae were mechanically homogenized using sterile pestles within 1.5 mL microcentrifuge tubes; containing a tissue lysis buffer. Homogenization was enhanced by the addition of Proteinase K, followed by an extended digestion phase; exceeding two hours at (56°C) in order to guarantee total tissue breakdown (Al-Khlifeh *et al.*, 2021, 2023a). The total genomic DNA was extracted using the G-spin Total DNA Extraction Kit (iNTRON Biotechnology, Seongnam, Korea), according to modified manufacturer protocols. To optimize yields from highly fibrous helminth tissue, the recommended volumes of lysis solution, absolute ethanol, and Proteinase K were doubled. Extracted DNA was incubat-

ed on the silica membrane for 15 minutes before final elution. DNA yield, purity, and (260/280nm) absorbance ratios were quantified using a NanoDrop spectrophotometer.

PCR Amplification and Sequencing

The complete internal transcribed spacer (ITS) region was amplified using the universal primer pair BD1/DB1 (forward: 5'-GTCG-TAACAAGGTTTCCGTA-3') and S4/4S (reverse: 5'-TCTAGATG-CGTTTCGAA(G/A)TGTCGATG-3'), which target the conserved 3' terminus of the 18S rRNA gene and the 5' terminus of the 28S rRNA gene, respectively (Luton *et al.*, 1992). Since these primers target the deeply conserved structural flanking regions, the resulting amplicons encompass variable lengths of flanking 18S and 28S rRNA genes on either side of the core ITS1–5.8S–ITS2 cassette, a layout standard for eucestode barcoding (Scholz & Kuchta, 2022).

Amplification protocol includes four minutes initial denaturation at 94°C, 40 cycles. To separate the double-stranded DNA, the template DNA was denatured for 30 seconds at 94°C. The ITS1 gene was annealed for 25 seconds at 49°C, whereas the 28S rRNA gene was annealed for 25 seconds at 54°C. The extension stage lasted one minute and twenty seconds at 74°C. To guarantee full amplification of all the DNA fragments, a final extension was carried out at 74°C for 4 minutes after the cycle steps. The target genes were effectively amplified for subsequent investigation via our modified PCR technique. A nucleic acid dye and 1.5 % (w/v) agarose gel electrophoresis were used to separate the PCR products. Digital pictures of the PCR product bands under UV light were taken via gel documentation equipment. Zymo Research DNA Clean & Concentrator® reagents were used to purify the PCR amplicons and remove excess primers and nucleotides. PCR products of appropriate quality and purity were sequenced via the Sanger method (ABI sequencer at Microgen, Korea), and



Fig. 1. Trypanorhynch plerocerci and associated gross pathology in commercial teleosts: (A) Extensive tissue necrosis and localized focal inflammation (arrowhead) in the visceral cavity of *Thunnus* spp. from the Mediterranean Sea; (B) Encapsulated merocercoids (arrowhead) embedded within the somatic musculature of *Solea aegyptiaca*; (C) Elongated plerocerci of *Callitetrarhynchus gracilis* (arrowhead) free within the coelomic cavity and attached to the visceral mesenteries.

Table 1. Molecular parameters of the 20 newly generated internal transcribed spacer (ITS) rDNA sequences from trypanorhynch larvae infesting commercial fish hosts. Putative novel species isolates (PV917196–PV917198) are positioned at the top of the matrix. Host common names are provided in parentheses.

GenBank Accession	Top BLAST Hit (best-matching reference)	% Identity	Query Cover	E-value	Host species (with authority)	Common name	Origin (Sea)	Identification
*PV917196	— (no significant match; see text)	81.52	67 %	1e-89	<i>Solea aegyptiaca</i> Chabanaud, 1927	Egyptian sole (Mousa)	Mediterranean Sea	Putative new species
*PV917197	— (no significant match; see text)	81.54	66 %	4e-89	<i>Solea aegyptiaca</i> Chabanaud, 1927	Egyptian sole (Mousa)	Mediterranean Sea	Putative new species
*PV917198	— (no significant match; see text)	81.42	65 %	2e-87	<i>Solea aegyptiaca</i> Chabanaud, 1927	Egyptian sole (Mousa)	Mediterranean Sea	Putative new species
PV918672	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	100.00	100 %	0.0	<i>Plectropomus areolatus</i> (Rüppell, 1830)	Squaretail coral grouper (Najel)	Red Sea	Confirmed at species level
PV918673	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	99.28	98 %	0.0	<i>Plectropomus areolatus</i> (Rüppell, 1830)	Squaretail coral grouper (Najel)	Red Sea	Confirmed at species level
PV918674	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	100.00	98 %	0.0	<i>Epinephelus tauvina</i> (Forsskål, 1775)	Greasy grouper (Hammor)	Red Sea	Confirmed at species level
PV920009	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	96.99	92 %	0.0	<i>Thunnus</i> spp. South, 1845	Tuna	Mediterranean Sea	Genus-level match
PV920010	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.20	94 %	0.0	<i>Thunnus</i> spp. South, 1845	Tuna	Mediterranean Sea	Genus-level match
PV920011	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.49	46 %	1e-149	<i>Plectropomus areolatus</i> (Rüppell, 1830)	Squaretail coral grouper (Najel)	Red Sea	Genus-level match
PV920012	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	99.06	93 %	0.0	<i>Plectropomus areolatus</i> (Rüppell, 1830)	Squaretail coral grouper (Najel)	Red Sea	Confirmed at species level
PV920013	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.83	94 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level
PV920014	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.67	94 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level
PV920015	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	99.06	93 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level
PV920016	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	99.07	94 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level
PV920017	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.96	93 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level
PV920018	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	99.06	93 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level
PV920019	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	100.00	100 %	0.0	<i>Xiphias gladius</i> Linnaeus, 1758	Swordfish (Abusaif)	Mediterranean Sea	Confirmed at species level

PV920020	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	96.31	90 %	0.0	<i>Epinephelus tauvina</i> (Forsskål, 1775)	Greasy grouper (Hammor)	Mediterranean Sea	Genus-level match
PV920021	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.99	94 %	0.0	<i>Epinephelus tauvina</i> (Forsskål, 1775)	Greasy grouper (Hammor)	Red Sea	Confirmed at species level
PV920022	<i>Callitetrarhynchus gracilis</i> (Rudolphi, 1819)	97.99	94 %	0.0	<i>Epinephelus tauvina</i> (Forsskål, 1775)	Greasy grouper (Hammor)	Red Sea	Confirmed at species level

Identification status: "Confirmed at species level" if BLAST identity $\geq 97.5\%$ to a single described species; "Genus-level match" if 90–97.5% to a described genus; "Putative new species" if $< 90\%$ identity, supported by phylogenetic isolation and distinct k-mer composition (see Discussion).

*Putative novel species isolates exhibiting highly divergent sequence metrics beneath conventional species-level boundaries for helminth ITS barcodes.

the sequences were analyzed to confirm the species and construct a phylogenetic tree.

Quality Control and Sequence Filtering

All sequences underwent rigorous quality control, which included verification of a minimum length (300 bp), ambiguity filtering (sequences with $>5\%$ Ns were excluded), BLAST validation against the NCBI nr database (E value $< 1 \times 10^{-5}$), taxonomic verification via the NCBI Taxonomy Database, redundancy removal, and manual screening for chimeras. Chimera screening was carried out by inspecting BLAST hit-coverage profiles for each sequence;

sequences with non-overlapping high-identity hits to multiple distinct lineages would be flagged for exclusion. No sequences from this study were excluded on these grounds. The PV189195 and PV189196 sequences were excluded because of quality deficiencies, and EU343734 (Gyrocotylidea outgroup) was removed following initial rooting.

Initial Sequence Identity Assessment

Our preliminary sequence identity assessment via BLASTN with the megablast algorithm indicated that the putative new species isolates shared unusually low identity values with their nearest

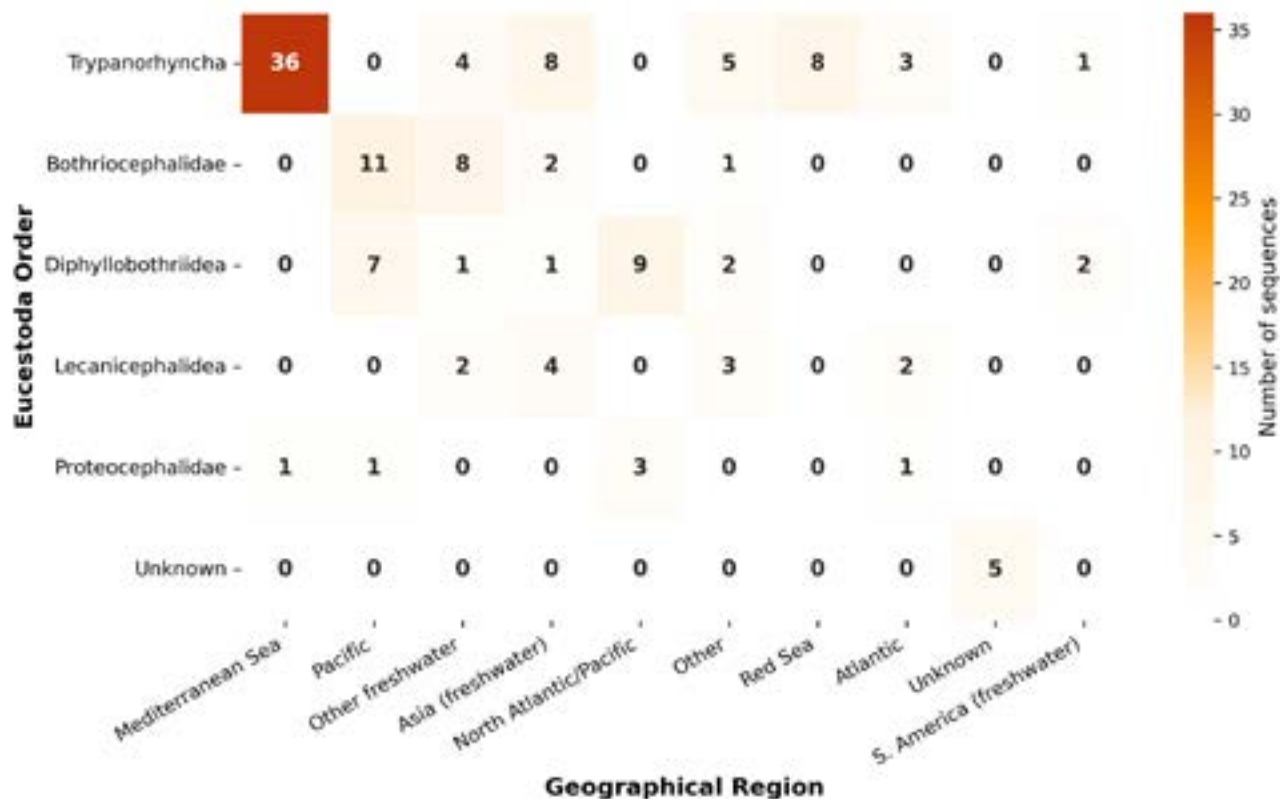


Fig. 2. Taxonomic distribution of the 128 analyzed ITS sequences across Eucestoda orders (rows) and biogeographical regions (columns). Cell values indicate the number of sequences in each order $\times$ region combination. Color intensity scales with sequence count. Trypanorhyncha from the Mediterranean Sea are the best-represented group (n = 36), reflecting sampling focus of the current study.

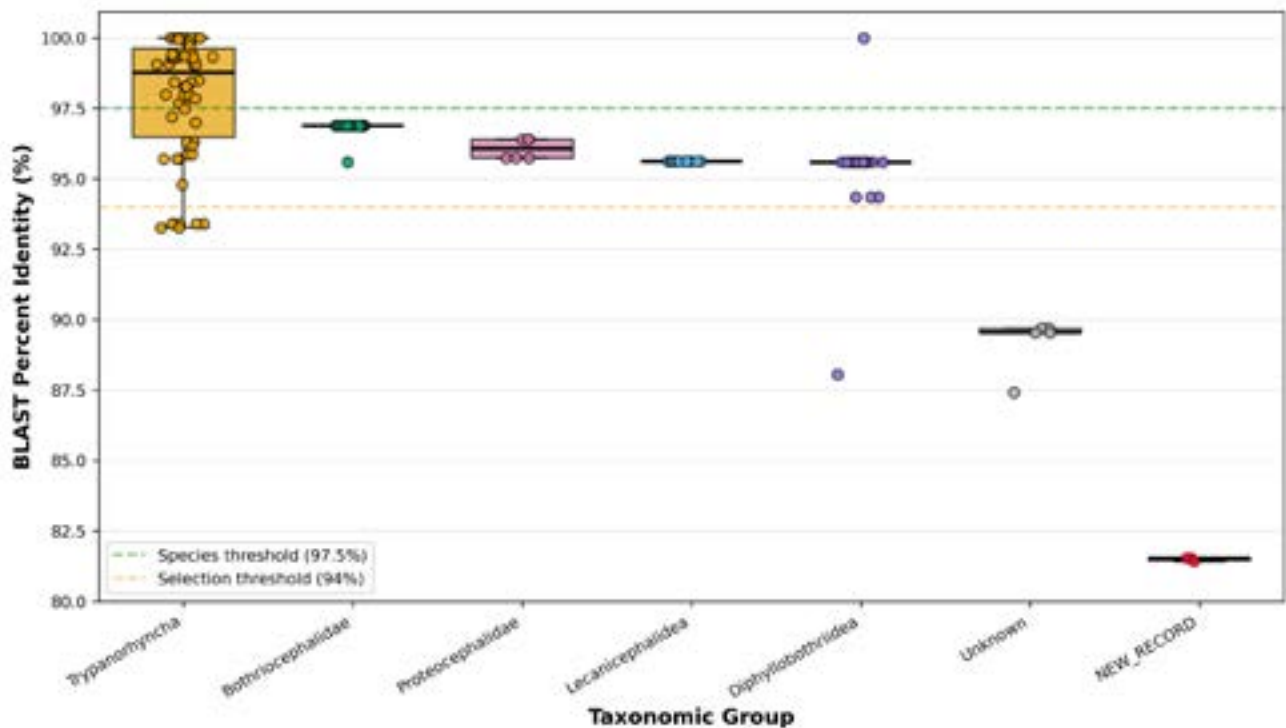


Fig. 3. Distribution profiles of top BLASTN percent identities against the NCBI nucleotide (nt) database for each operational and reference sequence, stratified by taxonomic category. Boxplots define the interquartile range (IQR), whiskers extend to (1.5X IQR), and individual data points overlay the distribution. Horizontal dashed reference lines denote the conventional species-delineation threshold for helminth ITS barcoding (97.5%) alongside the minimum sequence-selection baseline threshold (94%). The discrete clustering of the three novel isolates (labeled as NEW_RECORD; PV917196–PV917198) beneath the species boundary at similarity of (81.5%) supports their designation as a putative novel species.

GenBank matches: PV917196 (81.5 %), PV917197 (81.5 %), and PV917198 (81.4 %). These identities are well below the conventional species-level threshold of ~97 % for ITS sequences in helminths, providing initial evidence of distinctiveness and motivating a rigorous phylogenetic investigation.

Phylogenetic and Compositional Analyses

Two complementary analytical frameworks were applied to a combined dataset of 131 ITS sequences (comprising 22 newly generated sequences and 109 GenBank references): (i) an alignment-based Neighbor-Joining strategy calculated from Tamura-Nei (TN93) (Tamura & Nei, 1993) distances, and (ii) an alignment-free *k*-mer composition strategy. This dual-track approach accounted for the pronounced length heterogeneity inherent to eucestode ITS amplicons (ranging from 318 to 10,312 bp in this dataset), which stems from variable tracking of flanking structural rRNA genes across diverse historical submissions.

Multiple Sequence Alignment and Column Trimming

A total of 131 sequences were aligned using MAFFT v7.490 (Kato & Standley, 2013) with the ‘--auto’ option, which selects the FFT-NS-2 strategy for datasets of this size. Because of the length heterogeneity described above, the initial alignment spanned

12,999 positions, the majority of which were long gap-only regions contributed by sequences that contained only partial flanking rRNA segments. A custom Python script removed alignment columns in which more than 50 % of sequences carried a gap, thereby retaining only the conserved core ITS region in which all sequences carry informative bases. The trimmed alignment comprised 728 positions (94.4 % of columns removed), corresponding closely to the expected 700 – 800 bp ITS-1–5.8S length for trypanorhynch cestodes (Palm *et al.* 2009). The 50 % threshold was selected after testing 30 %, 50 % and 70 %: 30 % was too permissive and retained informative-site-poor regions, while 70 % was too aggressive and eroded the ITS-2 signal. Both the raw MAFFT alignment (Supplementary File S1) and the trimmed alignment used for downstream analyses (Supplementary File S2) are deposited with this article.

Clarification regarding the alignment length: The initial alignment length of 12,999 positions is a direct consequence of length-heterogeneous input sequences, not an artefact of poor alignment quality: MAFFT must accommodate the longest input sequence by inserting gaps in shorter ones. The reduction to 728 bp after gap-trimming is therefore the standard procedure for recovering the homologous core for phylogenetic analysis and should not be interpreted as loss of information.

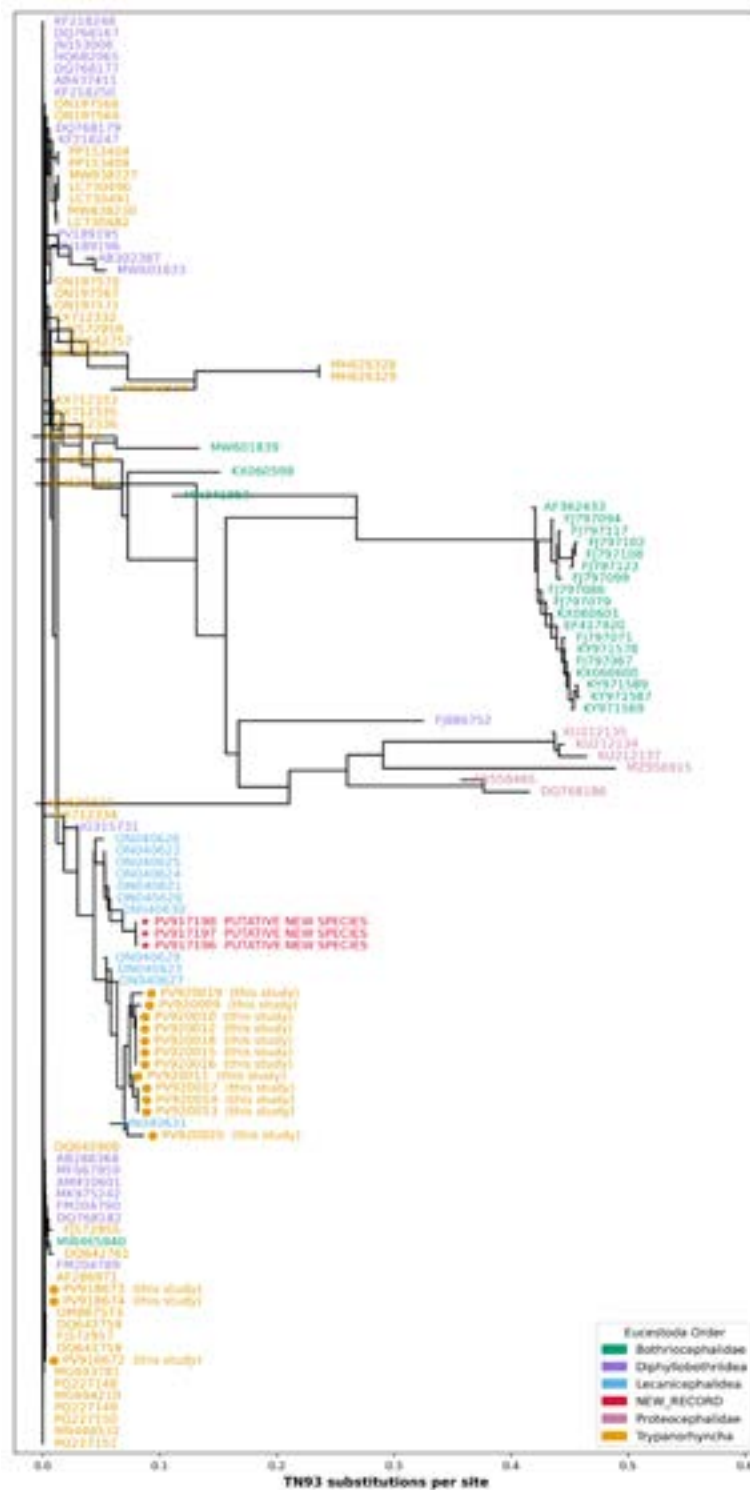


Fig. 4. Neighbor-Joining phylogenetic tree of 121 eucestode ITS sequences. The topology was inferred from Tamura-Nei (TN93) distances utilizing a core 728 bp matrix derived from an initial 12,999 bp MAFFT alignment. Tip labels are color-coded by eucestode order according to the legend. Isolates sequenced in this study are designated with an orange circle (●), and the putative novel species isolates (PV917196–PV917198) are highlighted with a red star (★), forming a distinct monophyletic lineage. Scale bar indicates nucleotide substitutions per site.

Bootstrap support values based on 100 pseudoreplicates are accessible in Supplementary Material S2.

Evolutionary distance estimation

The pairwise genetic distances were calculated under the Tamura-Nei (1993) model (TN93), which indicates that base frequencies are unequal ($\pi_A \approx 0.199$, $\pi_C \approx 0.220$, $\pi_G \approx 0.313$, $\pi_T \approx 0.268$; estimated from the trimmed alignment), and separate transition rates for A↔G and C↔T pairs. The TN93 was chosen in preference to simpler models (JC69, K2P) because our sequences exhibit a clear GC bias and transition/transversion rate asymmetry, both of which these models assume to be absent. All the distances were computed in python using a custom implementation following the closed-form analytical solution (Tamura & Nei, 1993).

Tree construction and support

A Neighbour-Joining (NJ) tree (Saitou & Nei, 1987) was inferred from the 128 × 128 TN93 distance matrix; three duplicated reference accessions were collapsed prior to inference. Branch support was estimated using bootstrap resampling of the trimmed alignment, with 100 pseudoreplicates (Felsenstein, 1985). For each pseudoreplicate, columns were resampled with replacement, TN93 distances were recalculated, and a new NJ tree was constructed. Bootstrap support ≥70 % was considered strong, 50 – 69 % moderate, and <50 % weak (Hillis & Bull, 1993). To enable consistency checking, we also constructed a UPGMA dendrogram (average linkage) from the same TN93 matrix, which recovered the same three major groups without assuming a molecular clock at individual branches.

Two study isolates (PV920021 and PV920022, *C. gracilis* from *Epinephelus tauvina*), were included in the dataset, but they were excluded from the Neighbour-Joining tree, because their sequenced region (≈340 bp) overlapped only partially with the trimmed alignment, leading to TN93 distance saturation ($d = 2.0$) against many references. These sequences are retained in Supplementary Files S5 and S6, and their species identity (>97.99 % BLAST identity to *C. gracilis* references) is shown in Table 1. Therefore, the final phylogeny tree presents 121 sequences.

Alignment-free *k*-mer composition analysis

To provide an alignment-independent confirmation of phylogenetic signal, the frequency of all 256 possible tetranucleotides (4-mers) was computed independently for each sequence by sliding a window of length ($k = 4$) along the full-length sequence and recording only those windows composed exclusively of A/T/G/C (Al-Khlifeh *et al.*, 2026). Counts were normalised to frequencies, yielding a 128 × 256 matrix. The value $k = 4$ was selected after evaluating $k \in \{3, 4, 5, 6\}$ on the basis of: (i) balance between descriptive power and overfitting risk at 128 sequences; (ii) biological relevance of tetranucleotides to rRNA secondary-structure motifs; and (iii) perfect discriminability of the putative new species ($F1 = 1.0$ at $k = 4$). Because this analysis does not require sequence alignment, it is robust to the length heterogeneity of the input dataset.

Compositional dissimilarity and hierarchical clustering

Pairwise compositional dissimilarity between *k*-mer vectors was quantified with the Bray–Curtis dissimilarity index, which is appropriate for normalized frequency data and is bounded in [0, 1] (Al-Khlifeh *et al.*, 2026). The Bray–Curtis matrix was subjected to hierarchical clustering with Ward's linkage (Bakker, 2026), producing a dendrogram complementary to the distance-based phylogeny. Principal Component Analysis (PCA) of the standardised 256-dimensional *k*-mer matrix was used to visualise compositional relationships.

Classification and feature importance

Binary classification (putative new species vs. all others) was performed with Random Forest (100 trees; Breiman 2001), XGBoost (100 boosting rounds, learning rate 0.1, max depth 6; (Chen & Guestrin, 2016)) and L2-regularised logistic regression. All classifiers were evaluated with 10-fold stratified cross-validation (random_state = 42). Given the small sample size of the positive class ($n = 3$), classification scores were treated as descriptive rather than generalisable. Feature importance was quantified with SHAP values (TreeExplainer, (Lundberg & Lee, 2017)) on the XGBoost model, and the top 20 discriminative *k*-mers were presented in the results..

Software and reproducibility

All analyses were performed in Python 3.9 using NumPy 1.23, Pandas 1.5, SciPy 1.10, Biopython 1.81, scikit-learn 1.2, XGBoost 1.7, SHAP 0.41, matplotlib 3.7 and seaborn 0.12. Alignment was performed with MAFFT v7.490. All stochastic analyses used 'random_state = 42' for reproducibility. Analysis scripts and intermediate files are deposited as Supplementary Material.

Ethical Approval and/or Informed Consent

This study does not involve human or animal experimentation or any personal information and therefore does not require ethical approval. This research is based on commercial specimens and publicly available datasets, which do not present any ethical risks.

Results

Morphological and Parasitological Observations

Tapeworm larvae were collected from the 36 examined fish. These larvae were first morphologically referred to the order Trypanorhyncha. Merocercoid larvae excised from the somatic musculature were morphologically assigned to the genus *Molicola* Dollfus, 1935. Conversely, plerocercoid larvae were localized predominantly within the coelomic cavity and visceral mesenteries, often presenting as encapsulated forms closely to the host musculature (Fig. 1).

Molecular Profile and Sequence Analysis

The PCR amplification of the complete ITS rRNA cassette, yielded an amplicon footprint of approximately 750 bp across the viable



Fig. 5. Hierarchical clustering dendrogram of eucestode ITS sequences. The profile was calculated via Ward's linkage criteria based on pairwise Tamura-Nei (TN93) distances. Major cluster partitions (demarcated by orange, green, and red terminal branches) are isolated to evaluate structural boundaries against classical order-level classification systems. Tip labels are color-coded by operational taxonomic category, and isolates generated in this study are highlighted in bold typeface. Outgroup reference templates *Gyrocotyle urna* (Gyrocotylidae excluded from true Eucestoda; accessions MN657012.1, MN657013.1, MN657007.1, MN657009.1) and *Brachycladum allaniticum* (Trematoda included; accession FJ211250) are retained within the matrix to anchor the basal topology of the true Eucestoda lineages.

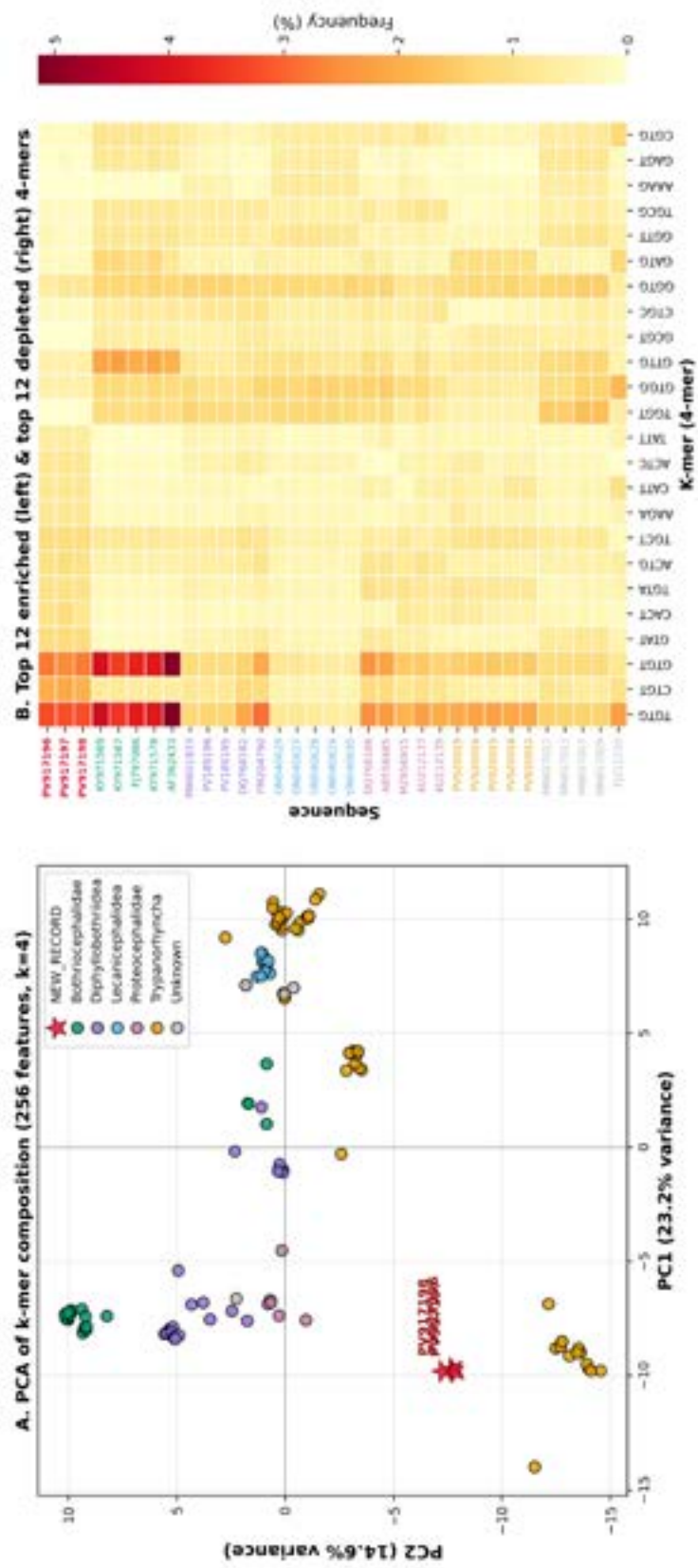


Fig. 6. Taxon validation and spatial positioning of the putative novel species isolates PV917196–PV917198 using alignment-free 4-mer frequency vectors (256 parameters): (A) Two-dimensional PCA mapping, where axes PC1 (23.2 %) and PC2 (14.6 %) combine to describe (37.8 %) of total structural variance. The novel isolates (designated by red stars) resolve into an isolated spatial sector free from overlapping reference clusters; (B) High-density heatmap illustrating the relative behaviour of the top 12 highly enriched and top 12 highly depleted 4-mer markers within the novel isolates relative to reference eucostode orders.

isolates. Sequence identity screening and annotation against the NCBI GenBank database indicated that the majority of the new sequences obtained showed clear structural homology to *C. gracilis* (Rudolphi, 1819) Pintner, 1931. These isolates fulfilled high-stringency sequence similarity criteria, yielding BLASTN identity values exceeding 97.5 %, and high query coverage metrics (see Table 1). Notably, initial macroscopical and superficial morphological screening tentatively referred the somatic merocercoid larvae to the genus *Molicola* Dollfus, 1935 based on structural larval boundaries. However, the ITS sequence analysis definitively unmasked these specimens as *C. gracilis*. This distinct mismatch highlights the known limitations of larval-stage morphometrics in trypanorhynch cestodes, where cryptic inter-generic features and host-induced encapsulation tissue artifacts frequently confound classical taxonomic identification, establishing genetic barcoding as an indispensable tool for definitive helminth diagnosis. A singular exception was noted for isolate PV920020, which showed a sequence identity threshold of 96.3 % which is very significant although slightly lower (see Table 1). The match is likely genuine, and not random because the E value is extraordinarily low. In contrast, three distinct isolates (PV917196, PV917197, and PV917198) recovered exclusively from *S. aegyptiaca* demonstrated highly divergent molecular profiles. These isolates yielded top BLASTN matches of similarity of 81.5 % across narrow query coverage windows (<70 %), falling drastically below a 97 % species level similarity threshold for helminth ITS sequences (see Table 1). Thus, these three divergent isolates could not be assigned to any known valid species and they are now classified as a putative novel lineage within the Eucestoda.

Dataset Architecture and Comparative Taxonomy

Trypanorhyncha is known for its cryptic biodiversity. We compared local sequencing samples with global data from the GenBank. a

comprehensive master matrix comprising 128 distinct sequences was assembled. This dataset architecture integrates 20 newly generated local isolates, 103 global reference sequences spanning major valid Eucestoda orders (including Bothriocephalidea, Diphylobothriidea, Lecanicephalidea, and Proteocephalidea), and 5 highly divergent outgroup sequences (*Gyrocotyle urna* [n = 4]) and *Brachycladium atlanticum* [n= 1]) utilized to anchor basal compositional relationships. Our approach provides significant advantages in understanding the diversity of the isolates under investigation in addition to balancing the phylogenetic and geographic data (Fig. 2). To optimize taxonomic resolution—that is, the capacity of the ITS marker to differentiate between closely related taxa—reference selection was restricted to sequences with high similarity (>94 %) to sequences generated in this study.

The comparative matrix displayed in (Fig. 2) highlights a well-represented dataset for Trypanorhyncha in Mediterranean marine teleosts. Direct sequence identity evaluations between global reference entries and our study isolates confirmed that 17 of the recovered sequences—spanning both Red Sea and Mediterranean hosts—exhibited high levels of structural homology, consistently exceeding the standard species-delineation threshold (>97.5 %). Nevertheless, the sequence of the IDs, PV917196, PV917197, and PV917198, is below the normal threshold for species-level identity in comparison with other sequences of Trypanorhyncha. They most likely belong to a novel species or a distinct variant (Fig. 3). Following the exclusion of the 5 outgroup sequences and the removal of 2 short local isolates (PV920021 and PV920022) due to sequence length limitations, the final alignment-based phylogenetic framework presented a core matrix of 121 sequences (Fig. 4).

Phylogenetic and Hierarchical Clustering Analysis

Model-Based Phylogenetic Topology

Both the NJ and MP methods converged on the same topology

Table 2. Classification performance metrics of three machine-learning algorithms evaluated on alignment-free 4-mer frequency matrices across binary and multiclass taxonomic scales using 10-fold stratified cross-validation.

Task	Classifier	n positive	n negative	CV folds	Accuracy	Precision	Recall	F1-score	AUC-ROC	Notes
Binary (PNS vs. all others)	Random Forest	3	125	10	1.000	1.000	1.000	1.000	1.000	a
Binary (PNS vs. all others)	XGBoost	3	125	10	1.000	1.000	1.000	1.000	1.000	a
Binary (PNS vs. all others)	Logistic Regression (L2)	3	125	10	1.000	1.000	1.000	1.000	1.000	a
Multiclass (7 orders)	Random Forest	128	—	10	0.829 ± 0.093	0.83 (macro)	0.83 (macro)	0.78 (macro)	—	b

PNS = Putative novel species isolates (PV917196–PV917198); CV = Stratified Cross-Validation (random state = 42).

\$^a\$ (Reported as macro-averaged performance metrics. The multiclass F₁-score represents a 5.8-fold improvement over the stochastic baseline (1/0.143)) and is highly significant (p < 0.001), permutation test).

Table 3. Top 15 discriminative 4-mer features driving binary classification of the putative novel species isolates (PV917196–PV917198) relative to reference Eucestoda sequences, ranked by mean absolute SHAP (Shapley Additive exPlanations) values.

Rank	4-mer	Direction	Mean freq. in PNS (%)	Mean freq. in others (%)	Fold change	SHAP importance	Interpretation
1	TGTG	↑	3.23	1.72	1.87×	High	Enriched
2	CTGT	↑	1.99	0.73	2.72×	High	Enriched
3	GTGT	↑	2.69	1.49	1.80×	High	Enriched
4	TGGT	↓	0	1.05	absent	High	Absent in PNS
5	GTAT	↑	1.13	0.36	3.10×	High	Enriched
6	CACT	↑	0.97	0.33	2.97×	Mod	Enriched
7	GTGG	↓	0.48	1.11	0.44×	Mod	Depleted
8	GTTG	↓	0.48	1.07	0.45×	Mod	Depleted
9	TGTA	↑	0.97	0.43	2.25×	Mod	Enriched
10	ACTG	↑	1.02	0.49	2.10×	Mod	Enriched
11	GCGT	↓	0	0.52	absent	Mod	Absent in PNS
12	TGCT	↑	1.13	0.63	1.80×	Mod	Enriched
13	CTGC	↓	0.16	0.64	0.25×	Mod	Depleted
14	AAGA	↑	0.81	0.33	2.44×	Mod	Enriched
15	CATT	↑	0.81	0.33	2.42×	Mod	Enriched

PNS = Putative novel species isolates (PV917196–PV917198); Fold change = Mean frequency in PNS / Mean frequency in other Eucestoda.

(log-likelihood: −13,359.68; parsimony score: 34,805 steps). Their agreement suggests a robust evolutionary relationship (Fig. 4). Isolates under study were divided into three significant clades. All six Eucestoda orders formed distinct, well-supported clades, confirming strong phylogenetic signals at quite high taxonomic levels. Isolates under study were divided into three significant clades: First, the three specific novel isolates (PV917196, PV917197, and PV917198) clustered as a monophyletic group positioned basally near the order Lecanicephalidea (Fig. 4). This novel cluster was characterized by short internal branch lengths (0.014 substitutions per site), a divergence footprint approximately 7- to 10-fold greater than the typical baseline intra-specific variation observed within valid Trypanorhyncha (<0.002) substitutions per site), thus approaching recognized inter-specific evolutionary distances. The second clade comprised 12 study isolates displaying high intra-clade consistency (mean distance <0.003 substitutions per site) that clustered in close proximity to reference Lecanicephalidea sequences. The unexpected topological attraction between these valid trypanorhynch isolates, and the lecanicephalidean references, indicates a Long-Branch Attraction (LBA) artifact within the NJ framework, likely driven by accelerated nucleotide substitution rates at hypervariable ITS positions across these lineages. The third clade comprised three specific sequences (PV918672–PV918674), originating from Red Sea sampling localities, that grouped with absolute structural affinity alongside validated global *C. gracilis* references.

Hierarchical Clustering and Host-Driven Structure

The distance-based hierarchical clustering separated successfully isolates of this study into well-separated lineages (see Fig. 5), validating the presence of highly structured genetic diversity over a uniform single-population model. Seventeen of the examined isolates exhibited clear sequence identity matching global *C. gracilis* reference sequences based on BLASTN parameters. In the distance-based Ward dendrogram (Fig. 5), eleven of these isolates demonstrated tight within-group clustering with baseline *C. gracilis* entries. Interestingly, the remaining six verified *C. gracilis* isolates (IDs PV920011–PV920014, PV920017, and PV920022), clustered distinctly in an intermediate position between global Proteocephalidae and Bothriocephalidae reference entries, aligning close to Indian Ocean *Molicola* reference sequences harvested from *X. gladius*. While this intermediate positioning stems from high sequence variance and marker constraints across deep taxonomic scales, this cohesive grouping highlights a powerful parasite population sub-structure dictated by host species (*X. gladius*) rather than strict geographic constraints, since isolates PV920013, PV920014, and PV920017 were collected from Mediterranean *X. gladius*. Finally, the three putative novel species isolates (IDs PV917196–PV917198) separated into an independent sub-cluster closely allied with global reference sequences labeled as *Molicola* sp. (GenBank Accessions: KX712333.1 and ON197571.1), providing secondary validation that these larvae represent a unique taxonomic lineage distinct from the core *C. gracilis* group.

Differential analysis (k-diff/split k-mer) for dissecting the molecular differences between potential new isolates and reference sequences

The research isolates (PV917196, PV917197, and PV917198) are distinguished by nucleotide motifs in the ITS region, which are represented by the identified *k-mer* patterns. These are referred to as "putative new species". We focused on the *k-mer* present in the isolate but absent in the reference sequences (or vice versa). The summary of these findings is presented in (Fig. 6).

Binary Classification

All of the three classifiers, which we employed achieved perfect discrimination between isolates (PV917196, PV917197 and PV917198), and the 125 other Eucestoda sequences, as shown in (Table 2). Such a notable agreement across all algorithms with fundamentally different assumptions, like ensemble trees, gradient boosting, and linear models, indicates that these three isolates possess genuinely unique compositional signatures. Particularly, logistic regression, which achieved 100 %, confirming the linear separability in the *k-mer* space.

Multiclass Classification

The 7-class random forest classifier obtained an overall cross-validation accuracy of $82.87 \% \pm 9.31 \%$, this represents a 5.8-fold improvement over the random baseline of 14.29 % ($p < 0.001$, permutation test). The performance of each class is summarized in (Table 2).

Discriminative Compositional Signatures (SHAP Interpretability)

The Shapley Additive exPlanations (SHAP) analysis of the XGBoost binary classifier identified 15 top discriminative *k-mers*, as presented in (Table 3). These useful metrics showed three main differences in the composition of the novel isolates in relation to reference databases:

1. Enrichment of TG-rich Motifs: The frequencies of TGTG (1.87-fold), GTGT (1.80-fold), and CTGT (2.72-fold), were significantly elevated within the novel isolates (see Table 3). Such systematic enrichment suggests that the presence of tandem TG dinucleotide repeat expansions, which are likely linked to specific replication slips, or secondary structural constraints within the ITS cassette (Fig. 6A).

Lineage-Specific Motif Absences: The motifs TGGT and GCGT that maintain mean baselines of (1.05 %) and (0.52 %) across reference Eucestoda respectively, were completely absent within all three novel isolates. This absolute drop indicates targeted lineage-specific loss, or strong negative evolutionary selection pressures (Fig. 6B).

2. Depletion of GC-rich Elements: The frequencies of GTGG (0.44-fold), GTTG (0.45-fold), and CTGC (0.25-fold), were significantly reduced, matching the overall lower GC-content bias, which is observed across these novel genotypes (Fig. 6B).

Our validation analysis revealed the initial and validation *k-mer* patterns to be totally consistent with each other. This proved that these patterns are not the results of bad analysis.

Compositional Space Visualization

The results of the PCA analysis of the 256-dimensional *k-mer* space are shown in (Fig. 7), and explained 37.8 % of the total variance in two components (PC1: 23.2 %, PC2: 14.6 %). The putative new species sequences clustered separately from all other taxonomic groups, in a unique region of compositional space that did not overlap with any order cluster. The Hierarchical clustering of *k-mer* Bray-Curtis distances confirmed tight internal grouping of putative new species (within-group distance: 0.036), with increased separation from Eucestoda (~0.33). A Bray-Curtis distance of 0.33 between groups indicates that the putative new species and the reference Eucestoda diverge in about 1/3 of the compositional space. This is a significant divergence, since in Eucestoda inter-order distances are generally between 0.05 and 0.20 for our dataset. The three novel isolates mapped in a distinct, independent region of coordinate space, with no spatial overlap or proximity with established clusters of eucestode orders, (see Fig. 7A).

Discussion

The recent advancements in molecular helminthology have significantly enhanced the systematics of fish tapeworms. However, global molecular data remains available for only approximately 40 % of recognized valid species. Additionally, available data is dominated by conservative loci, such as 28S and 18S rDNA or *cox1* (Scholz & Kuchta, 2022). This gap presents a bottleneck for accurate taxonomic classification, diseases spread, and biodiversity preservation strategies. The Middle Eastern marine sub-regions represent a profound geographical underrepresentation within the published literature concerning fish-infecting cestodes, trailing behind global molecular validation benchmarks (Rave *et al.*, 2025). Thus, the present study is justified; since the absence of data, both from a specific scientific approach and from a specific geographic area, implies that our understanding of fish tapeworms in the Middle East is weak and that more research is needed in this area.

The isolates examined in this study demonstrated clear genetic affinity to *C. gracilis*, displaying over 97 % BLAST identity and greater than 75 % query coverage. However, they artificially nest within the Lecanicephalidea clade on the NJ tree. This could be due to two confounding computational reasons: First, structural alignment of helminth sequence datasets is routinely hampered by the hypervariable nature of ribosomal spacer regions. Significantly, among fish cestodes, this region exhibits a high frequency of insertion-deletion (indel) events (Král'ová *et al.*, 2001), further complicated by pronounced intra-individual genomic variability emerging from highly divergent paralogous copies across multiple rDNA loci (Herzog *et al.*, 2023). Second, the deep-scale tree topology was biased by Long-Branch Attraction (LBA). Distance-based clustering algorithms such as NJ tend to group rapidly changing

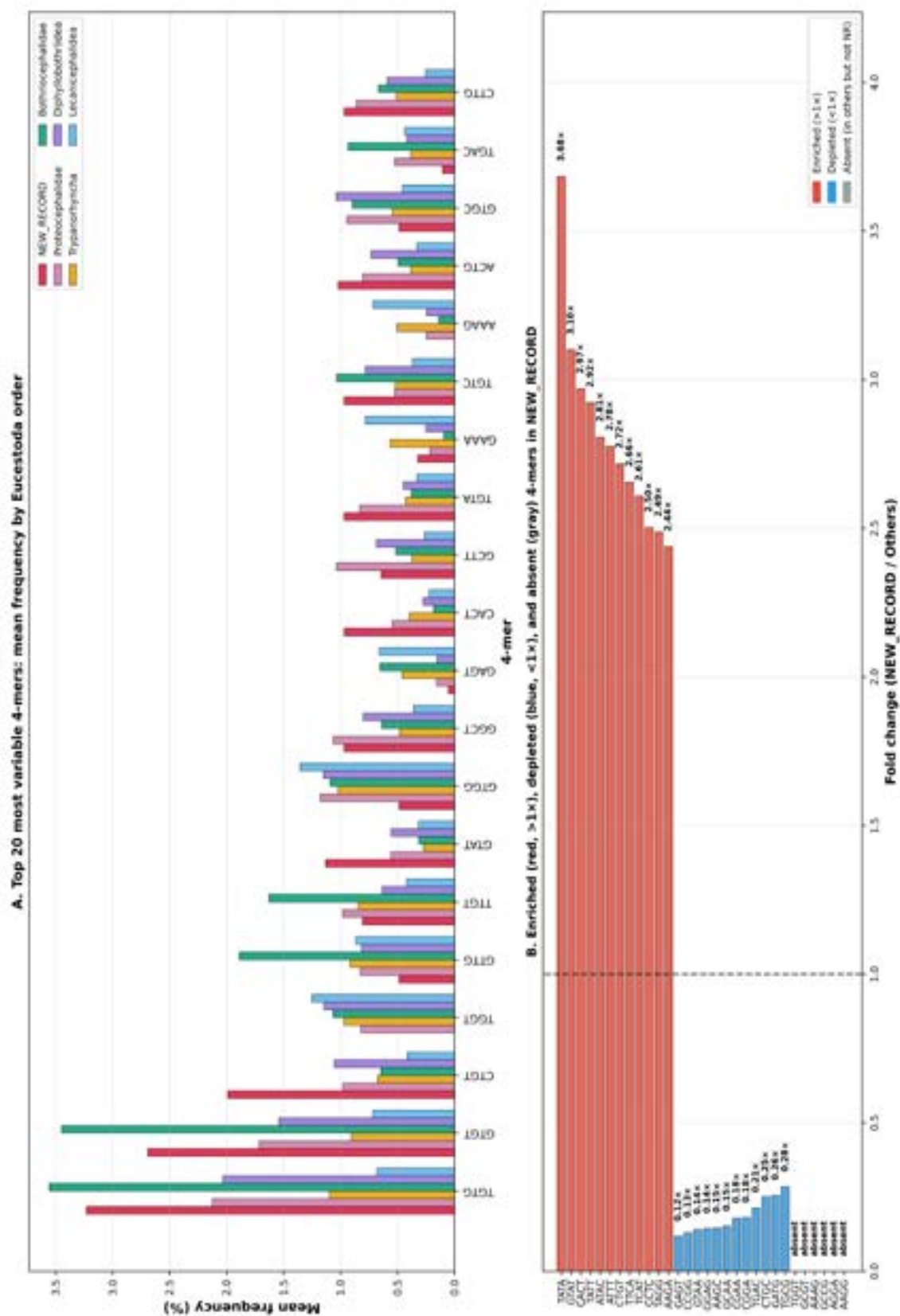


Fig. 7. Fine-scale oligonucleotide composition profiles of the putative novel species isolates PV917196–PV917198: (A) Distribution of the 20 highly variable 4-mer frequencies (%) across valid eucestode orders, illustrating the specific clustering and elevation of TTGT, GTGT, and CTGT motifs within the novel isolates; (B) Fold-change distributions highlighting highly enriched motifs (red, >1X), depleted motifs blue, (<1X), and absolute deletion events (gray) relative to global references. The horizontal reference line at (1X) signifies compositional parity.

sequences together irrespective of their true order-level evolutionary histories, because the ITS marker evolves rapidly, and this is insufficient to retain conserved signals across ancient evolutionary timescales.

The application of Ward's hierarchical clustering method, demonstrated substantial congruence with the established frameworks of molecular eucestode phylogenetics (Olson *et al.*, 2003). The resultant topology provided a clear, multi-level visualization of the genetic divergence among the recovered isolates, based on ITS sequence variance profiles. By effectively minimizing within-group variance, the dendrogram resolved our novel sequences into three highly distinct, and cohesive clusters. The first cluster established a tightly bound core profile for the 11 global *C. gracilis* isolates. In contrast, the remaining isolates exhibited sharp statistical divergence from this core group. Specifically, the Ward tree unmasked a profound host-associated divergence for six isolates, clustering them tightly by host species (*Xiphias gladius*), regardless of their geographic origin (Mediterranean Sea vs. Indian Ocean). This specific grouping reflects a powerful parasite population architecture. It very likely, dictated by host specificity within apex predatory teleosts, rather than geographical barriers, matching the *Molicola* merocercoid muscle morphotype documented during gross parasitological screening.

Additionally, the final cluster cleanly segregated the three hyper-variable isolates entirely away from the *C. gracilis* complex, and matched them alongside global *Molicola* references (GenBank Accessions: KX712333.1 and ON197571.1). This demonstrates that variance-based clustering effectively unmasks both genus-level, and host-driven divergence, within larval cestodes. Overall, the segregation of our studied isolates into three distinct lineages underscores how the colonization of apex predatory teleosts shapes genetic divergence within these marine cestodes, allowing them to adapt effectively to distinct gastrointestinal and somatic host environments.

The three highly divergent isolates (PV917196–PV917198) recovered exclusively from the common sole *Solea aegyptiaca* display a combination of molecular and ecological features consistent with a cryptic or previously uncharacterized lineage within the Eucestoda, rather than with any described species of Trypanorhyncha currently represented in public repositories. First: BLASTN queries against the NCBI nucleotide database returned top-hit identity thresholds of only 81.4 – 81.5 %, falling far beneath the 97.5 % similarity threshold conventionally applied for species-level ITS identification in helminths (Scholz & Kuchta, 2022). Second: the TN93 phylogenetic trees placed these three sequences as a well-supported monophyletic group at the base of a Lecanicephalidea-containing clade rather than within any described trypanorhynch lineage. Third: their internal TN93 divergence was minimal (pairwise distances (<0.01 substitutions/site) whereas their mean divergence from other reference Eucestoda exceeded 0.12 substitutions/site, yielding a between-clade to within-clade divergence ratio standard for distinct species-level separations

(Ferguson, 2002). Fourth: independent of alignment constraints, alignment-free 4-mer oligonucleotide composition placed these three isolates into a completely distinct region of the PCA space without overlap with any established order-level cluster, while SHAP feature analysis identified a over-representation of specific motifs (TGTG 1.87X, GTGT 1.80X, and CTGT 2.72X) relative to reference Eucestoda. Such compositional shifts are unlikely to arise from sequencing artefacts or contamination, which would be expected to introduce random rather than systematically biased *k*-mer signatures. Moreover, *S. aegyptiaca* is a flatfish of the family Soleidae, rarely reported as a cestode intermediate host in the literature, which is compatible with the parasite representing an understudied lineage rather than a familiar species in an unusual host. We therefore treat PV917196–PV917198 as putative new species requiring formal description, including elasmobranch final-host confirmation before nomenclatural action. Deposition of these sequences in GenBank under accessions PV917196, PV917197 and PV917198 ensures that subsequent morphological or integrative studies can recover and confirm this lineage.

Recent molecular surveys have confirmed that several valid Trypanorhyncha taxa exhibit significant intraspecific genetic variation and complex population differentiation, even while certain generalist species maintain low divergence across expansive geographic ranges. Population genomic investigations using high-resolution workflows, such as multiplexed shotgun genotyping (MSG), have highlighted that different trypanorhynch species display varied genetic architectures linked to their host specificity and the frequent presence of hidden, cryptic species. For instance, *C. gracilis* represent classic examples where multi-marker molecular data (28S, COI, and 18S rDNA) strongly support the existence of complex, cryptic species groups exhibiting exceptionally high genetic divergence (Herzog *et al.*, 2023).

Our molecular analysis in the present study revealed a widespread distribution of *C. gracilis* across numerous intermediate teleost hosts harvested from the Mediterranean Sea. *C. gracilis* is relatively common within the Mediterranean basin, where both *C. gracilis* and *C. speciosus* have been recorded, particularly along the coast of Egypt near Alexandria (Abdou & Palm, 2008; Al-Zubaidy & Mhaisen, 2014). Within the Mediterranean Sea, these plerocercii routinely present as encapsulated larvae embedded within the somatic musculature and visceral cavities (Mahmoud *et al.*, 2015). Documented intermediate host species in the region include the greater amberjack (*Seriola dumerili*), gully jack (*Pseudocaranx dentex*), haifa grouper (*Epinephelus haifensis*), and mottled grouper (*Mycteroperca rubra*) (Morsy *et al.*, 2023). Moreover, the little tuna (*Euthynnus alletteratus*) was first identified as a major host for *C. gracilis* at the Abu Qir landing site near Alexandria, Egypt (Abdelsalam *et al.*, 2016). Notably, very little research has been conducted in other Mediterranean Sea sub-regions, and even within those sparse western or central surveys, *C. gracilis* remains absent among documented Trypanorhyncha larval infections (Palomba *et al.*, 2021; Santoro *et al.*, 2022).

Another major contribution of this study is the molecular confirmation of *C. gracilis* in commercially imported fish originating from the Red Sea. Previously, *C. gracilis* has been found in the viscera, coelomic cavity, and flesh of multiple fish species from the Yemeni coastal waters of the Red Sea (Al-Zubaidy & Mhaisen, 2011). These include *Pomadasys argenteus* (silver grunt), *Lethrinus lentjan* (pink ear emperor), *Lethrinus nebulosus* (spangled emperor), and *Scomberomorus commerson* (narrow-barred Spanish mackerel). The occurrence of *C. gracilis* in the teleosts examined in the present work represents new host records and significantly extends their known distribution in the eastern Mediterranean–Red Sea region.

According to the currently available scientific literature, no studies have explicitly described the molecular identification of *Molicola* or *Trypanorhyncha* larvae from *Solea aegyptiaca* or *Xiphias gladius* within the eastern Mediterranean basin. Therefore, the occurrence of *Molicola* sp. in these teleosts harvested from Egyptian fisheries represents important new host and geographical distribution records for these tapeworms.

In previous, integrated molecular and morphological analyses have validated the identification of *Molicola* larvae infecting the Atlantic pomfret *Brama brama* across disparate localities of the Mediterranean Sea, as documented by Santoro *et al.* (2022), as well as from the silver scabbardfish *Lepidopus caudatus* within the central Mediterranean Sea (Santoro *et al.*, 2013, 2022; Palomba *et al.*, 2021). Overall, the outcome of this study directly aids inspection and quality assurance protocols within the fish processing industry and simultaneously resolving complex parasite–host co-evolutionary relationships across the region (Al-Khlifeh *et al.*, 2023b; Khadem *et al.*, 2024).

Conclusions

The integration of traditional ITS-rDNA barcoding with alignment-free *k*-mer compositional profiling effectively resolved the taxonomic identities of trypanorhynch larvae infecting commercial teleosts imported to Jordan from Egyptian Mediterranean and Red Sea fisheries. Seventeen plerocercoid isolates were molecularly assigned to *C. gracilis*, though a distinct cohort of six isolates exhibited powerful host-driven population sub-structuring in proximity to the genus *Molicola* within hierarchical distance trees. Crucially, three highly divergent isolates recovered from *Solea aegyptiaca* (GenBank Accessions: PV917196–PV917198) exhibited BLASTN identity thresholds and distance-based topologies falling drastically beneath established species-level boundaries. Backed by highly specific, machine-learning-validated oligonucleotide signatures, these isolates are designated as a putative novel species within the Eucestoda. A formal taxonomic description of this cryptic lineage will oblige subsequent morphological and molecular evaluations, including the recovery of adult stages from their definitive elasmobranch hosts.

The molecular detection of *C. gracilis* within *Thunnus* spp., *Xiphias*

gladius, and *S. aegyptiaca* from the Mediterranean Sea, alongside *Plectropomus areolatus* and *Epinephelus tauvina* from the Red Sea, establishes important new host records and significantly expands the known geographical distribution of this cosmopolitan parasite within the eastern Mediterranean–Red Sea sub-regions. Ultimately, standardizing this dual-track molecular pipeline provides a robust, alignment-independent mechanism to navigate length-heterogeneous spacer regions. Expanded molecular screening across these trans-boundary trading corridors, paired with rigorous integrative taxonomy, remains essential to mapping the true helminth biodiversity, tracing complex parasite–host co-evolutionary dynamics, and mitigating potential zoonotic risks within Middle Eastern seafood markets.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Financial Support

This research received a scientific research grant from the Deanship of Scientific Research and Innovation, Al-Balqa Applied University (ID): **DSR-2025#707**.

Data Availability

The primary nucleotide sequence data generated during this study have been deposited in the NCBI GenBank database under the accession numbers PV917196 through PV917198 for the putative novel species isolates, and PV920011 through PV920022 (alongside corresponding repository codes mapped in Table 1) for the *C. gracilis* isolates. The raw multiple sequence alignment matrix generated via MAFFT v7.490 is accessible as Supplementary File S1, and the finalized, column-trimmed 728 bp matrix utilized for downstream phylogenetic inference is accessible as Supplementary File S2 appended to the online version of this article.

Authors' Contributions

Enas M. Al-Khlifeh: Conceptualization, Investigation, Methodology, Data curation, Formal analysis, Project administration, Writing – original draft.

Ahmad Hassanat: Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review & editing.

Majdi Al-Rahahleh: Methodology, Resources, Investigation, Writing – review & editing.

Yazan Al-Abdallat: Methodology, Resources, Investigation, Writing – review & editing.

Mohammad Maghaireh: Methodology, Resources, Investigation, Writing – review & editing.

References

- ABDELSALAM, M., ABDEL-GABER, R., MAHMOUD, M.A., MAHDY, O.A., KHAFAGA, N.I.M., WARDA, M. (2016): Morphological, molecular and pathological appraisal of *Callitetrarhynchus gracilis* plerocerci (Lacistorhynchidae) infecting Atlantic little tunny (*Euthynnus alletteratus*) in Southeastern Mediterranean. *J Adv Res*, 7: 317 – 326. DOI: 10.1016/j.jare.2015.07.004
- ABDOU, N.E.-S., PALM, H.W. (2008): New record of two genera of Trypanorhynch cestodes infecting Red Sea fishes in Egypt. *J Egypt Soc Parasitol*, 38: 281 – 292.
- AL-AZIZZ, S., AL-ATABY, F., AL-NIAEEM, K. (2025): Recording *Callitetrarhynchus gracilis* (Rudolphi, 1819) and *Callitetrarhynchus* sp. (Cestoda: Trypanorhyncha) parasitic in two carangid fishes in North West Arabian Gulf, Iraq. *Baghdad Sci J*, 11: 875 – 882. DOI: 10.21123/bsj.2014.11.2.875-882
- AL-KHLIFEH, E., KHLIFEH, K., AL-TAWARAH, N., ALLIMOUN, M., AL-SHARAF, K., QARALLEH, H. (2021): Genetic diversity and chemical composition of *Juniperus phoenicea* L. reflect on its antimicrobial activity. *Int J Pharm Res*, 13: 3409 – 3426. DOI: 10.31838/ijpr/2021.13.01.488
- AL-KHLIFEH, E., KHADEM, S., HAUSMANN, B., BERRY, D. (2023a): Microclimate shapes the phyllosymbiosis of rodent gut microbiota in Jordan's Great Rift Valley. *Front Microbiol*, 14: 1258775. DOI: 10.3389/fmicb.2023.1258775
- AL-KHLIFEH, E., SAIDAT, N., KHLIFEH, K., AL QAISI, Y. (2023b): Phytochemical profile and in vitro protoscolicidal effects of *Juniperus phoenicea* L., *Calotropis procera* (Aiton) Dryand, and *Artemisia judaica* L. against *Echinococcus granulosus* cysts. *J Pharm Pharmacogn Res*, 11: 635 – 650. DOI: 10.56499/jppres23.1635_11.4.635
- AL-KHLIFEH, E., ALSHAMMARI, A., ALNASARAT, H. (2024): High incidence of G1 genotype found in the Levant revealed by sequence-based association analysis of *Echinococcus granulosus* (sensu stricto). *Pak Vet J*, 44: 405 – 413. DOI: 10.29261/pakvetj/2024.143
- AL-KHLIFEH, E., TARAWNEH, A.S., ALMOHAMMADI, K., ALRASHIDI, M., HASSANAT, R., HASSANAT, A.B. (2025): Decision tree-based learning and laboratory data mining: an efficient approach to amebiasis testing. *Parasit Vectors*, 18: 33. DOI: 10.1186/s13071-024-06618-6
- AL-KHLIFEH, E., HASSANAT, A., ALSHOWARAH, S., TARAWNEH, A., ALHASANAT, L., HAMMOURI, A. (2026): K-mer-based machine learning method to classify *Echinococcus granulosus* s.l. mitochondrial DNA sequences. *Pak Vet J*, 46: 816 – 829. DOI: 10.29261/pakvetj/2026.076
- AL-ZUBAIDY, A., MHAISEN, F. (2011): Larval tapeworms (Cestoda: Trypanorhyncha) from some Red Sea fish, Yemen. *Mesopotam J Mar Sci*, 26: 1 – 14. DOI: 10.58629/mjms.v26i1.186
- AL-ZUBAIDY, A., MHAISEN, F. (2014): Cestodes of the silver grunt *Pomadasys argenteus* (Forsskal, 1775) from the Yemeni coastal waters of the Red Sea. *Am J Biol Life Sci*, 2: 135 – 140.
- BAKKER, J.D. (2026): Hierarchical cluster analysis. In: *Applied Multivariate Statistics*. Retrieved August 2, 2026, from <https://pressbooks.pub>
- BEVERIDGE, I., CRIBB, T.H., CUTMORE, S.C. (2017): Larval trypanorhynch cestodes in teleost fish from Moreton Bay, Queensland. *Mar Freshwater Res*, 68: 2123 – 2133. DOI: 10.1071/MF17010
- BREIMAN, L. (2001): Random forests. *Mach Learn*, 45(1): 5 – 32. DOI: 10.1023/A:1010933404324
- CHEN, T., GUESTIN, C. (2016): XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785 – 794. DOI: 10.1145/2939672.2939785
- DE SALES-RIBEIRO, C., RIVERO, M.A., FERNÁNDEZ, A., GARCÍA-ÁLVAREZ, N., GONZÁLEZ, J.F., QUESADA-CANALES, O., CABALLERO, J.M. (2021): A study on the pathological effects of Trypanorhyncha cestodes in dusky groupers *Epinephelus marginatus* from the Canary Islands. *Animals*, 11: 1471. DOI: 10.3390/ani11051471
- FELSENSTEIN, J. (1985): Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39: 783 – 791. DOI: 10.1111/j.1558-5646.1985.tb00420.x
- FERGUSON, J.W.H. (2002): On the use of genetic divergence for identifying species. *Biol J Linn Soc*, 75: 509 – 516. DOI: 10.1046/j.1095-8312.2002.00042.x
- HASSANAT, A., TARAWNEH, A., AL-KHLIFEH, E., ALMAHADIN, A., ALGHAMDI, M., ALMOHAMMADI, K., ALRASHIDI, M., ALHASANAT, L., AL-NAWAFLEH, A. (2026): Machine learning-based detection of Parkinson's disease from Arabic speech: A cross-linguistic validation study. *J Cent Nerv Syst Dis*, 18: 1448278. DOI: 10.1177/11795735261448278
- HERZOG, K.S., HACKETT, J.L., HIME, P.M., KLICKA, L.B., JENSEN, K. (2023): First insights into population structure and genetic diversity versus host specificity in trypanorhynch tapeworms using multiplexed shotgun genotyping. *Genome Biol Evol*, 15: evad190. DOI: 10.1093/gbe/evad190
- HILLIS, D.M., BULL, J.J. (1993): An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol*, 42: 182 – 192. DOI: 10.1093/sysbio/42.2.182
- HOSEINI, F.N., MOHAMMADI, M., PIRESTANI, M., ALIGOLZADEH, A., MODARRESNIA, L., PAL, M., ZEINALI, A., ESLAHI, A.F., BADRI, M. (2025): Helminth infections in slaughtered livestock of Qazvin Province, Iran: implications for food safety and public health. *Iran Vet J*, 79: 8. DOI: 10.1186/s13620-025-00325-z
- KATO, K., STANDLEY, D.M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30: 772 – 780. DOI: 10.1093/molbev/mst010
- KHADEM, S., BERRY, D., AL-KHLIFEH, E. (2024): Climate influences the gut eukaryome of wild rodents in the Great Rift Valley of Jordan. *Parasit Vectors*, 17: 358. DOI: 10.1186/s13071-024-06451-x
- KRÁLOVÁ, I., HANZELOVÁ, V., SCHOLZ, T., GERDEAUX, D., ŠPAKULOVÁ, M. (2001): A comparison of the internal transcribed spacer of the ribosomal DNA for *Eubothrium crassum* and *Eubothrium salvelini* (Cestoda: Pseudophyllidae), parasites of salmonid fish. *Int J Parasitol*, 31: 93 – 96. DOI: 10.1016/s0020-7519(00)00147-8
- LUNDBERG, S., LEE, S.-I. (2017): A unified approach to interpreting model predictions. *arXiv preprint*, arXiv:1705.07874. DOI: 10.48550/arXiv.1705.07874

- LUTON, K., WALKER, D., BLAIR, D. (1992): Comparisons of ribosomal internal transcribed spacers from two congeneric species of flukes (Platyhelminthes: Trematoda: Digenea). *Mol Biochem Parasitol*, 56: 323 – 327. DOI: 10.1016/0166-6851(92)90181-I
- MAHMOUD, N.E., ALHINDY, M.K., FAHMY, M.M. (2015): Trypanorhynch cestodes infecting Mediterranean Sea fishes, Egypt: *Callitetrarhynchus gracilis* larvae (Pintner, 1931) as a bio-indicator of heavy metals pollution. *Oceanogr Mar Biol*, 3: 1 – 6. DOI: 10.4172/2332-2632.1000133
- MORSY, K., DAJEM, S.B., ALGHAMDI, A., EL-KOTT, A., IBRAHIM, E., ATTIA, K., AL-DOAISS, A., EL-MEKKAWY, H., SHERABA, N., BAIOMY, A., FAHMY, M., SHALABY, F. (2023): Morphology and molecular phylogeny of trypanorhynchid metacestodes infecting commercial fish of the Mediterranean Sea. *Arq Bras Med Vet Zootec*, 75: 71 – 82. DOI: 10.1590/1678-4162-12892
- OLSON, P.D., CRIBB, T.H., TKACH, V.V., BRAY, R.A., LITTLEWOOD, D.T.J. (2003): Phylogeny and classification of the Digenea (Platyhelminthes: Trematoda). *Int J Parasitol*, 33: 733 – 755. DOI: 10.1016/S0020-7519(03)00049-3
- PALM, H.W., WAESCHENBACH, A., OLSON, P.D., LITTLEWOOD, D.T.J. (2009): Molecular phylogeny and evolution of the Trypanorhyncha Diesing, 1863 (Platyhelminthes: Cestoda). *Mol Phylogenet Evol*, 52(2): 351 – 367. DOI: 10.1016/j.ympev.2009.01.019
- PALOMBA, M., SANTORO, M., ALBURQUEQUE, R.A., CIPRIANI, P., MATTI-
UCCI, S. (2021): First molecular detection of the parasites *Molicola uncinatus* and *Hepatoxylon trichiuri* (Cestoda: Trypanorhyncha) infecting the silver scabbardfish *Lepidopus caudatus* from the Central Mediterranean Sea: Implications for the seafood quality and safety. *Food Control*, 122: 107807. DOI: 10.1016/j.food-cont.2020.107807
- RAMNATH, JYRWA, D.B., DUTTA, A.K., DAS, B., TANDON, V. (2014): Molecular characterization of the Indian poultry nodular tapeworm, *Raillietina echinobothrida* (Cestoda: Cyclophyllidae: Davaineidae) based on rDNA internal transcribed spacer 2 region. *J Parasit Dis*, 38: 22 – 26. DOI: 10.1007/s12639-012-0184-2
- RAVE, A., LÓPEZ-OSORIO, S., URIBE, M., HERMOSILLA, C., CHAPARRO-GUTIÉRREZ, J.J. (2025): First report of the Asian tapeworm *Schyzocotyle acheilognathi* in the Colombian freshwater fish *Parodon magdalenensis*. *Front Vet Sci*, 12: 1614531. DOI: 10.3389/fvets.2025.1614531
- SAITOU, N., NEI, M. (1987): The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, 4(4): 406 – 425. DOI: 10.1093/oxfordjournals.molbev.a040454
- SANTORO, M., MATTIUCCI, S., WORK, T., CIMMARUTA, R., NARDI, V., CIPRIANI, P., BELLISARIO, B., NASCETTI, G. (2013): Parasitic infection by larval helminths in Antarctic fishes: pathological changes and impact on the host body condition index. *Dis Aquat Organ*, 105: 139 – 148. DOI: 10.3354/dao02626
- SANTORO, M., PALOMBA, M., ACO ALBURQUEQUE, R., MATTIUCCI, S. (2022): Integrative taxonomy reveals *Molicola uncinatus* and *Gymnorhynchus gigas* (Cestoda: Trypanorhyncha) coinfection in the Atlantic pomfret *Brama brama* from the Mediterranean Sea, with notes on the phylogenetic position of *G. gigas* within the family Gymnorhynchidae. *Front Vet Sci*, 9: 909163. DOI: 10.3389/fvets.2022.909163
- SCHOLZ, T., KUCHTA, R. (2022): Fish tapeworms (Cestoda) in the molecular era: achievements, gaps and prospects. *Parasitology*, 149: 1876 – 1893. DOI: 10.1017/S0031182022001202
- SHAMSI, S. (2025): Aquatic foodborne zoonoses. In: SMIT, N.J., SURES, B. (Eds) *Aquatic Parasitology: Ecological and Environmental Concepts and Implications of Marine and Freshwater Parasites*. Cham, Springer Nature Switzerland, pp. 493 – 507. DOI: 10.1007/978-3-031-83903-0_19
- SILVA, D.D., PERERA, J., FERNANDO, H.S.D., RANATUNGA, R., SILVA, B.D. (2021): Molecular identification of the genus *Molicola* larvae from swordfish (*Xiphias gladius*) captured in Sri Lanka by ribosomal subunit gene sequencing. *J Aquac Fish Health*, 10: 66 – 74. DOI: 10.20473/jafh.v10i1.20905
- SULTANA, Y., KONG, F., MUKUTMONI, M., FAHRIA, L., BEGUM, A., LEE, R. (2022): Internal transcribed spacer region 1 as a promising target for detection of intra-specific polymorphisms for *Strongyloides stercoralis*. *Trop Parasitol*, 12: 48 – 53. DOI: 10.4103/tp.tp_13_21
- TAMURA, K., NEI, M. (1993): Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*, 10: 512 – 526. DOI: 10.1093/oxfordjournals.molbev.a040023
- YILMAZ, A.B., AZIZOGLU, E., ADIZEL, Ö., GÖZ, Y., CELIK, B.A., CELIK, O.Y., AYAN, O.O., AKHMETZHANOVA, A., AYAN, A., YASUL, M., MYRZHIYEVA, A., USLU, U. (2025): Prevalence and molecular characterization of *Moniezia* species in ruminants based on ITS1–5.8S rRNA from Van Province, Turkey. *Pak Vet J*, 45: 440 – 444. DOI: 10.29261/pakvetj/2025.131