University of South Bohemia in České Budějovice

Faculty of Science

Department of Parasitology

MSc thesis

Molecular phylogeny of Nearctic proteocephalids of the *Proteocephalus*-aggregate (Cestoda: Onchoproteocephalidea)

Bc. Lucie Uhrová

Supervisor: RNDr. Jan Brabec, Ph.D.

Consultant: Prof. RNDr. Tomáš Scholz, CSc.

České Budějovice 2019

Uhrová L. 2019: Molecular phylogeny of Neartic proteocephalids of the Proteocephalus-aggregate (Cestoda: Onchoproteocephalidea). MSc. Thesis in English – Faculty of Science, University of South Bohemia, České Budějovice: 44 pp.

Annotation:

Phylogenetic relationships of species of the *Proteocephalus*-aggregate, a group of proteocephalidean tapeworms from freshwater teleosts with Holarctic distribution, were evaluated with emphasis on enlarged sampling of representative taxa from the Nearctic region. Molecular phylogenetic analyses based on 28S rDNA and cytochrome c oxidase subunit 1 supported validity of the newly sampled species and showed that Nearctic species of the *Proteocephalus*-aggregate do not form a monophyletic lineage, neither that species of *Proteocephalus*-aggregate parasitizing common fish host groups share a common evolutionary history. In addition, the current molecular data are not capable of resolving internal phylogeny of the group and alternative molecular markers should be aimed in future phylogenetic treatments of the group.

This project was financed by Grant Agency of the Czech Republic project no. P505/12/G112 (European Centre of Ichthyoparasitology; granted to T. Scholz).

Declaration [In Czech]

Prohlašuji, že svoji magisterskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své magisterské práce, a to v nezkrácené podobě – v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, dne 11.12.2019

.....

Lucie Uhrová

Acknowledgements

My thanks are due to Anindo Choudhury from St. Norbert College (USA) who collected and provided me samples which were analyzed in this study. I would like to express my gratitude to my supervisor Jan Brabec for his guidance, professional advice and help during writing my Master's thesis. Likewise, I would like to thank Tomáš Scholz, for useful remarks, advice and the opportunity to be part of the Laboratory of Helminthology. It was a pleasure to collaborate with all members of this laboratory surrounded by friendly environment and extraordinary support. Most importantly, I would like to thank my family and closest friends for their continuous support through my studies.

CONTENTS

1 INTRODUCTION

1.1 Neodermata

Neodermata is a group of obligatory parasitic flatworms that together with other groups of mostly free-living flatworms form a phylum Platyhelminthes, one of the major phyla belonging to the bilaterian animal supergroup Lophotrochozoa that also includes molluscs and annelids, among others (Halanych et al., 1995). While the interrelationships within the Lophotrochozoa are not yet fully resolved despite the use of phylogenomic data (Kocot et al., 2017), monophyly of the "true" Platyhelminthes (excluding acoelomorphs) has been well supported since the use of the first molecular data (e.g. Ruiz-Trillo et al., 1999). Molecular phylogenetic studies show that the Neodermata forms the most derived clade of the phylum Platyhelminthes (e.g., Littlewood et al., 1999; Egger et al., 2015; Laumer et al., 2015) and consists of three to four major clades: trematodes (Aspidogastrea + Digenea), cestodes (Amphilinidea + Eucestoda + Gyrocotylidea), and likely paraphyletic monogenean groups Monopisthocotylea and Polyopisthocotylea Fig. 1 (Littlewood, 2008; Perkins et al., 2010).

The term Neodermata was proposed by Ehlers (1984) on the basis of a common presence of neodermis, an apomorphic character of the taxon. Neodermis is a specialized epidermis (syncytial epithelium of mesodermal origin) with its cell nuclei immersed below the basal lamina. Neodermis (or tegument) covers the body of all representatives of the Neodermata and helps to protect the parasite from the environment and immunity system of the host, transport the nutrients or reject toxins (Rohde et al., 1993; Ehlers, 1985; Littlewood and Olson, 2001).

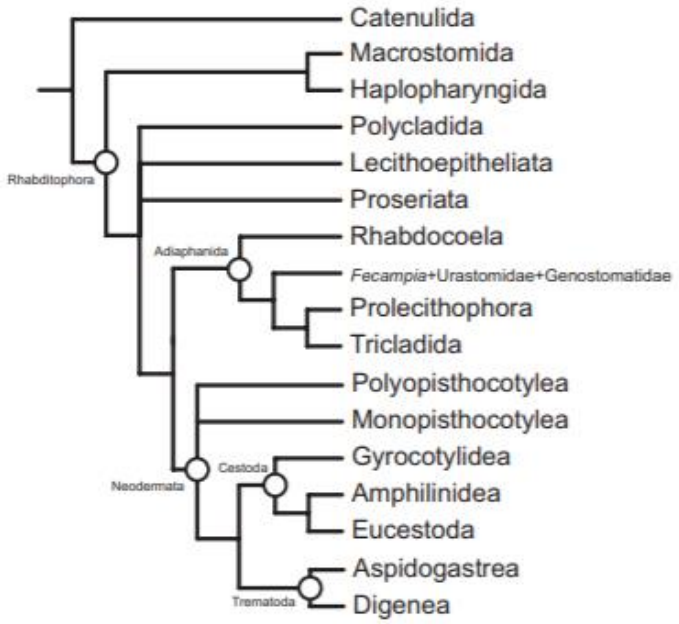


Fig. 1. Consensus view of the interrelationships of the Platyhelminthes (from Littlewood, 2008).

1.2 General features of cestodes

Cestodes (also tapeworms) are highly diversified parasites of all classes of vertebrates (Baily et al., 2014). Adult cestodes reside in the digestive tract of the definitive vertebrate host, but their larval stage(s) are found in body cavities of numerous invertebrates but also in tissues of vertebrates (Yamaguti, 1959). Cestodes are unique among other neodermatans by the absence of all parts of the digestive tract and the presence of a distinct type of tegument that differs from the other groups of the Neodermata by the presence of microtriches, specialized cylindrical extensions of the neodermis. Microtriches play an important role in absorption of nutrients from the lumen of the host intestine and might facilitate adhesion to the wall of the digestive tract (Chervy, 2009). With very few exceptions, tapeworms are generally hermaphroditic, which means that each worm individual possesses one or more sets of both female and male reproductive organs in a single body.

Body of an adult cestode follows a general body plan shared by most representatives, consisting of an anteriorly situated scolex and a posterior strobila. Scolex is mostly equipped with an attachment organ, whose type and morphology is often particular for a given tapeworm group, and thus serves as an important morphological character in cestode

taxonomy. The attachment organs can take numerous forms including grooves (bothria), hooks, suckers, tentacles, or others (Khalil et al., 1994; Roberts and Janovy, 2009). Posterior to the scolex and often anterior to the first segment of the strobila is a neck, a germinative zone where new proglottids are formed. The strobila can be either monozoic or polyzoic. Monozoic strobila consists of a single differentiated unit and bears only a single set of male and female reproductive organs (e.g., Caryophyllidea) whereas polyzoic strobila consists of many proglottids, each housing one or more sets of male and female reproductive organs (Roberts and Janovy, 2009). Tapeworm life cycles can be either indirect (heteroxenous) or direct (monoxenous). Vast majority of life cycles are indirect and include a definitive host (e.g. mammalian carnivore) and one or several intermediate hosts where larvae transform into the next larval stage and where asexual reproduction might occur. Life cycles can also include a paratenic host, which serves only as reservoir for transport to the definitive hosts (Goater et al., 2014). Larval stages are highly morphologically variable and their terminology was unified by Chervy (2002). We find only few exceptions among cestodes that display direct cycles in which the parasite develops within a single host individual, for example cyclophyllidean *Hymenolepis nana* (von Siebold, 1892), a human parasite or the caryophyllidean *Archigetes sieboldi* Leuckart, 1978 from freshwater oligochaetes (Olson et al., 2008; Saari et al., 2018).

Class Cestoda includes 4,810 valid species from 833 genera (Caira and Jensen, 2017). Estimates of the total number of cestodes are currently exceeding 6,000 species according to Caira and Jensen (2017). Diversity of the true tapeworms, a group called Eucestoda, is vast (Fig. 2) and its representatives parasitize about 20,000 vertebrate hosts (Caira and Jensen, 2017). Early branching lineages of tapeworms are generally found in aquatic vertebrates, both marine and freshwater. However, tapeworms managed to colonise all terrestrial tetrapods on several occasions during the evolution of the group, including reptiles and amphibians (e.g., Diphyllobothriidea and Onchoproteocephalidea), birds (e.g., Diphyllobothriidea, Cyclophylidea,) and mammals (e.g., Cyclophyllidea and Diphyllobothriidea; Caira et al., 2014). Several cestode species are important parasites of humans, who can serve as either an intermediate or a definitive host. While the infections with adult stages are generally asymptomatic, infections with cestode larval stages often result in severe diseases, e.g., echinococcosis (caused by the genus *Echinococcus*; Ash and Orihel, 2007) or sparganosis (genus *Spirometra*; Kuchta et al., 2014).

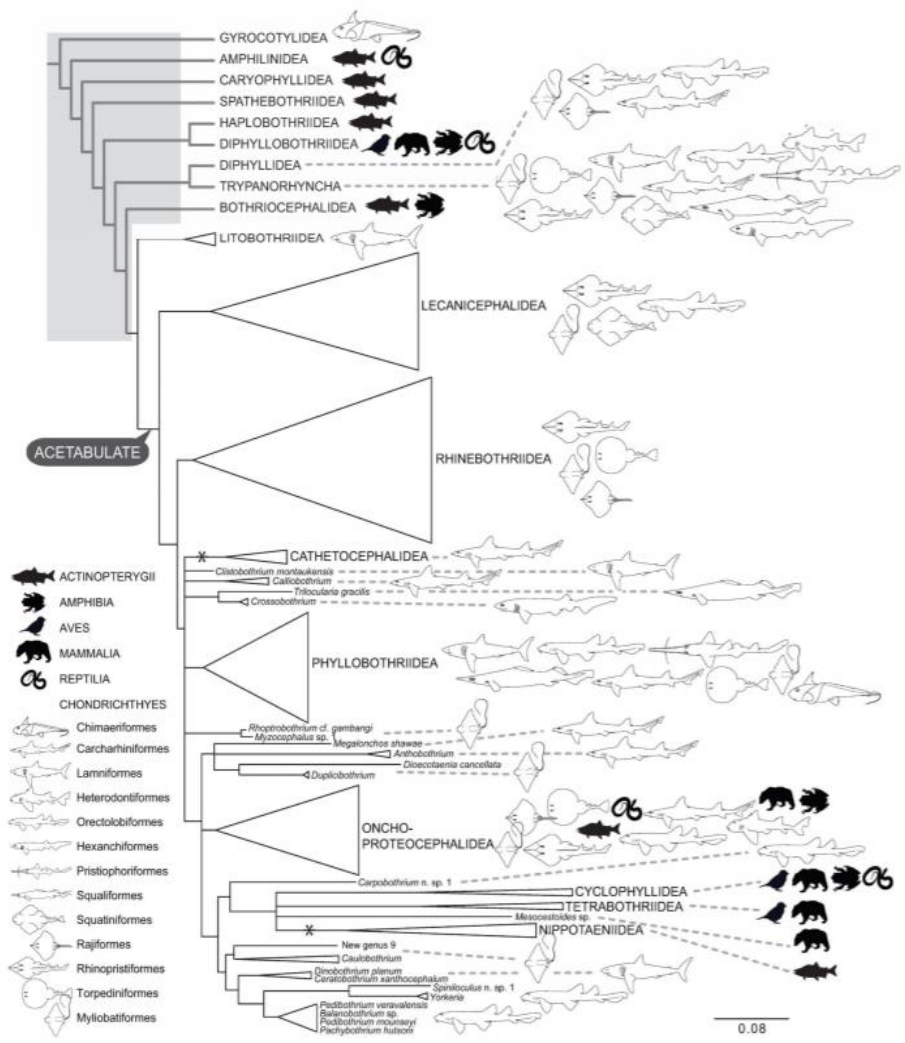


Fig. 2 Summary view of the phylogenetic relationships of the currently recognized 19 orders of tapeworms (Cestoda) based on the molecular data. The definitive host spectra of individual orders are depicted next to the taxon names (from Caira et al., 2014). Asterisk marks a secondary loss of acetabulum, an attachement muscular sucker.

1.3 Proteocephalidea (Cestoda: Onchoproteocephalidea)

The order Onchoproteocephalidea is a recently established group uniting some of the hook-bearing acetabulate cestodes traditionally placed in the family Onchobothriidae Braun, 1900 of the paraphyletic order Tetraphyllidea, parasites of elasmobranchs, and the former order Proteocephalidea, mostly parasites of bony fish, frogs, snakes and lizards (Caira et al., 2014).

Tapeworms of the order Proteocephalidea are found in freshwater fishes, reptiles and amphibians. One species, *Thaumasioscolex didelphidis* Cañeda-Guzmán, de Chambrier & Scholz, 2001, was also found in a mammal, the black-eared opossum in Mexico (Cañeda-Guzmán et al., 2001). The group includes 316 species in 67 genera and 12 subfamilies (de Chambrier et al., 2017). One of the genera with numerous species of proteocephalideans is *Proteocephalus* Weinland, 1858 which has been found paraphyletic by numerous studies including Zehnder and Mariaux (1999), Chambrier et al. (2004c, 2015) and Hypša et al. (2005) and has not yet been entirely revised. Proteocephalideans are globally distributed parasites and most of their diversity can be found in siluriform fishes of South America (Freze, 1965; Rego, 1994; Scholz and de Chambrier, 2003; de Chambrier et al., 2017). However, numerous species of proteocephalideans are also found in other teleost groups, for example Cypriniformes, Gasterosteiformes, Perciformes and Salmoniformes (de Chambrier et al., 2017).

Historically, classification of the Proteocephalidea at the subfamily level was established by W.N.F. Woodland (e.g., Woodland, 1925, 1933, 1935). This classification was later accepted by many authors and the order Proteocephalidea was subdivided into two families: the Proteocephalidae La Rue, 1911 and the Monticelliidae La Rue, 1911, each consisting of many subfamilies (Yamaguti, 1959; Freze, 1965; Schmidt, 1986; Rego, 1994). Rego's (1995) proposed new classification in which he did not recognize the validity of the family Monticelliidae. As a result, the group currently consists of a single family, the Proteocephalidae, further divided into 12 subfamilies: Acanthotaeniiae Freze, 1963; Corallobothriinae Freze, 1965; Ephedrocephalinae Mola, 1929; Gangesiinae Mola, 1929; Marsypocephalinae Woodland, 1933; Monticelliinae Mola, 1929; Nupeliinae Pavanelli & Rego, 1991; Peltidocotylinae Woodland, 1934; Proteocephalinae Mola, 1929; Rudolphiellinae Woodland, 1935; Sandonellinae Khalil, 1960 and Zygobothriinae Woodland, 1933 (Rego et al., 1999).

Proteocephalideans are polyzoic tapeworms with acraspedote, anapolytic proglottids each containing one set of genital organs. The acetabulate scolex of proteocephalideans is always equipped with four muscular suckers of variable shape, position and structure. Scolex may possess rostellum-like muscular organ with hooklets, an apical organ (glandular or glandulo-muscular), an apical sucker or a concentration of gland cells (Scholz and de Chambrier, 2003; de Chambrier et al., 2017). An important morphological characteristic of proteocephalideans is the type of uterine development. Two basic types of uterus were

described by de Chambrier et al. (2004c). Differences between these two types can be observed mainly in premature and mature proglottids (dotted part in Fig. 3). Type 1 is found in the subfamilies Acanthotaeniinae, Gangesiinae and proteocephalideans from reptilian families Viperidae and Elapidae. Type 2 is typical for the Holarctic *Proteocephalus*-aggregate and most of the Monticelliidae (sensu Rego, 1994).

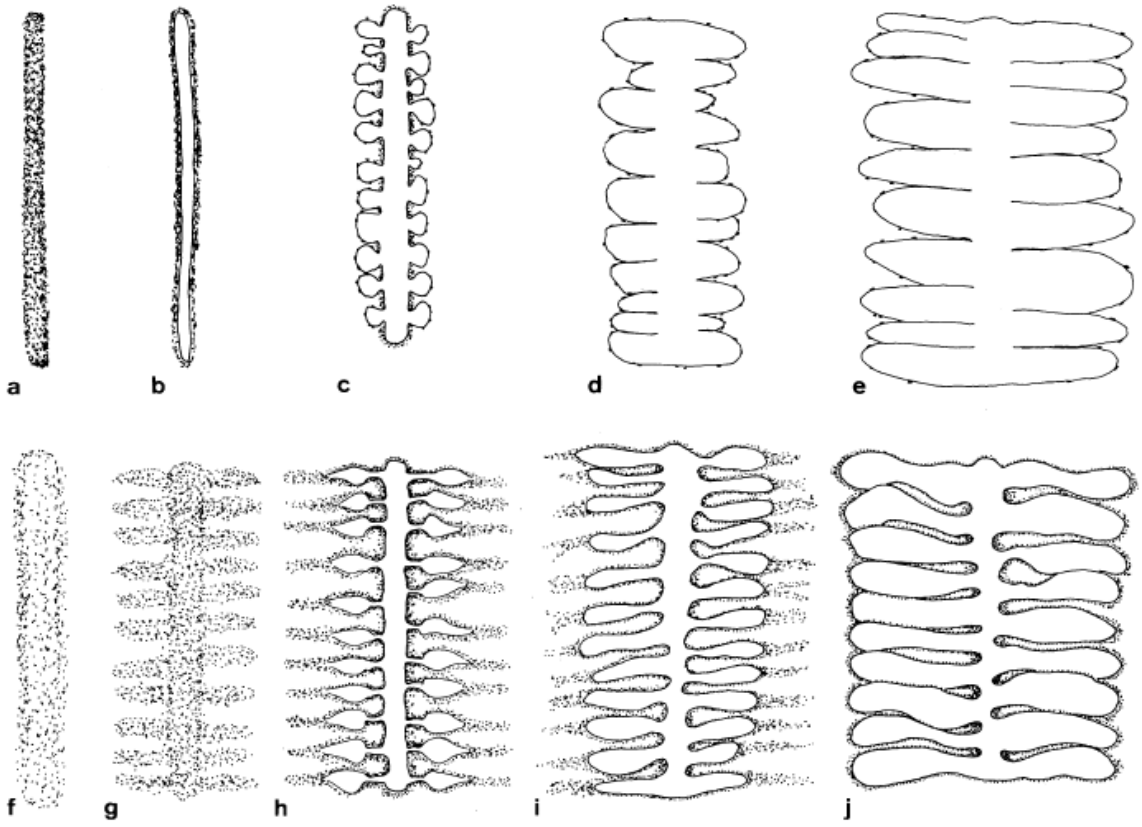


Fig. 3 Schematic view of two basic types of development of the uterus of proteocephalideans: type 1 (a-e) and type 2 (f-j). Immature, premature, mature pregravid and gravid proglottids are shown in the direction left to right (taken from de Chambrier et al., 2004c).

Life cycles of proteocephalideans have been studied in more detail in several species of proteocephalideans, for example *Proteocephalus longicollis* (see Scholz, 1999 for review). Proteocephalideans produce eggs that contain an oncosphere bearing three pairs of embryonic hooks (i.e. hexacanth). Plerocercoid larval stage develops after the oncosphere is eaten by an intermediate host, a planktonic crustacean (typically cyclopoid or diaptomid copepod). A plerocercoid that possesses the scolex similar to that of adult (see Chervy, 2002) develops in the body cavity of the crustacean. The final vertebrate host is usually infected directly after consuming infected crustaceans (Scholz and de Chambrier, 2003). However, in case of some

species (e.g., *Proteocephalus ambloplitis*), second intermediate (and also paratenic) hosts are included in the parasite's life cycle, typically when small fishes are eaten by carnivorous fish or in case of cannibalism (Freze, 1965).

Veterinary importance of proteocephalidean cestodes is negligible because they occur mostly in animals of limited economic importance. Exceptions include *Proteocephalus longicollis* (Zeder, 1800), a parasite of salmoniform fishes in the Holarctic region, which can endanger health and growth of the host, or larval stages of *P. ambloplitis* (Leidy, 1887), which can cause host castration (Williams and Jones, 1994). Despite the low pathogenicity, proteocephalideans might represent a model for studies of host parasite relationships because of their particularly narrow host-specificity (notably proteocephalideans of South American catfishes), making them a suitable model for studies of parasite-host coevolution (Škeříková et al., 2001; Scholz and de Chambrier, 2003).

Phylogenetic interrelationships of the proteocephalidean genera and their higher-level classification are still not sufficiently resolved, despite having been addressed with the use of molecular data over two decades. Analyses of partial sequences of the mitochondrial 16S ribosomal RNA gene (16S rDNA), nuclear ribosomal 18S and 28S rRNA genes (18S rDNA, 28S rDNA) and the associated internal transcribed spacer 2 (ITS2) by Zehnder and Mariaux (1999), Škeříková et al. (1998), de Chambrier et al. (2004c, 2015), Hypša et al. (2005), Scholz et al. (2007) have, however, significantly advanced our understanding of the phylogeny of the order as compared to the pre-molecular era. According to the most recent phylogenetic treatments of the group by de Chambrier et al. (2015), subfamilies Gangesiinae and Acanthotaeniinae, while being paraphyletic assemblages, form the earliest diverging lineages of the Proteocephalidea (de Chambrier et al., 2015). Gangesiinae are parasites of Siluriformes in the Palearctic and Indomalayan geographical regions, Acanthotaeniinae are found in reptiles in the Afrotropic, Indomalayan and Australian regions (de Chambrier et al., 2004c).

Sandonella sandoni (Lyndsdale, 1960), the type and only species of the subfamily Sandonellinae from osteoglossiform fish in Africa, is a molecularly highly diverged lineage, often found among the relatively early diverging groups, forming a sister group to a clade of the monotypic genera *Glanitaenia* de Chambrier, Zehnder, Vaucher & Mariaux, 2004 (Proteocephalinae) and *Paraproteocephalus* Chen in Dubinina, 1962 (Corallobothriinae), both from silurid catfishes in Palearctic region, and *Proteocephalus*-aggregate clade (Fig. 4) found in Holarctic teleosts (de Chambrier et al., 2004c, 2008, 2015).

According to the results of the latest group-wide phylogenetic treatment of the group by de Chambrier et al. (2015), the rest of the relatively more derived diversity of proteocephalideans splits into two main lineages (Fig. 4), both mostly found in siluriform fishes. The first lineage consists of proteocephalideans from Afrotropic siluriforms of the subfamilies Corallobothriinae, Marsypocephalinae and Proteocephalinae together with a clade composed of *Scholzia emarginata* (Diesing, 1850), *Proteocephalus hemioliopteri* de Chambrier & Vaucher, 1997 (both Proteocephalinae) and *Zygobothrium megacephalum* Diesing, 1850 (Zygobothriinae) from Neotropical catfish and remaining representatives of the Corallobothridae from the Nearctic region, parasites of channel catfish (Ictaluridae). Second lineage, and also the most derived and species-rich one, of the Proteocephalidea (Fig. 4) includes vast majority of the Neotropical proteocephalideans (mainly parasites from siluriform fishes) as well as representatives of the polyphyletic genus *Ophiotaenia*, parasites of snakes and amphibians from different zoogeographical regions. Interrelationships within this latter clade are still highly unresolved (de Chambrier et al., 2015).

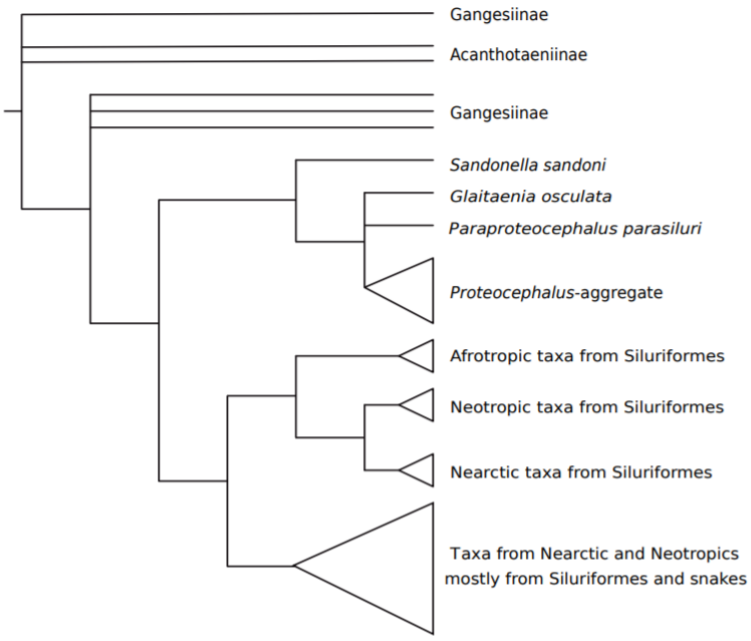


Fig. 4 Phylogenetic relationships of the proteocephalidean cestodes based on 28S rDNA (simplified based on de Chambrier et al., 2015).

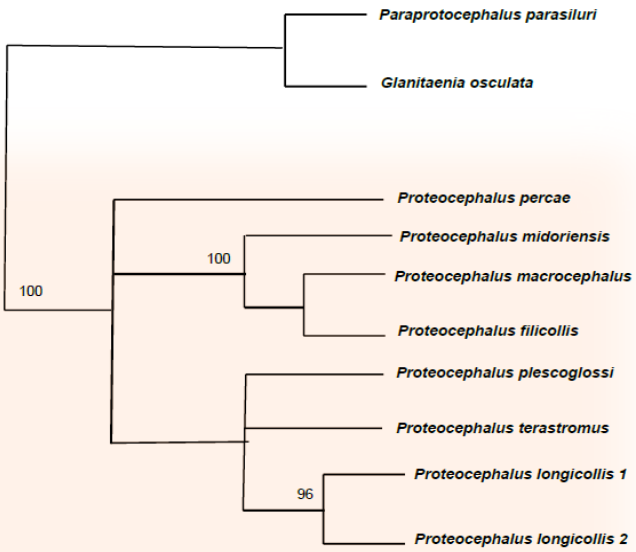
Systematics of the Proteocephalidea predating the molecular data, based on the morphological characteristics including the scolex morphology or the relative position of genital organs in relation to the longitudinal musculature are of limited value. In contrast, new

characters, such as the relative ovary size, egg structure and the pattern of the uterus development were proposed to be potentially useful for defining subgroups supported by the molecular data (de Chambrier et al., 2015).

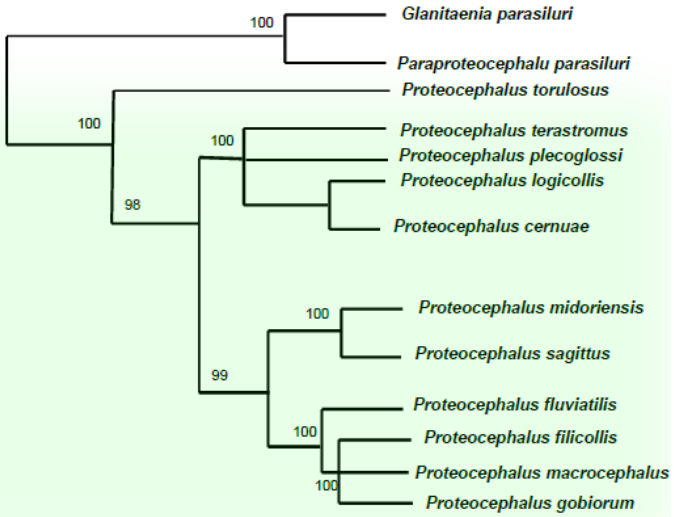
1.4 *Proteocephalus*-aggregate sensu de Chambrier et al. (2004)

The *Proteocephalus*-aggregate is a monophyletic group of tapeworms of teleosts of the Holarctic region. Representatives of the *Proteocephalus*-aggregate clade have similar morphology, life cycles and host specificity, but they do not seem to form lineages specific to the Palearctic or Nearctic geographic regions (Scholz and Hanzelová, 1998; de Chambrier et al., 2004c, 2015; Hypša et al., 2005; Scholz et al., 2017). Although the total number of valid species is a subject of ongoing research, Scholz et al. (2007) recognized validity of 14 species from the Palearctic region and few years later, two additional species from North America, *P. fluviatilis* and *P. pinguis*, were added to this group based on the analyses of de Chambrier et al. (2015). Members of the *Proteocephalus*-aggregate group keep generic name based on the fact that *P. ambiguus* (Dujardin, 1845), the type species of the genus (Scholz, 2007; de Chambrier, 2015), belongs to this clade. Previous analyses of the evolutionary histories of the group representatives and their fish hosts did not reveal any apparent congruency between the phylogeny of the fish hosts and species of the *Proteocephalus*-aggregate, thus indicating host-switching might have occurred frequently within this clade (Škeříková et al., 2001; Scholz et al., 2007). Monophyly of species of the *Proteocephalus*-aggregate is supported by their similar morphology in comparison to the remaining members of the paraphyletic genus *Proteocephalus* (see Scholz et al., 2007). However, systematics of members within this group is still insufficiently known and the interrelationships at both the species and genus levels are largely unresolved. The previous molecular phylogenetic studies were based mainly on the 28S rDNA (de Chambrier et al., 2004c, 2015) but also a few studies utilised partial protein-coding genes in addition to the ribosomal rRNA genes (Hypša et al., 2005, Scholz et al., 2007, 2017) also see Fig. 5.

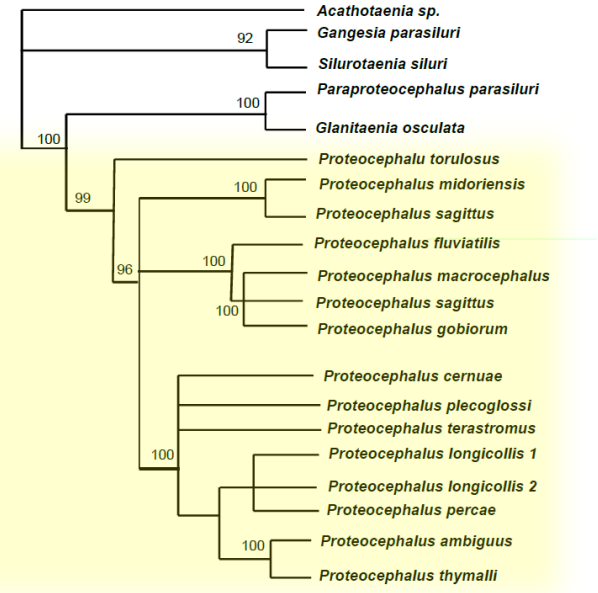
A)


Paraprotocephalus parasiluri
Glanitaenia osculata
Proteocephalus percae
Proteocephalus midoriensis
Proteocephalus macrocephalus
Proteocephalus filicollis
Proteocephalus plescoglossi
Proteocephalus terastromus
Proteocephalus longicollis 1
Proteocephalus longicollis 2
100
100
96

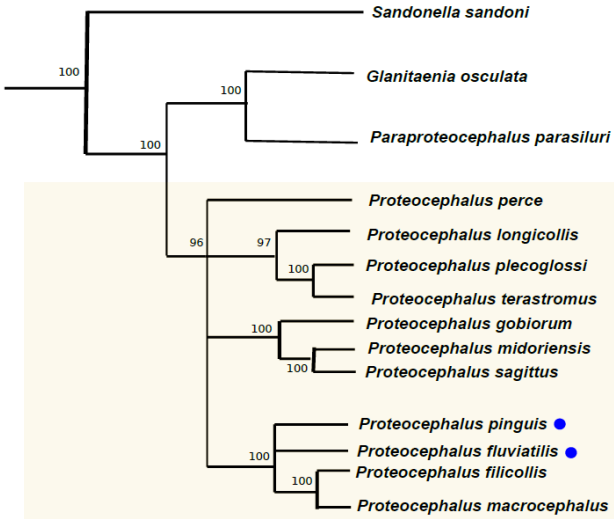
B)


Glanitaenia parasiluri
Paraproteocephalu parasiluri
Proteocephalus torulosus
Proteocephalus terastromus
Proteocephalus plecoglossi
Proteocephalus logicollis
Proteocephalus cernuae
Proteocephalus midoriensis
Proteocephalus sagittus
Proteocephalus fluviatilis
Proteocephalus filicollis
Proteocephalus macrocephalus
Proteocephalus gobiorum
100
100
100
98
100
99
100
100

C)


Acathotaenia sp.
Gangesia parasiluri
Silurotaenia siluri
Paraproteocephalus parasiluri
Glanitaenia osculata
Proteocephalu torulosus
Proteocephalus midoriensis
Proteocephalus sagittus
Proteocephalus fluviatilis
Proteocephalus macrocephalus
Proteocephalus sagittus
Proteocephalus gobiorum
Proteocephalus cernuae
Proteocephalus plecoglossi
Proteocephalus terastromus
Proteocephalus longicollis 1
Proteocephalus longicollis 2
Proteocephalus percae
Proteocephalus ambiguus
Proteocephalus thymalli
92
100
100
99
100
96
100
100
100

D)


Sandonella sandoni
Glanitaenia osculata
Paraproteocephalus parasiluri
Proteocephalus perce
Proteocephalus longicollis
Proteocephalus plecoglossi
Proteocephalus terastromus
Proteocephalus gobiorum
Proteocephalus midoriensis
Proteocephalus sagittus
Proteocephalus pinguis
Proteocephalus fluviatilis
Proteocephalus filicollis
Proteocephalus macrocephalus
100
100
100
96
97
100
100
100
100
100
100

E)

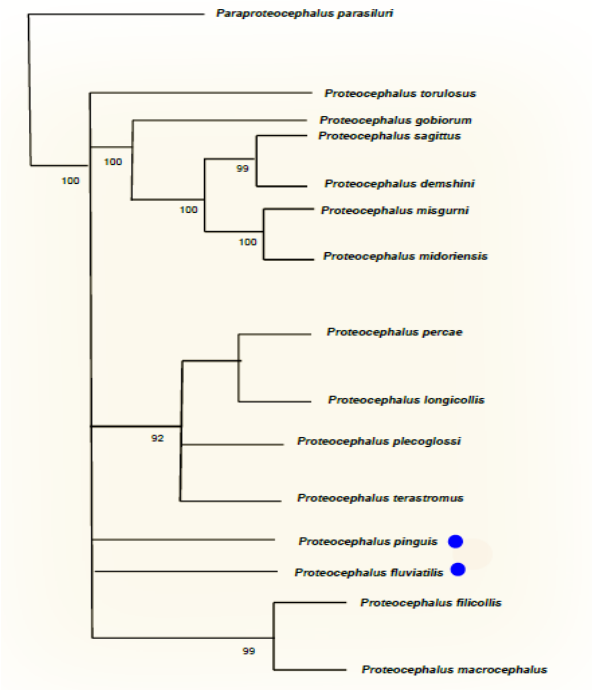


Fig. 5 Molecular phylogenetic estimates of the interrelationships of the Proteocephalus-aggregate redrawn from previous publications: A, analysis of 28S rDNA by de Chambrier et al. (2004c); B, analysis of V4 region of 18S RNA, ITS2, 28S rDNA and 16S rDNA by Hypša et al. (2005); C, analysis based on V4 region of 18S rDNA + 5.8S + ITS2 by Scholz et al. (2007); D, analysis of 28S rDNA by de Chambrier et al. (2015); E, analysis of a 4-gene dataset consisting of 18S + 28S rDNA, rrnL and cox1 by Scholz et al. (2017). Blue dots mark Proteocephalus-aggregate species with Nearctic distribution

1.5 Morphological features of the *Proteocephalus*-aggregate

Morphological features of the *Proteocephalus*-aggregate include various-shaped and well-marked, anapolytic strobila, mostly acraspedote or exceptionally slightly craspedote. Scolex possesses four suckers situated dorsoventrally, two by two. Apical organ can be present or absent. Inner longitudinal musculature is well-developed. Ovary is bilobed, near the posterior margin of proglottids. Vagina tubular, opening anterior, antero-dorsal or dorsal to cirrus-sac. Cirrus-sac is thick-walled opening into small genital atrium. Testes oval to spherical, in one central field (de Chambrier et al., 2004c; Scholz et al., 2007). This clade also displays development of the uterus type 2 (Fig. 3), which is unique to the species of the *Proteocephalus*-aggregate and most representatives of the former Monticelliidae sensu Rego,

1994 (de Chambrier et al., 2004c, 2015). In immature proglottids, the uterine stem creates an undifferentiated longitudinal concentration of chromophlilic cells; in premature proglottids the uterine stem evolves dense elongate digitations that could ramify or develop into the central field of undifferentiated chromophilic cells. In mature proglottids there is a sequential appearance and extension of a lumen from the base to the apex, which consists of chromophilic cells. In gravid uterus, diverticula occupy almost the entire width of proglottids and includes many chromophilic cells (de Chambrier et al., 2004c, 2015).

2 AIMS OF STUDY

1) Selection and amplification protocol optimisation of an alternative molecular marker to supplement molecular phylogenetic analyses of proteocephalideans with focus on representatives of the *Proteocephalus*-aggregate group.
2) Molecular characterisation of proteocephalideans from the Nearctic region collected during the field trips between 2012 and 2017.
3) Comparing the evolutionary history of the Nearctic species of the *Proteocephalus*-aggregate and their definitive fish hosts.

3 MATERIAL AND METHODS

3.1 Origin of the parasite material

Newly sequenced specimens were obtained from colleagues and originate from freshly dissected fish hosts. Table 1 shows a list of specimens sequenced *de novo* including details on their hosts and collection localities. Specimens for molecular analysis were fixed in 70–96% molecular grade ethanol. Table 1 also includes details on the sequences downloaded from GenBank.

Table 1. List of taxa used in the molecular phylogeny. Newly generated sequences are in bold.

Species	Host	Country	28S rDNA	Cox1
OUTGROUP				
Glanitaenia osculata	*Silurus glanis*	Switzerland	KX768937	KX768943
Paraproteocephalus parasiluri	*Silurus asotus*	Russia	KX768938	n/a
INGROUP				
Proteocephalus demshini	*Barbatula toni*	Russia	KX768942	KX768950
Proteocephalus filicollis	*Gasterosteus aculeatus*	United Kingdom	AJ388636	n/a
Proteocephalus fluviatilis	*Micropterus dolomieu*	Japan	KP729390	KX768945
Proteocephalus gobiorum	*Apollonia fluviatilis*	Ukraine	KP729393	KX768944
Proteocephalus longicollis	*Micropterus dolomieu*	Canada	**MN061862**	**MN061850**
Proteocephalus longicollis	*Coregonus clupeaformis*	USA	**MN061863**	**MN061851**

Proteocephalus longicollis	*Sander vitreus*	Wisconsin, USA	**MN061864**	**MN061852**
Proteocephalus longicollis	*Coregonus lavaretus*	Germany	JQ639165	n/a
Proteocephalus luciopercae	*Sander vitreus*	Wisconsin, USA	**MN061853**	**MN061841**
Proteocephalus luciopercae	*Sander vitreus*	Wisconsin, USA	**MN061854**	**MN061842**
Proteocephalus luciopercae	*Sander vitreus*	Ontario, Canada	**MN061855**	**MN061843**
Proteocephalus luciopercae	*Sander vitreus*	Ontario, Canada	**MN061856**	**MN061844**
Proteocephalus macrocephalus	*Anguilla anguilla*	Czech Republic	AJ388609	n/a
Proteocephalus macrocephalus	*Anguilla anguilla*	United Kingdom	EF095261	JQ268552
Proteocephalus midoriensis	*Lefua echigonia*	Japan	AJ388610	n/a
Proteocephalus misgurni	*Misgurnus anguillicaudatus*	Russia	KX768941	KX768949
Proteocephalus percae	*Perca fluviatilis*	Switzerland	AJ388594	KX768947
Proteocephalus pearsei	*Perca flavescens*	USA	**MN061857**	**MN061845**
Proteocephalus pearsei	*Esox niger*	USA	**MN061858**	**MN061846**
Proteocephalus pinguis	*Esox lucius*	USA	KP729395	n/a
Proteocephalus pinguis	*Esox lucius*	Minnesota, USA	**MN061859**	**MN061847**
Proteocephalus pinguis	*Esox lucius*	Minnesota, USA	**MN061860**	**MN061848**
Proteocephalus	*Esox lucius*	Minnesota,	**MN061861**	**MN061849**

pinguis		USA		
Proteocephalus plecoglossi	*Plecoglossus altivelis*	Japan	KX768939	KX768946
Proteocephalus sagittus	*Barbatula barbatula*	Czech Republic	KP729391	KX768948
Proteocephalus tetrastomus	*Hypomesus nipponensis*	Japan	AJ388635	n/a
Proteocephalus torulosus	*Alburnus neretvae*	Bosna and Herzegovina	**Unpublished**	**Unpublished**
Proteocephalus torulosus	*Squalius tenellus*	Bosna and Herzegovina	**Unpublished**	**Unpublished**

3.2 DNA isolation

Total genomic DNA was extracted from ethanol-fixed tissue samples of adult tapeworms. Small pieces of strobila were placed in the 1.5 ml eppendorf tubes and cut into pieces with sterile scissors. Tissue was lysed and genomic DNA purified using commercially available kit E.Z.N.A. Tissue DNA Kit (Omega Bio-tek) according to the manufacturer's instructions. DNA was eluted in 200 µl of elution buffer and stored at -20 °C.

3.3 Design of primers

Molecular characterisation of the representatives of the *Proteocephalus*-aggregate was done through sequencing two molecular targets commonly utilised in systematic studies of tapeworms. Amplification and sequencing of the D1–D3 domains of the 28S rDNA made use of primers published previously see Table 2, complete sequence of the cox1 gene was obtained either using the primers from de Chambrier et al. (2019) or with a set of newly designed primers (Table 2). The newly developed cox1 primers were designed on the basis of three complete sequences of mitochondrial genomes available in 2017 in the Laboratory of Helminthology, Institute of Parasitology, Biology Centre of the Czech Academy of Sciences. All three representative species were newly characterised under the scope of a separate project and included *Proteocephalus luciopercae, P. pinguis* and *P. torulosus*.

Table 2. List of primers Forward (F) and reverse (R), used for 28S rDNA and cox1 amplification and sequencing.

Gene	Primer	Sequence (5´-3´)	Reference
28S rDNA	LSU5 (F)	TAG GTC GAC CCG CTG AAY TTA AGAC	Olson et al. 2003
	1500R (R)	CGA AGT TTC CCT CAG GAT AGC	Olson et al. 2003
	300F (F)	CAA GTA CCG TGA GGG AAA GTT G	Littlewood et al. 2000
	400R (R)	GCA GCT TGA CTA CAC CCG	Littlewood et al. 2000
	900F (FS)	CCG TCT TGA AAC ACG GAC CAA	Lockyer et al. 2003
cox1	B2-16SR (R)	GCA TGA TRC AAA AGG CACA	de Chambrier et al. 2019
	B2-TrpF (R)	TAG ACT AAR TGT TTT CAA AAC A	de Chambrier et al. 2019
	B3-16SR (R)	GCA AAA GGC AAR CAA ACC TA	**newly designed**
	B3-TrpF (F)	GTT TTC AAA ACA TTC AGC GGY	**newly designed**

3.4 PCR amplification

DNA fragment corresponding to the D1–D3 domains of the 28S rDNA was amplified by PCR using universal tapeworm primers LSU5 and 1500R (Table 2) originally published in Olson et al. (2003). To amplify the complete sequence of the mitochondrial cox1, primer pairs B2-TrpF and B2-16SR (de Chambrier et al., 2019) or B3-TrpF and B3-16SR were used (Table 2). In case of both molecular targets, 25µl PCR reactions consisted of 5 µl of 5x GoTaq Flexi Buffer (Promega), 2 mM $MgCl_2$, 0.25 µg BSA (Promega), 200 µM dNTPs (Promega), 200 nM of each of forward and reverse primers, 0.5 U of GoTaq G2 Flexi DNA Polymerase (Promega) and 1 µl of genomic DNA. Amplification of 28S rDNA and cox1 took place in T100 Thermocycler (Bio-Rad) under the conditions shown on Fig. 6.

PCR results were checked by electrophoresis on 1% TRIS acetate EDTA (TAE) agarose gel stained with GelRed (Biotium) in approximate ratio 1 : 1,000,000. Two µl of each PCR product was mixed with 0.3 µl sample loading buffer (H_2O solution including 0.25% bromphenol blue, 0.25% xylene cyanol, 30% glycerol). Size of the PCR product was estimated according to the GeneRuler 1kb Plus DNA Ladder marker (Fermentas) and documented on Gel Logic 112 Imaging System (Carestream Molecular Imaging).

Single PCR products were purified enzymatically following the protocol of Werle et al. (1994) using 6.7 U of Exonuclease I and 1.5 U of FastAP Thermosensitive Alkaline Phosphatase (both Thermo Scientific). Incubation protocol was 37 °C/40 min, 80 °C/20 min.

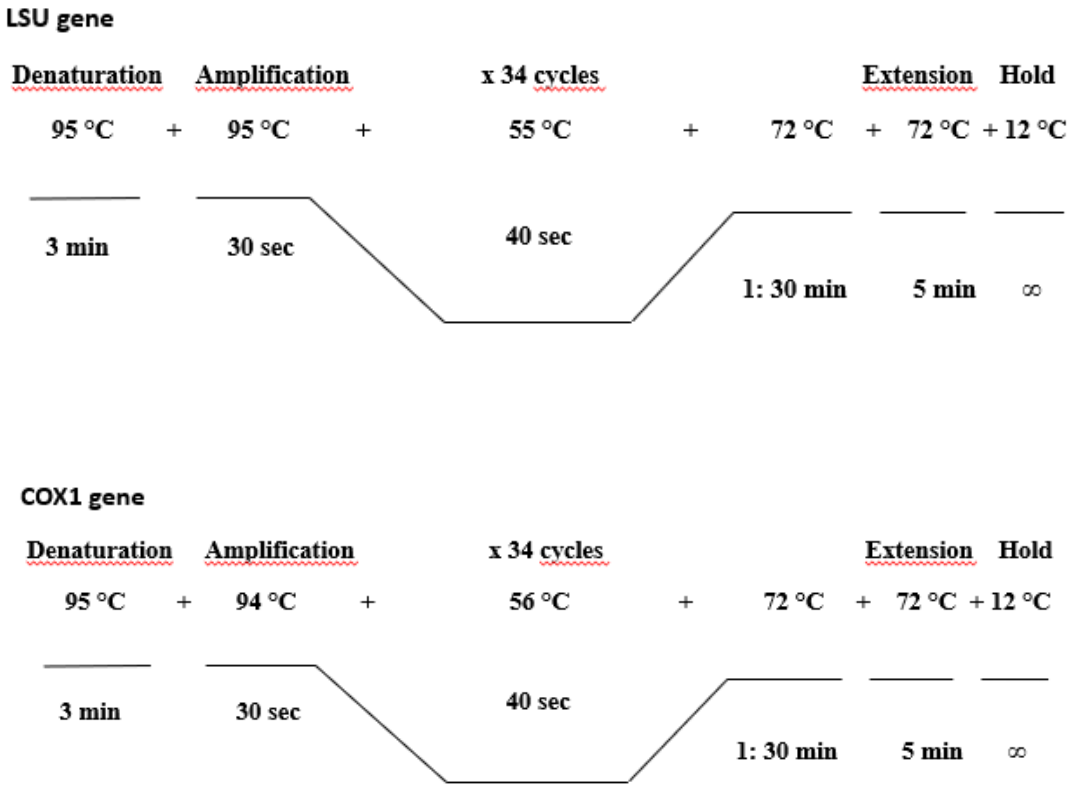


Fig. 6 PCR thermocycling profiles used for amplification of 28S rDNA (above) and cox1 genes (below).

3.5 Sequencing

Purified PCR fragments were sequenced by Sanger method at SEQme s.r.o. (Dobříš, Czechia). PCR products were sequenced from both strands using PCR and sequencing primers (Table 2). Contiguous sequences were assembled and manually corrected in Geneious 9.1.6. (Biomatters). Identity of sequences was checked by Basic Local Alignment Search Tool (BLAST) through website National Center for Biotechnology Information (NCBI; www.ncbi.ncbi.nih.gov/BLAST/).

3.6 Phylogenetic analyses

Newly obtained sequences of both molecular targets were combined with 28S rDNA and cox1 sequences of representatives of the *Proteocephalus*-aggregate tapeworms available from previous studies (de Chambrier et al., 2015 and Scholz et al., 2017). Outgroup taxa selection was informed by the tree published in de Chambrier et al. (2015) and resulted in selection of *Glanitaenia osculata* (28S rDNA, cox1) and *Paraproteocephalus siluri* (28S rDNA) as outgroups, because these two taxa form the closest, clearly separate sister clade of the *Proteocephalus*-aggregate group (de Chambrier et al., 2015).

Multiple sequence alignments of 28S rDNA and cox1 dataset were created using the E-INS-i algorithm of the program MAFFT 7.017 (Katoh and Standley, 2013) implemented in Geneious. All datasets were analysed under the maximum likelihood (ML) criterion using the on-line version of the program IQ-TREE 1.6.5. (Trifinopoulos et al., 2016). The following datasets and their respective partitioning schemes were analysed: 28S rDNA (single partition), cox1 (single partition), concatenated dataset of 28S rDNA + cox1 (two partitions). IQ-TREE was also used to estimate the best-fit model of sequence evolution for each of the partition using the built-in ModelFinder algorithm. Models were chosen according to the corrected Akaike Information Criterion and were as follows: GTR + F + I + G4 (28S rDNA); GTR + F + I + G4 (cox1); GTR + F + I + G4: 28S rDNA, GTR + F + I: cox1 (concatenated dataset). Nodal supports were estimated through running 1,000 nonparametric standard bootstrap replicates within IQ-TREE.

In addition to the originally characterised data consisting of the 28S rDNA and cox1 sequences described above, phylogenetic analyses of two extended datasets (each including five representatives of the *Proteocephalus*-aggregate plus outgroup species) were run. The data used for construction of the two extended datasets originated from the complete sequences of nuclear ribosomal RNA operons and mitochondrial genomes obtained within a sequencing project undertaken in the Laboratory of Helminthology, Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, independent of this thesis. Complete sequences of 6 nuclear 18S + 28S rDNA genes (4 unpublished; Brabec et al., in prep.) formed the first extended dataset. Complete sequences of 12 concatenated mitochondrial protein-coding genes (PCGs) of 4 representatives (*P. longicollis*, *P. luciopercae*, *P. pinguis* and *P. torulosus*; Brabec et al., unpublished) plus 2 partial mitochondrial PCG data sets (consisting

of 3 PCGs of *P. macrocephalus* and *Acanthotaenia* sp.) were used for construction of the second extended dataset. Both extended dataset were created using the E-INS-i algorithm of the program MAFFT 7.017 implemented in Geneious. The rRNA and mtDNA extended datasets were partitioned into 2 and 12 partitions, respectively, according to the PCG boundaries. The best-fitting models of sequence evolution and the partitioning scheme were chosen according to the corrected Akaike Information Criterion and were as follows: rRNA operon: TIM3 + F + I: 18S rDNA + 28S rDNA; mtDNA: TIM + F + I + G4: cox3, GTR + F + I + G4: cob + cox1, GTR + F + G4: nad4L + nad3, TVM + F + I + G4: nad4, GTR +F + I: atp6, TVM + F + G4: nad2, TIM + F + I+ G4: nad1, TIM2 + F + G4: cox2, TIM3 + F + I + G4: nad6, GTR + F + I + G4: nad5. ML trees were estimated in IQ-TREE.

4 RESULTS

4.1 28S rDNA phylogenetic analysis

The 28S rDNA data matrix subjected to phylogenetic analysis consisted of 14 newly generated and 16 previously characterised sequences retrieved from GenBank reaching the total aligned length of 1,495 characters. Length of individual sequences subjected to alignment ranged from 1,260 to 1,495 bp. Out of the 1,338 unambiguously aligned nucleotide sites, 107 were parsimony-informative and 1,172 represented constant sites. Resulting topology obtained from ML analysis of the 28S rDNA data is shown on Fig. 7.

Individual species of *Proteocephalus*-aggregate represented by more than one specimen (i.e., *P. longicollis*, *P. luciopercae*, *P. macrocephalus*, *P. pearsei*, *P. pinguis* and *P. torulosus*) formed monophyletic lineages, mostly supported by relatively high nodal bootstrap support values with the only exception of the lineage of *P. longicollis*. *Proteocephalus demshini*, *P. misgurni* and *P. sagittus*, all parasites of Palearctic loaches, formed a well-resolved group, with a strong sister-lineage relationship to *P. gobiorum*. This five-species group then formed the earliest-branching clade of the 28S rDNA tree, but without any statistical support from the bootstrap analysis. Another lineage receiving high nodal supports from the 28S rDNA data was *P. filicollis* + *P. macrocephalus* that together formed a well-supported clade with *P.*

pearsei. Supports for the internal nodes defining relationships between the remaining species-level lineages were generally weaker.

The ML tree recognized *P. longicollis* and *P. percae* as sister species, uniting them in a clade with two other species, *P. plecoglossi* and *P. tetrastomus*. *Proteocephalus pinguis* and *P. fluviatilis* formed together a sister lineage to the clade consisting of *P. filicollis*, *P. macrocephalus* and *P. pearsei*. Specimens of *P. luciopercae* and *P. torulosus* formed separate lineages without a clear position within the *Proteocephalus*-aggregate group.

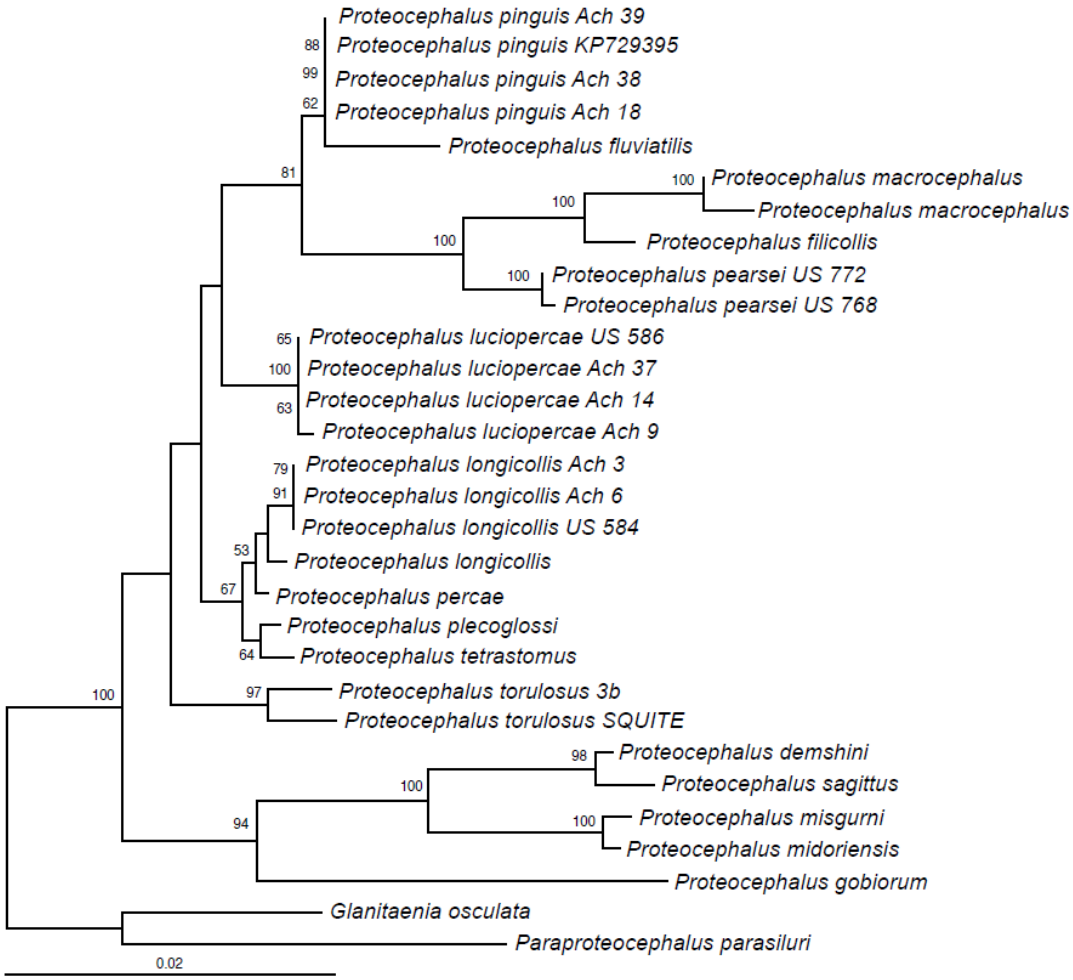


Fig. 7 Resulting ML phylogeny based on the 28S rDNA gene dataset constructed in the program IQ-TREE using the GTR + F + I + G4 model. Nodal support based on standard bootstrapping (1000 repetitions). Bootstrap support values below 50 not shown.

4.2 Cox1 phylogenetic analysis

The cox1 sequence matrix consisted of 23 (14 newly obtained) sequences of 1,551 unambiguously aligned nucleotide positions. Total length of individual sequences ranged from 441 to 1,551 bp (8 sequences were short). Out of the 1,551 aligned characters, 529 were parsimony-informative and 938 constant. The topology resulting from the ML analysis of the nucleotide dataset is shown on Fig. 8.

All individual species of the *Proteocephalus*-aggregate represented by more than one specimen (i.e., *P. longicollis, P.luciopercae, P. pinguis*, *P. pearsei* and *P. torulosus*) formed monophyletic linages supported by relatively high bootstrap values. Analogously to the analysis of the 28S rDNA data, cox1 data failed to confidently resolve the internal branching pattern within the *Proteocephalus*-aggregate group. The internal nodes of the cox1 tree received relatively lower bootstrap support values than the internal nodes of the 28S rDNA tree, leaving the interrelationships between individual species dubious. Only one group, parasites of Palearctic loaches including *P. demshini*, *P. misgurni* and *P. sagittus*, formed a relatively well-resolved group. This is a finding analogous to the one found on the 28S rDNA ML tree: *P. demshini, P. sagittus* and *P. misgurni* also formed a separate, well-resolved group, however, without a close relationship to *P. gobiorum* that formed the most basal lineage on the cox1 tree. A clade consisting of all representatives of *P. longicollis* and *P. percae* formed a sister lineage to *P. plecoglossi*, a grouping also present on the 28S rDNA tree. The ML tree based on cox1 data also resolves *P. pinguis* as a sister group of *P. fluviatilis*, but without any statistical support from the bootstrap analysis. Phylogenetic position of the specimens of *P. luciopercae, P. macrocephalus* and *P. pearsei* remained unsupported by bootstrap support values.

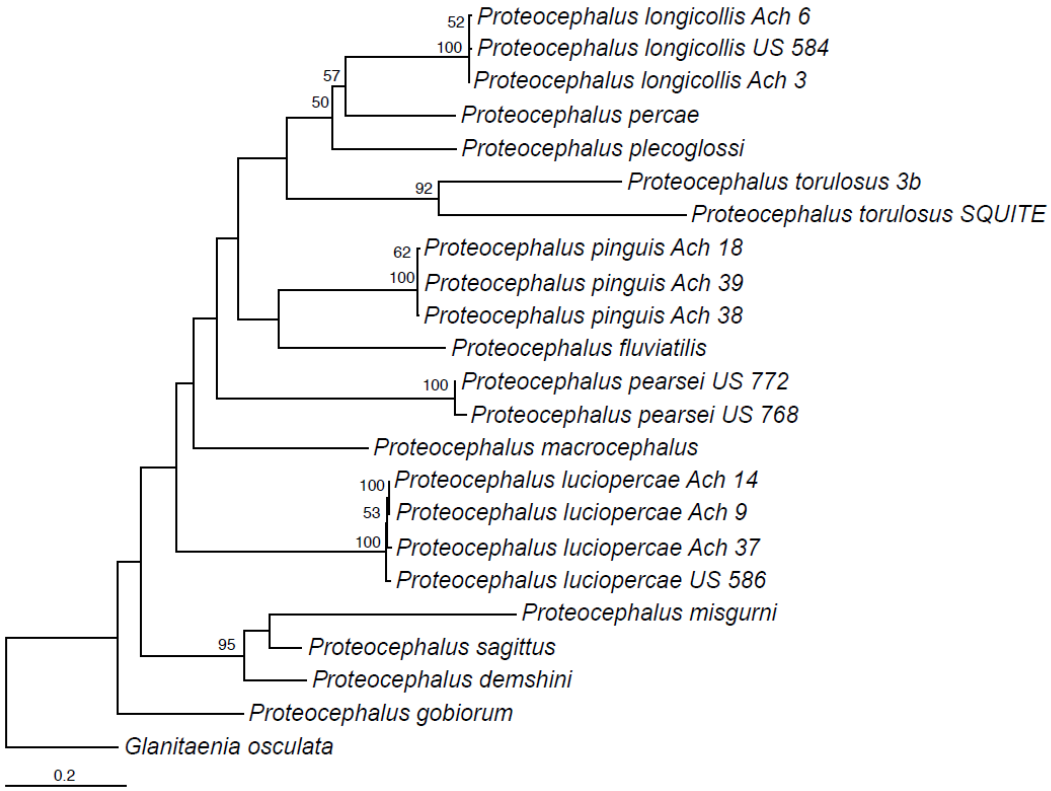
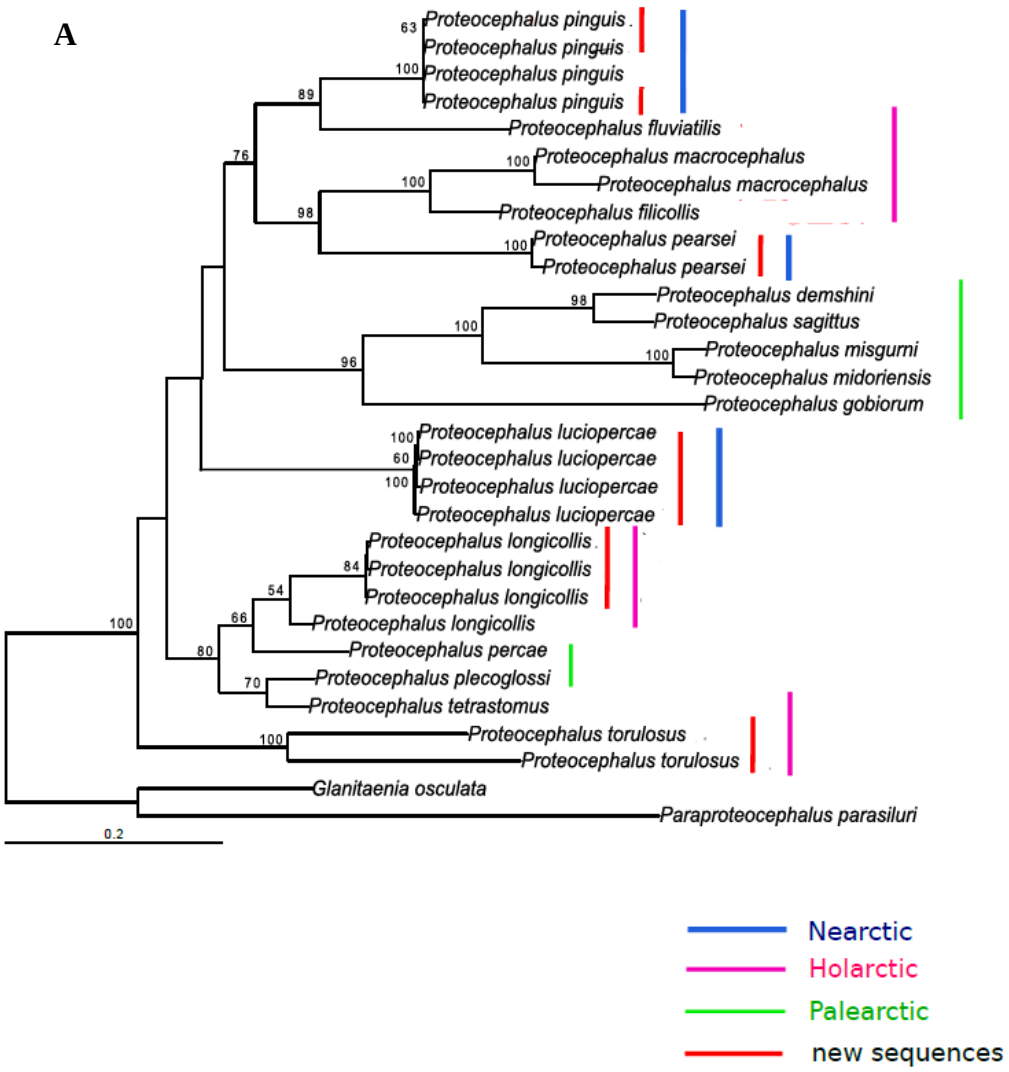


Fig. 8 Resulting ML phylogeny based on the cox1 gene dataset constructed in the program IQ-TREE using the GTR + F +I + G4 model. Nodal support based on standard bootstrapping (1000 repetitions). Bootstrap support values below 50 not shown.

4.3 Phylogenetic analysis of the concatenated dataset

The concatenated matrix consisting of 28S rDNA and cox1 sequences had total length of 2,889 characters. Topology resulting from ML analysis is shown in Fig. 9. All species of the *Proteocephalus*-aggregate except *P. longicollis* (i.e., *P. luciopercae*, *P. macrocephalus*, *P. pinguis*, *P. torulosus*) formed well-supported monophyletic lineages. All parasites of Palearctic loaches formed a well-resolved group with a sister-lineage relationship to *P. gobiorum*, supported by a high bootstrap support value. Other well-supported groups include *P. macrocephalus* + *P. filicollis* having a strong sister-lineage relationship to *P. pearsei*. Another relatively well-supported clade was the one uniting *P. pinguis* and *P. fluviatilis*. Support for internal nodes defining interrelationships within the remaining species-level lineages was largely unresolved, despite concatenation of the 28S rDNA and cox1 data.

Proteocephalus pinguis + *P. fluviatilis* formed a sister-lineage to the clade consisting of *P. filicollis*, *P. macrocephalus* and *P. pearsei* on the concatenated ML tree. As on the 28S rDNA tree, *P. longicollis* and *P. percae* formed a group, sister to a clade of *P. plecoglossi* and *P. tetrastomus*. The basal lineage was formed by *P. torulosus*, but unsupported by bootstrap support values.



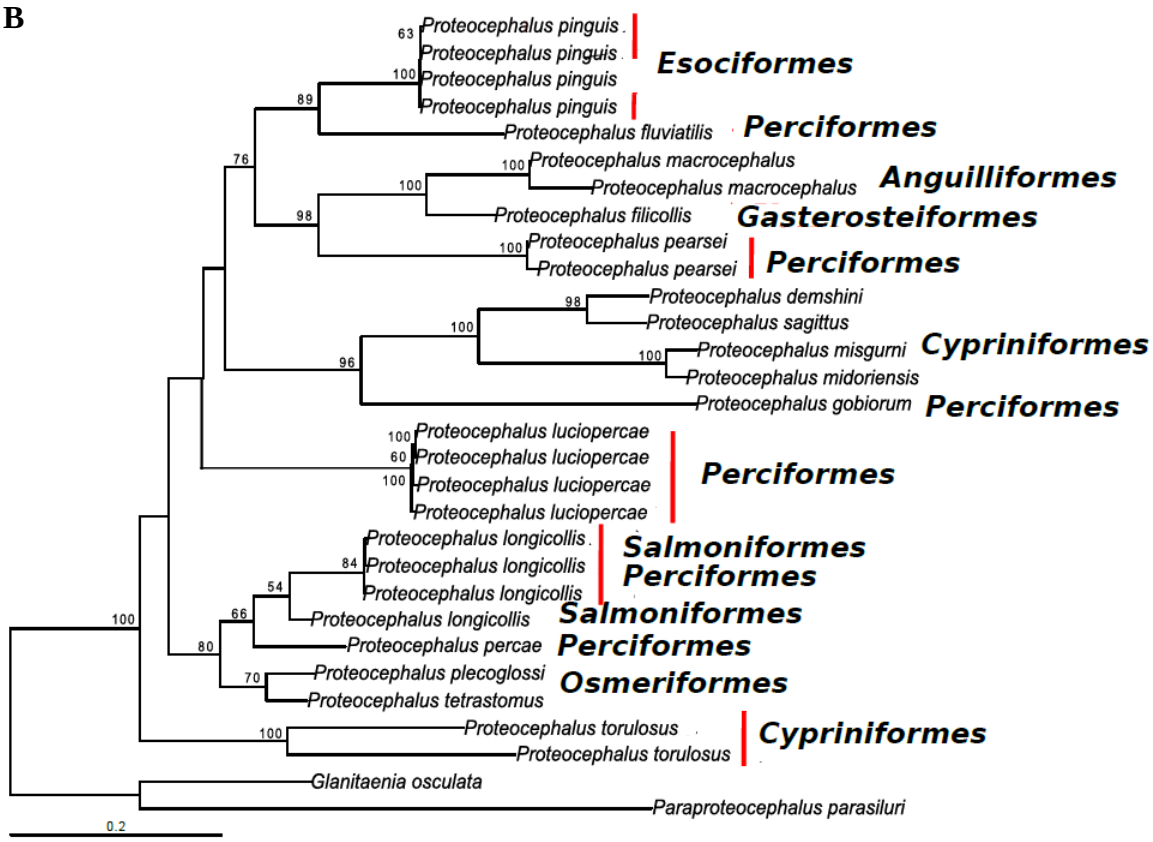


Fig. 9 Resulting ML phylogeny based on the concatenated 28S rDNA + cox1 gene dataset constructed in the program IQ-TREE using the GTR + F + I + G4 (28S rDNA) and GTR + F + I (cox1) models corresponding to the individual partitions. Nodal support based on bootstrapping (1000 repetitions). Bootstrap support values below 50 not shown. Grey dots depict the absence of the apical sucker. A – Geographical distribution B – Host specify.

4.4 Phylogenetic analyses of the extended datasets

The extended rRNA operon dataset consisted of complete 18S and 28S rDNA genes per each of the five representatives of the *Proteocephalus*-aggregate plus an outgroup taxon and had total number of 6,051 characters. The second extended dataset consisting of 12 mitochondrial PCGs had total length of 9,951 characters and was built from 4 complete sets of unpublished mitochondrial PCGs and 2 previously characterised partial mitochondrial

genome sequences corresponding to 3 PCGs (nad1, nad3 and cox1). Topologies resulting from the ML analyses of the extended datasets are shown on Figs. 10 and 11.

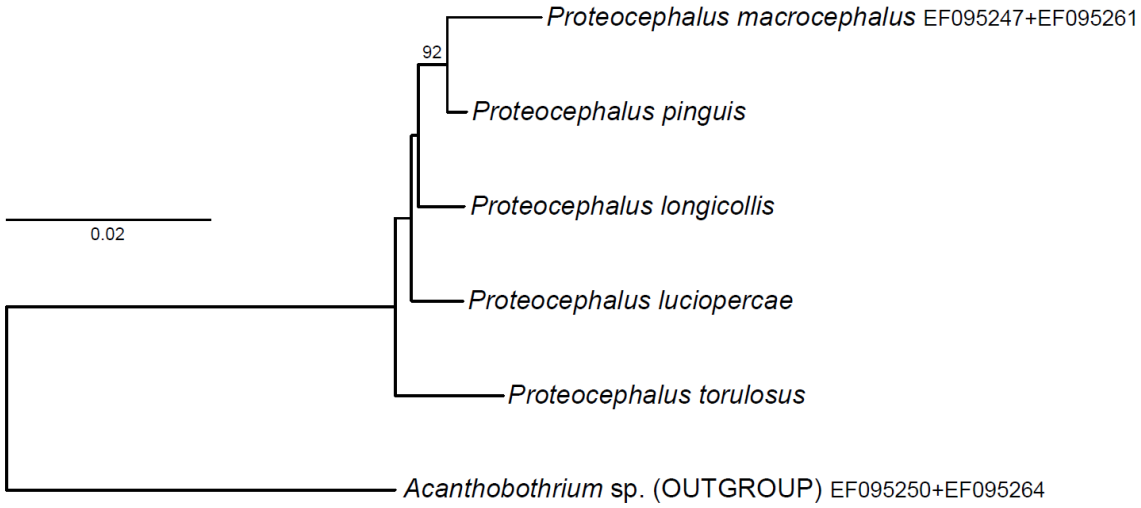


Fig. 10 Resulting ML phylogeny based on the concatenated dataset of complete 18S and 28S rDNA genes constructed in IQ-TREE. Nodal support based on standard bootstrapping (1000 repetitions). 18S+28S rDNA genes were analysed as a single partition under TIM3 + F + I model. Bootstrap support values below 50 not shown.

ML estimates based on both extended datasets (18S + 28S rDNA, mitochondrial PCGs) recovered *P. macrocephalus* and *P. pinguis* as a well-supported clade and *P. torulosus* as a basal taxon, however, without statistical support in neither of the two analyses. The relative position of the remaining ingroup taxa, i.e., *P. longicollis* and *P. luciopercae*, differed between the two analyses, with rDNA recovering *P. longicollis* as a sister-lineage to *P. macrocephalus* + *P. pinguis*, and mitochondrial data resolving *P. longicollis* and *P. luciopercae* as a monophyletic clade.

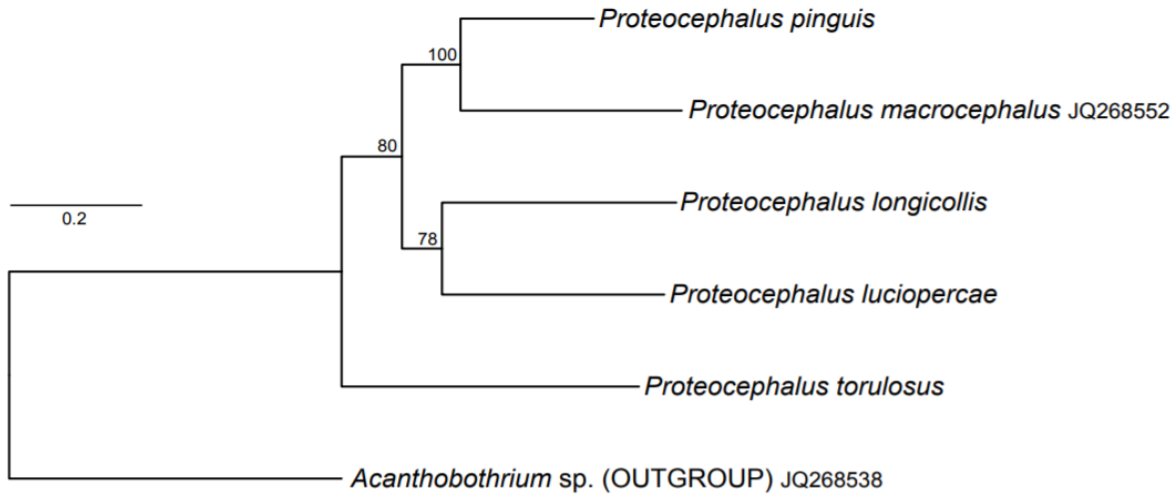


Fig. 11 Resulting ML phylogeny based on the concatenated dataset of mitochondrial PCGs constructed in IQ-TREE. Nodal support based on standard bootstrapping (1000 repetitions). Dataset was divided into 10 partitions (see Methods for details). Bootstrap support values below 50 not shown.

5 DISCUSSION

Proteocephalus-aggregate has been considered to represent a monophyletic lineage, one of several lineages originally belonging to the polyphyletic genus *Proteocephalus*, since Zehnder and Mariaux (1999) studied phylogenetic relationships of the order Proteocephalidea using molecular data for the first time. Since that time, molecular-analyses concerning the *Proteocephalus*-aggregate clade were published in several publications (de Chambrier et al. 2004c, 2015; Hypša et al., 2005; Scholz et al., 2007, 2017). The current study brings novel molecular data consisting of partial 28S rDNA gene sequences to allow direct comparisons to the previously characterised specimens as well as the complete sequences of mitochondrial cox1 gene characterised for the first time in its entirety, thus increasing the length of the sequence of the few previously published cox1 sequences approximately twice. Phylogenetic position of two of the species, *P. luciopercae* and *P. pearsei*, was thus assessed for the first time; both of these species represent host-specific parasites of percid fishes from Nearctic region (Scholz et al., 2019). In addition to *P. luciopercae* and *P. pearsei* sequences, the current analyses supplement first mitochondrial data represented by the complete cox1 sequences for few additional species of the *Proteocephalus*-aggregate from different groups of fish hosts, including *P. longicollis*, *P. pinguis*, and *P. torulosus*.

In spite of the inclusion of the additional data, the internal phylogeny of the *Proteocephalus*-aggregate clade remains grossly unresolved. However, both of the single-gene phylogenies and the tree resulting from the concatenated dataset of 28S rDNA plus cox1 sequences recognize each of the *Proteocephalus*-aggregate species as a monophyletic lineage and in few cases also support relationships between particular congeneric species with relatively high statistical support values. For example, the phylogenetic analysis of 28S rDNA and the concatenated dataset resolve the clade composed of *P. filicollis*, *P. macrocephalus* and *P. pearsei* as a monophyletic clade. In contrast, the analysis of cox1 does not support this particular phylogenetic scenario, but at the same time, does not support any other alternative scenario with high bootstrap support values. The 28S rDNA and the concatenated trees also further recognize a clade of *P. pinguis* and *P. fluviatilis* as a sister to the clade consisting of *P. filicollis*, *P. macrocephalus* and *P. pearsei*. Interestingly, both of the extended dataset analyses based on the complete sequences of the nuclear ribosomal RNA operon and the mitochondrial genomes of few representatives of the *Proteocephalus*-aggregate favour a close

relationship between *P. pinguis* and *P. macrocephalus*. However, it has to be noted here that the extended datasets consist of only very few taxa and that the closest relatives of *P. pinguis* and *P. macrocephalus* as seen on the single-gene trees, including *P. filicollis*, *P. fluviatilis* and *P. pearsei*, are completely missing in these analyses. It would be interesting to see, for example, if the extended mitochondrial PCG dataset would support the sister-clade relationship between *P. fluviatilis* and *P. pinguis* with a higher statistical support values than the ones observed on the single-gene trees.

All of the current analyses also favour close relationships of *P. longicollis* and *P. percae* and their sister-clade relationship to a clade of *P. plecoglossi* plus *P. tetrastomus* (the latter being based on 28S rDNA data only). The statistical support for this four-species clade is slightly higher in the concatenated analysis relative to the single-gene analyses. *Proteocephalus fluviatilis* is a parasite of bass with Holarctic distribution and *P. pinguis* utilises pikes (*Esox lucius*) from North America. The previous papers by de Chambrier et al. (2015) and Scholz et al. (2017) did not resolve their relationship and the tree presented in Scholz et al. (2019) supports this scenario by a relatively low bootstrap value. Both of these specimens also show affinity to the well-supported clade consisting of Nearctic *P. pearsei*, a parasite of yellow perch, and *P. filicollis* and *P. macrocephalus*, circumboreal parasites of three-spined sticklebacks and eels, respectively. Studies by Scholz et al. (2007, 2017 and 2019) also suggested that *P. percae* is a closely related species to *P. longicollis*, parasites of percid and salmoniform fishes, respectively, from the Holarctic region. Both species share some morphological features (e.g., presence of a well-developed, ring-like vaginal sphincter, thick-walled and relatively long cirrus sac, a vestigial apical sucker, etc. – see Scholz and Hanzelová, 1998). However, their phylogenetic relationship never received high nodal support values and should not be considered confident (Scholz and Hanzelová, 1998, 1999; Scholz et al., 2007, 2017, 2019; Škeříková et al., 2001).

Identity of the earliest-diverging lineage of the *Proteocephalus*-aggregate group is still not clear and the internal nodes close to the root of the group are never resolved with statistical support from bootstrap analyses. While the current analysis based on cox1 shows *P. gobiorum* as a basal group, 28S rDNA data find species of the *Proteocephalus*-aggregate from loaches (*P. demshini*, *P. gobiorum*, *P. midoriensis*, *P. misgurni* and *P. sagittus*) as the basal taxon. Concatenated analysis then supports *P. torulosus* as the earliest-diverging species of the *Proteocephalus*-aggregate, a scenario which is in agreement with the results of many previous studies (e.g., Škeříková et al., 2001; Hypša et al., 2005; Scholz et al., 2007, 2017).

Considered from the host spectrum, species of *Proteocephalus* are abundant parasites of North American and European freshwater fishes (Hoffman, 1999), but the knowledge about their host-specificity, host co-evolution, distribution and ecology is still not fully observed. The *Proteocephalus*-aggregate group is also unique among proteocephalideans in its specificity to other fish groups than siluriforms. While the majority of proteocephalideans parasitized siluriform fishes, species within this clade are hosted by other groups of fish. Hypša et al. (2005) suggested that phylogenetic relationships among Holarctic species reflect a probable host-switch from siluriforms to other fish groups more common in the Holarctic region rather than a co-evolution. There is no indication of co-evolution because closely-related parasites are hosted by phylogenetically relatively unrelated hosts (Škeříková et al., 2001; de Chambrier et al., 2004c; Hypša et al., 2005). For example, *P. macrocephalus*, parasite of the Anguilliformes, one of the earliest-branching groups of bony fishes, forms a clade with *P. filicollis* from the Gasterosteidae, one of the highly derived clades among teleosts (Near et al., 2013). Newly collected 28S rDNA data of representatives of *P. longicollis* collected in North America suggest a possible presence of a genetic distance between the North American lineage and the lineage from Europe, the latter represented by a single specimen from *Coregonus lavaretus* characterised previously. Previous phylogenetical analyses also showed presence of a genetic difference between representatives of *P. longicollis*. For example, de Chambrier el al. (2004c) characterised two samples from Europe from one genus of fish host – *Coregonus pollan* and *Coregonus* sp. (Salmoniformes). In another study by Scholz et al. (2007), *P. longicollis* specimens from European *Coregonus albula* and *Coregonus widergreni* (Salmoniformes) also formed two genetically distinct lineages. Specimens of *P. longicollis* in the current study originate from a wide range of unrelated definitive hosts including *Coregonus clupeaformis* and *Coregonus lavaretus* (Salmoniformes), *Micropterus dolomieu* and *Sander vitreus* (Perciformes), and different localities in the Holarctic region. Wide range of hosts and different localities can reflect why isolates of one specimen are genetically dissimilar. Maybe more than one species is presented and more data from additional host species and geographical localities along with detailed focus on the morphology will be required to assess if *P. longicollis* represents a genetically highly diversified species or a complex of species as discussed in Scholz et al. (2007).

Current phylogenetic analyses show that *P. pearsei* from the American yellow perch is not closely related to *P. percae*, a parasite of the European perch. Both species are also able to infect other hosts the perciform fishes, including *P. pearsei* from *Esox lucius* in the current

study that may represent a postcyclic host (Scholz and Hanzelová, 1998). There are two specimens of *P. torulosus* in the current study, one from *Alburnus neretvae* (Cypriniformes) and second from *Squalius tenellus* (Cypriniformes). While both of the fish hosts are endemic to Bosna and Herzegovina and Croatia, the two isolates of *P. torulosus* display very high genetic variability. *Proteocephalus torulosus* is reported from many species of cyprinid fishes (e.g. *Abramis, Barbus*) (Scholz et al., 2003) and more sampling is needed for comparing genetic variability of this species. Other species of the *Proteocephalus*-aggregate display much narrower host specificity. For example *P. fluviatilis* from centrarchids, especially smallmouth and largemouth bass in North America, *P. pearsei* from *Perca flavescens*, or *P. luciopercae* from *Sander vitreus* or *S. canadensis* (Scholz et al., 2019).

Distribution of important morphological characters on the resulting phylogenetic trees in this thesis suggests, in spite of the lack of statistical support for the internal nodes, that some of the morphological features are homoplastic and evolved independently within the *Proteocephalus*-aggregate group. For example, both *Proteocephalus*-aggregate sister-lineage representatives *Glanitaenia osculata* and *Paraproteocephalus parasiluri* possess a well-developed apical sucker (e.g., Scholz et al., 2007; de Chambrier et al., 2015), a morphological characteristic also present in most of the *Proteocephalus*-aggregate species but absent in *P. torulosus*, *P. luciopercae*, *P. gobiorum* and the clade from Palearctic loaches including *P. demshini*, *P. misgurni*, *P. midoriensis* and *P. sagittus*. In agreement with previous studies (e.g., Zehnder and Mariaux, 1999; Scholz et al., 2007, 2017) the close phylogenetic position of *P. longicollis* and *P. percae* based on molecular data is in accordance with their similar morphology including the shape of the scolex and apical sucker or the well-developed vaginal sphincter (Scholz and Hanzelová, 1998). The phylogenetic proximity of *P. filicollis* and *P. macrocephalus* is also supported by common morphological features, such as the presence of a vestigial (muscular) apical sucker, absence of the longitudinal tegumental wrinkles or the weakly developed longitudinal musculature (Škeříková et al., 2001).

Phylogenetic relationships of the *Proteocephalus*-aggregate remain not fully understood and future efforts might focus on collecting fresh parasite material representing a few of the taxa still not represented in the 28S rDNA and cox1 data, especially *P. ambiguus* (type species of the genus *Proteocephalus*) from Gasterosteiformes, *P. thymalli* (Annenkova-Chlopina, 1923) from Salmoniformes or *P. cernuae* (Gmelin, 1790) from Perciformes. Moreover, it would be interesting to collect and molecularly characterise multiple isolates per each of the *Proteocephalus*-aggregate species. Most notably the species with a wide Holarctic

distribution, including *P. macrocephalus*, *P. longicollis* and *P. torulosus*, would benefit from a much wider sampling to better understand their relatively high intraspecific genetic variability that becomes apparent with the analyses of additional molecular data. It is hoped that this thesis will stimulate larger focus on research of freshwater parasites in the Nearctic region, especially for molecular phylogenetics.

6 CONCLUSION

In the present study, interrelationships of the Holarctic cestodes of the *Proteocephalus*-aggregate group were examined using newly collected cestode specimens and sequences of partial 28S rDNA and complete cox1 genes. Phylogenetic trees were estimated based on both single gene datasets as well as the concatenated dataset. While all of the newly represented species of the *Proteocephalus*-aggregate formed monophyletic lineages, mutual interrelationships between the individual species of the group remained largely dubious, due to the lack of statistical support for majority of the internal nodes of the *Proteocephalus*-aggregate trees inferred from individual or concatenated gene datasets. Despite neither of the two genetic loci used in this thesis seemed to be highly informative to confidently resolve the interrelationships of the *Proteocephalus*-aggregate, the species from Nearctic region did not seem to share a common evolutionary history. Further data will be needed to confirm this scenario.

This study extends previous molecular data for with 28 new sequences of five species of *Proteocephalus*-aggregate. Phylogenetic position of two of the species from North America, *P. luciopercae* and *P. pearsei*, was analysed for the first time. Newly designed cox1 primers from this study can be used in future studies to characterise further specimens of proteocephalidean tapeworms, including taxa outside of the *Proteocephalus*-aggregate group. This thesis contributes to the knowledge of the taxonomy of the North American freshwater fish cestode parasites and brings data that will be useful in future research projects aiming to fill existing gaps in our knowledge of the parasite fauna of the North America.

7 REFERENCES

Ash L. R., Orihel T. C. 2007: Ash & Orihel's Atlas of Human Parasitology. *American Society of Clinical Pathologists Press*: 540 pp.

Baily G., Garcia H. H. 2014: Other Cestode Infections: Intestinal Cestodes, Cysticercosis, Other Larval Cestoda Infections. *Manson´s Tropical Infectious Diseases* 23. Saunders Ltd: 1360 pp.

Caira J. N., Jensen K. 2017: Planetary Biodiversity Inventory (2008-2017): Tapeworms from Vertebrate Bowels of the Earth. *Yurchak printing*, Inc., Landisville, Pennsylvania. University of Kansas, Natural History Museum 25: 463 pp.

Caira J. N., Jensen K., Waeschenbach A., Olson P., Littlewood D. T. J. 2014: Orders out of chaos – molecular phylogenetics reveals the complexity of shark and stingray tapeworm relationships. *International Journal of Parasitology* 44: 55-73.

Cañeda-Guzmán I., de Chambrier A., Scholz T. 2001: *Thaumasioscolex didelphis* n. gen and n. sp. (Eucestoda: Proteocephalidae) from black-eared opossum Didelphis marsupialis from Mexico, the first proteocephalidean tapeworm from mammal. *Journal of Parasitology* 87: 639-647.

de Chambrier A., Mariaux J., Sene A., Mahmoud Z.N., Scholz T. 2008: *Sandonella sandoni* (Lynsdale, 1960) Enigmatic and morphologically unique cestode parasitic in the osteoglossiform fish Heterotis niloticus in Africa. *Journal of Parasitology* 94: 202-211.

de Chambrier Al., Mariaux J., Scholz T., Kuchta R. 2017: Onchoproteocephalidea I. Caira, Jensen, Waeschenbach, Olson and Littlewood,2004. Planetary Biodiversity Inventory (2008-2017): Tapeworms from Vertebrate Bowels of the Earth, Chapter 14. *University of Kansas, Natural History Museum*. 251-277 pp.

de Chambrier A., Scholz T., Brabec J. 2019: Revision of Acanthotaenia von Linstow, 1903 (Cestoda: Proteocephalidae), parasites of monitors (Varanus spp.), based on morphological and molecular data. *Helminthology* 118: 1761-1783.

de Chambrier A., Waeschenbach A., Fisseha M., Scholz T., Mariaux J. 2015: A large 28S rDNA-based phylogeny confirms the limitations of established morphological characters for classification of proteocephalidean tapeworms (Platyhelminthes, Cestoda). *Zoo Keys* 500: 25-59.

de Chambrier A., Zehnder M. P., Vaucher C., and Mariaux J. 2004c: The evolution of the Proteocephalidea (Platyhelminthes, Eucestoda) based on an enlarged molecular phylogeny, with comments on their uterine development. *Systematic Parasitology* 57: 159–171.

Egger B., Lapraz F., Tomiczek S. M., Dessimoz Ch., Girstmair J., Škunca N., Rawlinson K. A., Cameron Ch. B., Beli E., Torado M. A., Gammoudi M., Norena C., Telford M. J. 2018: A transcriptomic-Phylogenomic Analysis of the Evolutionary Relationships of Flatworm. *Current Biology* 25: 1347-1353.

Ehlers U. 1984: Phylogenetisches System der Platyhelminthes. *Verh Nat wiss Ver Hambg* 27: 291-294.

Ehlers U. 1985: Das Phylogenetische System der Platyhelminthes. *G. Fischer Verlag*, Stuttgart: 317 pp.

Freze V. I. 1965: Essentials of Cestodology. Vol V. Proteocephalata in fish, amphibians and reptiles. *Izdatel´stvo Nauka*, Moscow, Russia: 538 pp.

Goater T. M., Goater C. P., Esch G.W. 2014: Parasitims, Second edition. *Cambridge university press*, USA: 497 pp.

Halanych K., Bacheller J., Aguinaldo A., Liva S., Hillis D., Lake J. 1995: Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* 267: 1641-1643.

Hypša, V., Škeříková A., and Scholz T. 2005: Phylogeny, evolution and host-parasite relationships of the order Proteocephalidea (Eucestoda) as revealed by combined analysis and secondary structure characters. *Parasitology* 130: 359–371.

Chervy L. 2002: The terminology of larval cestodes or metacestodes. *Systematic parasitology* 52: 1–33.

Chervy L. 2009: Unified terminology for cestode microtriches: a proposal from the International Workshops on Cestode Systematics in 2002–2008. *Folia Parasitologica* 56: 199–230.

Katoh K., Standley D. M. 2013: MAFT Multiple Sequence Alignment Software Version 7: Improvements in performance and Usability. *Molecular Biology and Evolution* 30: 772-780.

Khalil L. F., Jones A., Bray R. A. (Eds.) 1994: Keys to the Cestode Parasites of Vertebrates. *CAB International*, Wallingford, U.K.: 751 pp.

Kocot M., Struck T., Merkel., Waits., Todt Ch., Brannock P., Weese D., Cannon J., Moroz L., Lieb B., Halanych K. 2017: Phylogenomics of Lophotrochozoa with Consideration of Systematic Error. *Systematic Biology* 66: 256-282.

Kuchta R., Scholz T., Brabec J., Wicht B. 2014: Chapter 16 Diphyllobothrium, Diplogonoporus and Spirometra. In: Xiao, Ryan, Feng (eds.) *Biology of Foodborne Parasites*. Section III Important Foodborne Helminths. *CRC Press*: 520 pp.

Laumer Ch. E., Hejnol A., Giriber G. 2015: Nuclear genomic signals of the microturbellarian roots of platyhelminth evolutionary innovation. *eLife Sciences* 2015: e05503.

Littlewood D.T.J. 2008: Platyhelminth systematics and the emergence of new characters. *Parasite* 15: 333-341.

Littlewood D.T.J., Curini-Galletti M., Herniou E. A. 2000: The interrelationships of Proseriata (Platyhelminthes: Seriata) tested with molecules and morphology. Molecular Phylogenetics and Evolution 16. 449-466.

Littlewood D. T. J., Olson P. D. 2001: Small subunit rDNA and the Platyhelminthes:signal, noise, conflict and compromise. *Interrelationships of the Platyhelminthes* (D. T. J. Littlewood and R. A. Bray, Eds.). Taylor and Francis, London and New York: pp. 262-278.

Littlewood, D. T. J., Rohde, K., Clough, K. 1999: The interrelationships of all major groups of Platyhelminthes: phylogenetic evidence from morphology and molecules. *Biological Journal of the Linnean Society* 66: 75–114.

Lockyer A.E., Olson P.D., Litlewood D.T.J. 2003: Utility of complete large and small subunit rRNA genes in resolving the phylogeny of the Neodermata: implications and a review of the cercomer theory. *Biological Journal of the Linnean Society* 78: 155-171.

Near, J. T., Dornburg A., Eytan R. I., Keck B. P., Smith W. L., Kuhn L. K., Moore J. A., Samantha P. a., Burbrink F. T., Friedman M., Wainwright P. C. 2013: Spiny-rayed fish phylogeny and diversification. *Proceedings of the National Academy of Sciences* 110: 12738-12743.

Olson P.D., Cribb T.H., Tkach V.V., Bray R. A., Littlewood D. T. 2003: Phylogeny and classification of the Digenea (Platyhelminthes: Trematoda). *International Journal for Parasitology* 33: 733-755.

Olson P. D., Poddubnay L. G., Littlewood D. T. J., Scholz T. 2008: On the Position of *Archigetes* and Its Bearing on the Early Evolution of the Tapeworms. *Journal of Parasitology* 94: 898-904.

Perkins E.M., Donnellan S. C., Bertozzi T., Whittington I. D 2010: Closing the mitochondrial circle on paraphyly of the Monogenea (Platyhelminthes) infers evolution in the diet of parasitic flatworms. *International Journal for Parasitology* 40. 1237-1245.

Rego A. A. 1994: The Order Proteocephalidea. In Keys to the Cestode Parasites of Vertebrates. Khalil L.F., Jones A., Bray R. A. (eds). *Cab International, Wallingford*, UK: pp. 257-293.

Rego, A. A. 1995: A new classification of the cestode order Proteocephalidea Mola. *Revista Brasileira de Zoologia* 12: 791–814.

Rego A. A. 1999: Scolex morphology of proteocephalid cestodes parasites of Neotropical freshwater fishes. *Memorias do Instituto Oswaldo Cruz* 94: 791-814.

Roberts L. S., Janovy J. J. 2009: Foundations of Parasitology, Eight Edition (G. D. Schmidt and L. S. Roberts, Eds.). *McGraw Hill Companies*, Inc., Boston, USA: 701 pp.

Rohde K., Hefford C., Ellis J. T., Baverstock P. R., Johnson A. M., Watson N. A., Dittmann S. 1993: Contributions to the phylogeny of Platyhelminthes based on partial sequencing of 18S ribosomal DNA. *International Journal for Parasitology* 23: 705–724.

Ruiz-Trillo I., Riutort M., Littlewood D.T.J., Herniou E., Baguna J. 1999: Acoel Flatworms: Earliest Extant Bilaterian Metazoans, Not Member of Platyhelminthes. *Science* 283: 1919-1923.

Saari S., Näreaho A., Nikander S. 2018: Canine Parasites and Parasitic Diseases 1, Chapter 4 – Cestoda (Tapeworms). *Academic Press*: 250 pp.

Schmidt, G. D. 1986: CRC Handbook of Tapeworm Identification. *CRC Press*, Inc., Boca Raton, FL, USA: 675 pp.

Scholz T. 1999: Life cycles of species of Proteocephalus parasites of fishes in the Paleartic Region: a review. *Journal of Helminthology* 73: 1-19.

Scholz T., de Chambrier A. 2003: Taxonomy and biology of proteocephalidean cestodes: current state and perspectives. *Helminthologia* 40: 66-77.

Scholz T., de Chambrier A., Schimazu T., Ermolenko A., Waeschenbach A. 2017: Proteocephalid tapeworms (Cestoda: Onchoproteocephalidea) of loaches (Cobitoidea): Evidence for monophyly and high endemism of parasites in the Far East. *Parasitology International* 66: 871-883.

Scholz T., Hanzelová V. 1998: Tapeworms of the genus Proteocephalus Weinland, 1858 (Cestoda: Proteocephalidae), parasites of fishes in Europe, Studie AV ČR No. 2/98. *Academia*, Prague, Czech Republic: 119 pp.

Scholz T., Hanzelová V. 1999: Species of Proteocephalus Weinland, 1858 (Cestoda: Proteocephalidae) from cyprinid fishes in North America. *Journal of Parasitology* 85: 150-154.

Scholz, T. Hanzelová V., Škeříková A, Shimazu T., and Rolbiecki L. 2007: An annotated list of species of the *Proteocephalus* Weinland, 1858 aggregate sensu de Chambrier et al. (2004) (Cestoda: Proteocephalidea), parasites of fishes in the Palaearctic Region, their phylogenetic relationships and a key to their identification. *Systematic Parasitology* 67: 139–156.

Škeříková A., Hypša V., Scholz T. 2001: Phylogenetic analysis of Europea species of Proteocephalus (Cestoda: Proteocephalidea): compatibility of molecular and morphological data and parasite-host coevolution. *International Journal of Parasitology* 31: 1121-1128.

Trifinopoulos J.M., Nguyen L. T., von Haeseler A., Minh B. Q. 2016: W-IQ-TREE: a fast online phylogenetic tool for maximul likelihood analysis. *Nucleic Acids Research* 44: W232-W235.

Wiliams H., Jones A. 1994: Parasitic Worms of Fish. *Revista do Instituto de Medicina Tropical de São Paulo* 36: 559-561.

Woodland W. N. F. 1925: On three new proteocephalids (Cestoda) and a revision of the genera of the family. *Parasitology* 17: 370-394.

Woodland W. N. F. 1933: On a new subfamily of proteocephalid cestodes – the Othinoscolecinae - from Amazon siluoid fish *Platystomatichthys sturio* (Kner.) *Parasitology*: 491-500.

Woodland W. N. F. 1935: Additional cestodes from the Amazon siluroids pirarará, dorad, and sudobim. *Proceedings of the Zoological Society of London* 104: 851-862.

Yamaguti S. 1959: The Cestodes of Vertebrates Systema Helminthum. *Interscience Publishers*, Inc., New York. 860 pp.

Zehnder, M. P. and J. Mariaux. 1999: Molecular systematic analysis of the order Proteocephalidea (Eucestoda) based on mitochondrial and nuclear rDNA sequences. *International Journal for Parasitology* 29: 1841–1852.