School of Doctoral Studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

Population genomics of parasite adaptation: New insights into diversification and speciation of parasites

Ph.D. thesis

Masoud Nazarizadeh

Supervisor: doc. RNDr. Jan Štefka, Ph.D.
Department of Parasitology, Faculty of Science, University of South Bohemia
Institute of Parasitology, Biology Centre of the Czech Academy of Sciences

České Budějovice 2024

This thesis should be cited as:
Nazarizadeh, M., 2024: Population genomics of parasite adaptation: New insights into diversification and speciation of parasites. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 230 pp.

☆ Annotation

This thesis investigates the ecological and genetic dynamics of parasitic species, primarily focusing on *Ligula intestinalis* tapeworms and *Cimex* bedbugs, aiming to enhance our understanding of their evolutionary biology, genetics, and adaptation mechanisms. In Chapter I, it elucidates the prevalence and population genetic structure of *L. intestinalis* plerocercoids in various cyprinoid fish species across Czech water bodies, focusing on how different hosts and environments influence genetic diversity and distribution. Chapter II explores the historical dispersal and host-switching events that have shaped the evolutionary history of the *L. intestinalis* species complex on a global scale. This includes identifying major evolutionary lineages, their host specificity, and the influence of biogeographic and demographic events on lineage diversity. Chapter III provides a highly contiguous genome assembly for *L. intestinalis* and examines RNA transcription patterns to understand the regulation of biological functions during its life cycle, establishing a comprehensive reference transcriptome for both plerocercoid and adult stages. Chapter IV explores the mechanisms underlying possible ecological speciation in *L. intestinalis*, particularly under a sympatric setting, to determine whether host specificity influences population structure in the absence of geographical separation. Chapter V shifts focus to the establishment of the tropical bedbug, *Cimex hemipterus*, in Central Europe, using genetic analyses to reveal limited mtDNA variability and the presence of pyrethroid resistance mutations. Finally, Chapter VI investigates genetic and transcriptomic divergence in *C. lectularius* populations associated with human and bat hosts, exploring how host switching influences genetic diversity and adaptation. This genomic and transcriptomic approach provides valuable insights into the evolutionary dynamics and adaptive strategies of parasitic species.

☆ Declaration

I hereby declare that I am the author of this dissertation and that I have used only those sources and literature detailed in the list of references.

České Budějovice, 12.07. 2024

…….

Masoud Nazarizadeh

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre of the ASCR, supporting doctoral studies in the Parasitology study program.

☆ **Financial support**
The thesis was supported by following grant awards:

- Czech Science Foundation (GA19-04676S)
- Ministry of Education, Youth and Sports (LTAUSA18032)
- Grant Agency of the University of South Bohemia

☆ Acknowledgements

I am deeply grateful to the numerous individuals whose support and guidance were instrumental throughout my PhD journey; their contributions are too vast to enumerate on a single page.

Foremost, I extend my heartfelt thanks to my PhD supervisor, Dr. Jan Štefka, for his unwavering calm and steady support. Dr. Štefka not only opened doors to numerous opportunities but also introduced me to the fascinating world of science and population genetics. His mentorship has been valuable in shaping my academic career. I must admit that without his significant support and guidance, I would not have been able to complete my PhD journey.

Special thanks to Dr. Ondřej Balvín from Life Science University, who has been a valuable mentor throughout my PhD thesis. His guidance and insights have been vital to my research and academic growth. I would also like to express my gratitude to Dr. Emiliano Trucchi from Marche Polytechnic University in Ancona, Italy, and to Dr. Peter D. Olson from the Natural History Museum in London, for their support and guidance during my internship.

I am deeply grateful to Somayeh Rasouli Dogaheh for her significant support and encouragement throughout my academic pursuits. I would also like to extend my sincere thanks to Anna Mácová and Roman Hrdlička for their valuable assistance during my fieldwork. Their expertise and friendship greatly enriched my research experience and contributed significantly to the success of my projects. Furthermore, I extend my deepest thanks to my colleagues from České Budějovice: Serafim Nenarokov, Valeria Gabkova Juricova, Jakub Vlček, Jana Martinů, Dragomir Damnjanovic, and Milena Nováková, for their camaraderie and support throughout my research journey.

I am also grateful to the University of South Bohemia, Faculty of Science, for providing a stimulating and supportive environment where I could thrive academically over many rewarding years. Similarly, my appreciation goes to the staff and colleagues at the Biology Centre, Institute of Parasitology, for their valuable contributions to my work.

☆ List of publications and manuscripts and author's contribution

I. **Nazarizadeh, M**., Peterka, J., Kubečka, J., Vašek, M., Jůza, T., de Moraes, K. R., ... & Štefka, J. (2022). Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies. *Folia Parasitologica*, *69*, 1-10.
Masoud Nazarizadeh concluded the research under the supervision of JŠ. MN and JŠ drafted the text with contributions by JP and PB. MN contribution 70%.

II. **Nazarizadeh, M**., Nováková, M., Loot, G., Gabagambi, N. P., Fatemizadeh, F., Osano, O., ... & Štefka, J. (2023). Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite–The *Ligula intestinalis* species complex. *Molecular Phylogenetics and Evolution*, *180*, 107677.
Masoud Nazarizadeh and JŠ defined the research objective and drafted the manuscript. MNaz. analysed the data under the supervision of JŠ and ET. MN contribution 60%.

III. **Nazarizadeh, M**., Nováková, M., Drábková, M., Catchen, J., Olson, P. D., & Štefka, J. (2024). Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *Ligula intestinalis*. *Proceedings of the Royal Society B*, *291*(2015), 20232563.
Masoud Nazarizadeh carried out the analysis of data and initiated the drafting of the manuscript under the guidance of JŠ and PO, and with contributions from JC. MN contribution 60%.

IV. **Nazarizadeh, M**., Nováková, M., Vlček, J., & Štefka, J. (2024). Host-associated genetic differentiation in the face of ongoing gene flow: ecological speciation in a pathogenic parasite of freshwater fish. *bioRxiv*, 2024-01.
JŠ initiated the research topic. Masoud Nazarizadeh participated in field collection, analysed the data and drafted the manuscript under the supervision of JŠ and JV. MN contribution 70%.

V. Balvín, O., Sasínková, M., Martinů, J., **Nazarizadeh, M**., Bubová, T., Booth, W., ... & Štefka, J. (2021). Early evidence of establishment of the tropical bedbug (*Cimex hemipterus*) in Central Europe. *Medical and Veterinary Entomology*, *35*(3), 462-467.
Masoud Nazarizadeh participated in writing and analyses. MN contribution 20%.

VI. **Nazarizadeh**, M., Sasínková, M., Martinů, J., Nováková, M., Štefka, J., Křemenová, J., Bartonička, T., & Balvín, O. (2024). Host Influence on Genetic Diversity and Adaptability in *Cimex lectularius*: Insights from Genome-Wide SNP and Transcriptome Analyses. Manuscript.
Masoud Nazarizadeh carried out all the analyses under the supervision of OB and JŠ. MN drafted the text with contributions by OB and JŠ. MN contribution 60%.

☆ Co-author agreement

Jan Štefka, the supervisor of this thesis and co-author of all presented manuscripts, fully acknowledges the contribution of Masoud Nazarizadeh as the first author and his contributions as stated above.

…….

doc. RNDr. Jan Štefka, Ph.D.

Contents

Introduction

Introduction
Population Genomics of Parasite Adaption: New insights into diversification and speciation of parasites

Population Genomics

Population genomics is generally described as the study of multiple genomes and loci with the aim of uncovering the evolutionary processes involved in shaping the patterns of intraspecific variation. These processes may include natural selection, mutation, gene flow and genetic drift (Luikart et al., 2003; Safran et al., 2016). According to this interpretation, population genomics focuses on comparing locus-specific forces such as mutation and selection to genome-scale forces like genetic drift and demographic processes to make better inferences about the genomic and evolutionary basis of adaptation, fitness, gene flow, speciation, inbreeding and outbreeding (Hunter et al., 2018). In other words, population genomics studies loci and genomes simultaneously in an attempt to find answers to the key questions of population genetics (Garner et al., 2016; Hohenlohe et al., 2018). Based on the conceptual framework developed by Hohenlohe et al. (2010), population genomics studies evolutionary processes and patterns of genetic diversity across genomes by applying population genetic statistics to numerous loci. For instance, gene expression, population differentiation and genome diversity among populations are evaluated to identify loci under selection or alleles linked with a phenotypic trait (Hohenlohe et al., 2018; Rosenfeld et al., 2016).

Population genomics provides tools to deepen our understanding of processes not only in free living but also in symbiotic (or parasitic) organisms (Korhonen et al., 2022; Li et al., 2021; Nazarizadeh et al., 2024; Olson et al., 2020). It significantly contributed towards our understanding of host-symbiont coevolution, which can be demonstrated, for instance, in the system of the Eurasian Alcon large blue butterfly (*Phengaris alcon*) and its two hosts: a gentian plant (*Gentiana pneumonanthe*) and an ant (*Myrmica scabrinodis*) (De Kort et al., 2018). The butterfly's eggs are laid in the ovules of gentian flowers to feed and later develop into caterpillars. Because of the damage caused by caterpillars to the ovules, phenological flowering shifts have been observed in *G. pneumonanthe* to avoid the times of high parasite pressure (Valdés and Ehrlén, 2017). Due to their chemical mimicry, the caterpillars are mistakenly adopted by worker *Myrmica* ants and are

nursed as superior larvae in the nest. As a result of this host-parasite relationship, coevolutionary changes have occurred between the surface chemistry of *P. alcon* caterpillars and *Myrmica* ants (Nash et al., 2008). In a study applying restriction associated DNA (RAD) sequencing to analyze the effect of habitat fragmentation on *P. alcon* as well as its host species (De Kort et al., 2018), abiotic variables, such as elevation, were found to be responsible for some of the genetic variation among host populations. Moreover, it was revealed that a considerable proportion of population genetic variation in the butterfly was driven by the genetic diversity of host taxa. Also, this study proposed that the genetic structure between host and parasite species has synchronized as a result of mutual coevolutionary selection. Thus, habitat fragmentation and associated landscape genomic structure substantially affected the functioning of the mountain Alcon butterfly as well as its host species.

Phylogeography and historical biogeography of parasites

Parasites account for a significant portion of the global biodiversity, inhabiting nearly every free living species and spreading across various geographical regions. They affect not only around one billion people, with neglected tropical diseases, but also numerous domestic and wild animals (Fenwick, 2012; Nadler and De León, 2011; Ondrejicka et al., 2014). Nevertheless, they can enhance our understanding of the geographical distribution and evolutionary patterns of their hosts. Recognized species often consist of distinct genetic and morphological groups that are differentiated through speciation or the evolution of separate metapopulation lineages (De Queiroz, 2007). The evolutionary dynamics of parasites is influenced by several factors including host colonization, ecological interactions, speciation, and changes in the host's genetics, and more recently also by human activities and environmental changes, which may lead to parasite transfer between different host species (Hoberg and Brooks, 2013, 2008).

Host-parasite systems provide an excellent opportunity to study coevolutionary and ecological processes throughout the history of life. Host-parasite interactions are studied by various evolutionary approaches, which produce varying results (Martínez-Aquino et al., 2014; Morand and Krasnov, 2010). In some host-parasite systems, the evolutionary interrelation (co-speciation) between parasites and hosts is clearly evident, while in others, it is hardly detectable. In particular, assemblages that include cestode parasitic worms and pinniped species show that parasite diversification relies more on the parasite's capacity to infect new hosts and less on coevolution

with the host taxa. This often leads to increased incongruence among phylogenies (Hoberg and Adams, 2000, 1992).

Co-speciation events are characterized by various phylogenetic relationships, including host switches, duplications, and losses, alongside the co-speciation itself. A host switch, also known as a host shift, occurs when a parasite moves to a new host, diverging from its original lineage, which can lead to speciation under the right conditions of opportunity and compatibility (Clay and Kover, 1996). Duplication, in contrast, involves a parasite speciating within the same host, without switching hosts. Losses, or sorting failures, may lead to the extinction of a parasite if it cannot keep up with the evolutionary changes of its host, a concept underscored by the Red Queen Hypothesis (Hoffman, 1991). This hypothesis describes a continuous cycle of adaptation and counter-adaptation between interacting species, emphasizing the dynamic nature of host-parasite relationships and the ongoing evolutionary arms race that can shape biodiversity (Clay and Kover, 1996; Liow et al., 2011).

The Stockholm Paradigm is a contemporary approach addressing the interactions between climate change and emerging diseases, emphasizing the evolutionary perspective. It underlines the importance of recognizing that pathogens can adapt and find new hosts in changing environments, especially under conditions influenced by climate change and human activities (Brooks et al., 2019). This paradigm challenges traditional views that focus solely on direct host-pathogen adaptations and instead considers the broader ecological and evolutionary contexts. According to the paradigm, the movement of species due to environmental changes and human actions creates new interactions among species, including pathogens, hosts, and vectors. This increases the opportunities for diseases to emerge and spread. For instance, as global temperatures rise and habitats are altered or destroyed, species are forced into new areas, often bringing them into contact with other species and pathogens they would not have encountered otherwise. This can lead to new diseases emerging or existing ones spreading or becoming more virulent. The Stockholm Paradigm synthesizes four main theories: Taxon Pulse, Ecological Fitting, Oscillation, and Geographic Mosaics of Coevolution. These concepts interact to describe how fauna, including parasites and their hosts, adapt and diversify through ecological and evolutionary processes (Agosta and Klemens, 2008; Araujo et al., 2015; Cháves-González et al., 2022). Taxon Pulse describes adaptive shifts through geographical or ecological expansions, facilitating biotic mixing and increasing

host-parasite contact opportunities. This driver underlines the dynamics of niche perturbations and geographical colonization that promote recurrent host invasions. Ecological Fitting, on the other hand, triggers host switching or colonization through resource tracking or exploitation of new "sloppy fitness spaces" provided by novel hosts, effectively leveraging the opportunities and compatibilities established by Taxon Pulse (Cháves-González et al., 2022) .

Oscillation explains the cyclical pattern wherein specialist parasites become generalists and subsequently evolve back into specialists. This alternation in host range dependency is influenced by resource usage, potentially narrowing ecological associations and determining host exploitation patterns. The interaction between Oscillation and Ecological Fitting is crucial for understanding how host ranges expand and contract over time. The concept of Geographic Mosaics of Coevolution (GMC) offers a framework for examining real-time coevolutionary processes between hosts and parasites across different environmental conditions. GMC posits that environmental variability leads to genetic diversification and distinct evolutionary trajectories, creating hotspots of coevolutionary activity amid broader areas of lesser interaction, and influencing trait evolution through genomic alterations and gene flow (Agosta and Brooks, 2020).

Evolutionary studies of host-parasite systems are mainly aimed at determining the evolutionary relationships and geographical ranges of host species using probabilistic biogeographical models (Ronquist and Sanmartín, 2011) and phylogenetic trees of parasite and host species (Page, 2003). Geographical factors may strongly affect evolutionary host-parasite relationships through creating geographic co-differentiation between parasites' and hosts' evolutionary trajectories (Wafa Bouzid et al., 2008; Zardoya and Meyer, 1996). Unfortunately, biological models that illustrate diversification events in host-parasite relationships are rare (Althoff et al., 2014; Caira and Jensen, 2001). Historical biogeography of parasite species, which are defined as broadly distributed parasitic taxa that correspond to a monophyletic group of host species (Nieberding et al., 2008; Pérez-Ponce de León and Choudhury, 2005), provide exceptional possibilities for studying host-parasite diversification events. The findings of such studies enhance our knowledge of the spatial genetic structure of populations, which allows us to understand present and historical demographic processes, e.g. glacial persistence of taxa in refugia, postglacial colonization, and population bottlenecks (Emerson and Hewitt, 2005; Provan and Bennett, 2008). Additionally, recent methodological and theoretical advances gave rise to comparative phylogeography, which

compares the evolution and genetic structures among co-distributed species with shared ancestral histories or ecological niches (Bermingham & Moritz, 1998; Engelbrecht, Matthee, Du Toit, & Matthee, 2016; Maia Da Silva et al., 2007; Matthee, Engelbrecht, & Matthee, 2018; Nieberding & Morand, 2006).

Host-parasite congruent phylogeography

Some early studies showed a strong correlation between parasite and host species' diversity in different regions (Sapegina, 1988), but no general pattern has been established. For example, species richness among fish species distributed in different tributaries of the Ohio River was positively associated with species richness of unionid freshwater mussels. Ectoparasitic larvae of these mussels infect fish hosts in order to successfully disperse (Vaughn and Taylor, 2000). In a different study, species richness of the bumblebee (*Bombus terrestris*) was negatively correlated with species richness of its parasite (*Crithidia bombi*) (Durrer, 1996; Schmid-Hempel, 2001). Yet, this study estimated parasite species richness for each host species, and not for each site; therefore, the results of the two studies cannot be compared. Another study revealed no correlation between species richness of parasitic flea species and their mammalian hosts (Yunzhi et al., 2002). However, other reports have all confirmed the positive correlation between host-parasite species diversity (Dallas et al., 2020; Kamiya et al., 2014; Wood and Johnson, 2016).

Similar results have been found for parasitic helminths, such as trematodes, which generally require a vertebrate primary host and an intermediate (usually an aquatic snail) host. Species richness of these parasites is likely governed by different host species at different phases of their life cycle. Trematode larval diversity in intermediate hosts (snails) showed a positive association with local diversity of primary host species (birds), regardless of whether the microhabitats were studied together or individually (Hechinger and Lafferty, 2005). In other words, species richness of primary hosts determines the species richness of parasites in intermediate hosts. A larger-scale study found that European trematode parasite diversity was positively linked to regional diversity of primary hosts, and not intermediate hosts (Thieltges et al., 2011). This condition is more evident in parasite species which only need a single host species. For example, mammalian host species diversity was positively linked to parasitic flea species diversity across different Holarctic regions (Krasnov et al., 2004). In the same vein, species richness of different generalist and specialist parasites was strongly correlated with host species richness in 29 species of carnivores (Dunn et al., 2009).

The genetic structure of parasite populations has been suggested to represent their evolutionary potential (McDonald and Linde, 2002), meaning that those with higher gene flow, higher mutation

rates and larger Ne[1] present the greatest risk of resistance gene breakdown. In addition to phylogenies and studies of co-speciation events (Huyse, *et al* 2005), studies of population genetic structure are also essential to examine the origin of parasites evolving resistance to drugs and determine the key elements involved in parasite evolution (Huyse, *et al* 2005). These findings are vital to management and control of parasitic diseases.

Other studies on phylogeographical patterns of parasite species have uncovered valuable information about the cryptic phylogeography of host taxa, their historical migration patterns, past glacial refugia, and inter-population gene flow. In one of these studies, the phylogeography of *Heligmosomoides polygyrus*, an intestinal parasitic nematode, was compared to that of the wood mouse, *Apodemus sylvaticus*, a host for *H. polygyrus.* The results found well-defined allopatric Quaternary refugia for the parasite and the host species in Italy. Also, it was revealed that both taxa colonized North Africa from South of Spain (Nieberding et al., 2005, 2004). Another example is provided by a study on the phylogeography of *Puumala hantavirus* (PUUV) and its host species, the bank vole (*Clethrionomys glareolus*), which validated a previously suggested historical migration path for *C. glareolus* from the East European plain toward northern Europe after the last glacial maximum (Asikainen et al., 2000; Dekonenko et al., 2003).

Phylogeographic studies contribute significantly to understanding species introductions and their historical movements. For example, research on *Bowlesia incana*, a flowering plant from the Umbelliferae family and a host for the parasitic butterfly *Greya powelli*, revealed that *B. incana* was introduced from South America to North America far earlier than previously thought—more than 250 years ago (Pellmyr et al., 1998). Similarly, a study on the giant liver fluke *Fascioloides magna*, which significantly impacts both free-living and domestic ruminants, demonstrated complex migration patterns across North America and documented its introduction into Europe from distinct North American origins. The analysis used microsatellite loci to illustrate the intricate pathways of host-driven dispersal and subsequent parasite migration. This study also revealed that the European populations of *F. magna* showed distinct genetic lineages compared to their North American counterparts, indicating multiple introductions into Europe (Juhásová et al., 2016; Králová-Hromadová et al., 2011).

[1] Effective population size

Host-parasite incongruent phylogeography

Although numerous parasite taxa have been known to produce phylogeographic patterns similar to their hosts (Benovics et al., 2020; Johnson et al., 2006; Martinů et al., 2018; Štefka et al., 2009), host-parasite phylogeographic incongruities have also been detected. A large-scale study was done to compare the phylogeography of *Greya enchrysa*, *Greya subalba* and *Greya politella* moths in the Americas to the phylogeographic patterns of their hosts and parasitoids (Saxifragaceae plants and *Agathis* braconid wasps, respectively) (Althoff and Thompson, 1999; Brown et al., 1997). *G. politella* is a generalist floral parasite which has had multiple evolutionary host shifts and its phylogeographic structure shows no signal of co-differentiation with a specific host. Also, the phylogeographies of *Agathis n. sp.* and *Agathis thompsoni,* two generalist wasps, did not match the phylogeography of their host moth species, *G. enchrysa* and *G. subalba.* In the same vein, phylogeography of the nest parasite Horsfield's bronze-cuckoo (*Chrysococcyx basalis*) was compared to that of two genera of its passerine hosts, *Acanthiza* and *Malurus*. The phylogeography of *C. basalis* shows that, as the climate improved after the Pleistocene, the species experienced recent expansions in its distribution and host ranges, suggesting no signs of evolution of stable host races in *C. basalis* (Joseph et al., 2002).

Moreover, in polyxenous parasites, co-phylogeographic patterns may be complicated by the inclusion of intermediate host(s). For example, the comparison of a host-specific parasitic cestode (*Paranoplocephala arctica*) with its host, collared lemming (*Dicrostonyx* spp.) in the Holarctic revealed two distinct clades for each taxon across the Palearctic and Nearctic, divided by the Bering Strait. Nonetheless, *P. arctica* phylogeographic analyses uncovered further differentiation events at the Bering Strait and in the Arctic Archipelago that mismatched the phylogeographic patterns of its mammalian host species. Such host-parasite incongruities are likely a result of *P. arctica* residing in invertebrate intermediate hosts, which decreases its spatial dependency on its primary mammalian host (Wickström et al., 2003).

In addition to these examples, the research by Wasimuddin et al. (2016) further illustrates phylogeographic incongruities within a host-parasite relationship across a hybrid zone. Their study examined the whipworm *Trichuris muris* in the European house mouse hybrid zone, an area where different mouse taxa hybridize. Contrary to the expectation of co-divergence due to the whipworm's direct life cycle and the close association with its host, genetic analyses showed no

significant congruence in the phylogeographic patterns between the parasite and the hybridizing mouse populations. This suggests that *T. muris* might not be as host-specific as previously thought, capable of infesting non-specific hosts within the zone. Such findings indicate that hybrid zones, by facilitating gene flow among divergent taxa, can further complicate the co-phylogeographic patterns of hosts and parasites, potentially leading to significant phylogeographic incongruence.

Ecological speciation and pattern of parasite diversity

It has been widely acknowledged that identifying the evolutionary processes and mechanisms involved in the speciation of taxa is crucial to understanding the patterns of biodiversity. The total number of parasite species constitutes about 30% of total eukaryotic diversity, and at least 10% of the extant metazoan species (De Meeûs and Renaud, 2002; Morand et al., 2015; Poulin and Morand, 2000). Given their high specialization and diversification potential, parasites serve as exemplary biological models for the studies of speciation processes. Also, certain parasite species occur in environments fit for sympatric speciation. Finally, given that parasites are such ubiquitous organisms, understanding the interactions between parasites and their hosts are vital to evolutionary and ecological studies (Morand et al., 2015).

Ecological speciation is a key process which leads to the emergence of reproductive barriers between populations and consequently speciation due to divergent natural selection under different ecological conditions (Nosil, 2012). Ecological speciation emerges as a consequence of selecting traits with genetic correlation to reproductive isolation, where the formation of new species is a by-product of adaptive divergence. Even though ecological speciation is understood as one of the primary mechanisms of sympatric speciation, the process can play role under all geographical settings, including parapatric and allopatric layouts (Nosil, 2012; Rundle and Nosil, 2005).

Population genetics and population ecology of parasites are closely connected, and their population structures are affected by host specificity, host mobility, parasite infra-population size, parasite reproductive mode, and parasite life-cycle complexity (Huyse et al., 2005). The significance of these factors depends on a particular species of the parasite. Thus, to understand the determining processes in diversification of parasites, comparative phylogenetic and population genetic approaches are needed (Bueter et al., 2009).

The advent of novel genomic approaches allowed us to investigate the evolutionary patterns and population diversity of organisms at an unprecedented depth. These advanced tools, including massively parallel sequencing (MPS), enabled the use of genomic methods for non-model taxa (McCormack et al., 2013), producing significant amounts of genetic data as well as exceptional insights into the adaptive and neutral processes involved in evolution of populations (Davey et al., 2011; Ellegren et al., 2012). As one of the pioneers of genomic methods in studying ecological adaptation, Nosil et al. (2012) used genome-wide data to examine an adaptive trait in genomic divergence for the evolution of ecomorphs. Various populations of *Timema* walking sticks from different plant hosts and from different regions were compared and a particular set of gene outliers (FST outlier method) was detected along with comparisons between the hosts.

Parasitological studies have been largely focused on the evolutionary origin of resistance in parasites. For instance, recent multi-locus microsatellite research on human lice ecotypes (the body louse and the head louse) has revealed remarkable geographic population structures and patterns of inbreeding (Ascunce et al., 2013). Most importantly, this study uncovered insecticide resistance in lice populations possibly as a consequence of selection for resistant traits along with outlier loci potentially involved in resistance evolution. Therefore, their results highlighted the significance of population genetics in the evolution of parasites. Further research using high-throughput single nucleotide polymorphism sequencing technologies is required to generate a more complete genome map and consequently more accurate inferences about the evolutionary processes of parasite evolution (Ascunce et al., 2013). Recently, a genome-wide analysis of the human head louse (*Pediculus humanus capitis*) collected across the globe revealed structured genetic populations based on geographical regions. The study showed significant genetic variation, especially between African and non-African lice, related to human colonisation of the planet (House et al., 2023).

Early genomics researchers assumed that divergent selection of certain traits possibly hinders gene flow in genetic regions related to these traits (Nadeau et al., 2012), hence forming genomic islands of speciation. Yet, recent genomic studies using multiple models, including *Heliconius* butterflies, *Ficedula* flycatchers (Burri et al., 2015; Cruickshank and Hahn, 2014), suggest that both the processes of recombination and natural selection create such patterns of genome-level diversity (Wolf and Ellegren, 2017). Therefore, more powerful tools such as analyses of whole-genome

sequencing and genetic linkage can be used to determine whether genomic regions under selection are isolated or contiguous across chromosomes, and to measure the decrease in genetic divergence in regions away from the selected sites. In addition, analyses based solely on SNP divergence in genomes are not strong enough to infer singular, historical events (Burri 2017), e.g. host switching in parasites. Detection of selection signatures becomes more challenging with the ongoing divergence of populations as a result of high recombination rates and rapid accumulation of mutations. In spite of these issues, genomic studies are currently biased toward the use of SNP data, which leads to an overestimation of the significance of gene variation at coding exons (Stern and Orgogozo, 2008). Studies of neutral and adaptive variation at the genomic level must be combined with analyses of changes in gene regulation in order to uncover adaptive changes in gene expression (Hornett and Wheat, 2012). However, among studies on parasite-host coevolution, those using transcriptomic genomic methods are still scarce.

Population Genomics of Speciation

Speciation is one of the most essential evolutionary processes, it contributes to the diversity of taxa by accompanying population divergence with reproductive isolation. Speciation genomics is a relatively new field, aimed at understanding how genomic variation is involved in the process. This field uses extensive amounts of genome-wide data to determine loci and genomic features that promote speciation (Nosil and Feder, 2012; Seehausen et al., 2014) or hamper it, for example by gene flow reducing genetic variation between different populations.

Formation of barrier loci (loci that contribute to the reduction of gene flow) is influenced by the homogenizing effects of gene flow. However, selection can promote the accumulation of these loci, especially among ecologically adapted forms, counteracting the gene flow and supporting divergence. Understanding the mechanisms of reproductive isolation and the accumulation of genetic differences are fundamental questions in speciation genomics studies. Divergence can persist despite gene flow, allowing the movement of alleles across a variety of taxa. This is observable by examining species along the extremes of divergence or speciation, ranging from weakly diverged species to well-distinguished taxa (Feulner et al., 2015; Nadeau et al., 2013; Riesch et al., 2017; Seehausen et al., 2014). While species may exhibit unique characters despite ongoing gene flow, these differentiations may not necessarily have accumulated under current gene flow conditions. Instead, secondary contact and hybridization may occur in populations with

narrow spatial isolation (parapatric populations), or where continuous contact is under divergent selection. Genomic differentiation under these scenarios, whether through hybridization in zones of short temporal secondary contacts or other forms, requires a detailed examination using genome scans and other molecular techniques that elucidate genomic differences (Nosil and Feder, 2012; Payseur and Rieseberg, 2016).

Genomic signatures of speciation

High-resolution genomic sequencing tools enable studies of population diversification based on screens of genome-wide differentiations (Seehausen et al., 2014). Reduction of gene exchange under divergent selection between different environments creates barriers to gene flow, which eventually lead to ecological speciation (Nosil, 2012). According to the theoretical model of Dobzhansky-Muller incompatibility, speciation may also occur intrinsically due to fitness reduction of hybrids as a result of hybrid genic incompatibilities at multiple loci (Coyne and Orr, 2004; Dobzhansky, 1936; Muller, 1940). Nevertheless, the Dobzhansky-Muller model of incompatibility suffers from several shortfalls; for instance, alleles with epistatic interactions could also mediate local adaptation, thus leading to fitness reduction of hybrids. According to the genic model of speciation, genetically differentiated regions arise at a few genomic regions due to divergent selection advancing reproductive isolation. Such barrier loci show resistance to gene flow through intrinsic incompatibility or divergent selection, whereas the remaining genomic regions are balanced under the homogenizing effect of gene flow (Wu, 2001).

Studies have revealed that barrier loci could form in populations that are geographically isolated or could accumulate despite gene flow. In the case of the geographic isolation, gene flow eliminates genetic differences that have been established in the time of allopatry, as a result of secondary contact between populations along the genome, except at sites that include barrier loci. However, if gene flow is allowed, genetic differentiation remains restricted to regions unlinked to barrier loci. The diverging populations, under both scenarios, will have few highly-differentiated loci surrounded by low-differentiated sites in their genomes at the beginning of speciation (i.e. ‘genomic islands of divergence’ or ‘differentiation islands’) (Nosil and Feder, 2012; Turner et al., 2005). Moreover, limited gene flow in the proximity of genomic islands of differentiation enables further accumulation of barrier loci at adjacent regions, thus increasing the size of these islands during speciation (Via, 2012). With the progressive accumulation of even more barrier loci, either

near or far from the existing genomic islands of differentiation, reproductive isolation becomes stronger and therefore genetic differentiation increases throughout the entire genome. In this case, empirical (Riesch et al., 2017) and theoretical (Feder et al., 2012) studies suggest that the increase in genetic differentiation is not linear and may exhibit a tipping point in the number of differentiated regions or the strength of reproductive isolation. Consequently, populations may eventually move from having limited islands of differentiation to showing large genome-wide differentiations (Nosil et al., 2017). Despite the controversy of the "genomic islands of speciation" concept, findings of the studies on genomic islands have helped researchers identify genomic paradigms of genetic differentiation among diverging taxa and among closely related species in order to understand the underlying mechanisms of reproductive isolation.

Genome scans to identify barrier loci

Numerous studies have highlighted heterogeneous genomic patterns in islands of differentiation (Nadeau et al., 2012; Turner et al., 2005; Via et al., 2012), yet interpreting such patterns is a difficult task, particularly to specify whether speciation has been the dominant force in the formation of islands of differentiation (i.e. 'speciation islands') or whether different processes unlinked to evolutionary mechanisms of reproductive isolation have shaped these islands (i.e. 'incidental islands') (Cruickshank and Hahn, 2014). The homogenizing effect of gene flow between taxa across most of the genomic regions unlinked to loci associated with reproductive isolation has served as an important role in shaping differentiation islands in the genic model of speciation. Nevertheless, numerous studies have reported similar results for heterogeneous genomic differentiation patterns among populations that apparently have no contemporary gene flow and are geographically isolated (Renaut et al., 2013; Vijay and Ravichandran, 2016).

Retention of ancestral polymorphism and stochasticity in changes in allele frequencies are the likely causes of heterogeneity in genetic differentiation of closely related taxa even if current gene flow is absent, particularly if linked neighbouring sites are affected by selection (Nachman and Payseur, 2012; Noor and Bennett, 2009). Two types of selection may occur under these conditions: negative or positive selection. Selection is not necessarily the sole mechanism of emerging reproductive isolation. The reason behind this is that recurring positive or negative selection eliminates polymorphisms at the regions targeted by selection as well as at nearby regions in linkage disequilibrium (LD). This process, called 'linked selection' (Cutter and Payseur, 2013)

generates regions of locally decreased Ne, which further speeds up lineage sorting, increases differentiation (FST), and reduces genetic diversity (π). Considering that the effect of linked selection is known to be a function of local variation in recombination rates and the density of sites targeted for selection, the amplitude of lineage sorting and therefore genomic differentiation will be heterogeneous along the genome. Hence, understanding the underlying mechanisms of evolutionary and genetic processes involved in formation of genomic islands of differentiation is crucial (Rajora, 2019).

To distinguish between these scenarios, comparison of patterns of differentiation across genomes of several closely related taxa with varying divergence times is needed. It is assumed that conserved genomic structure and recombination rates between closely related taxa (e.g. density and distribution of genes) generate shared patterns of genomic differentiation in the strength of linked selection. A great deal of studies have discussed that genomic islands of differentiation during early and developed stages of speciation (less divergent and more divergent taxa, respectively) probably indicate 'incidental islands', whereas population-specific islands of differentiation indicate candidate 'speciation islands' (Andrew and Rieseberg, 2013; Burri et al., 2015; Poelstra et al., 2014; Renaut et al., 2013). Also, during the earliest stages of speciation, 'speciation islands' are differentiated from heterogeneous genome diversity formed under linked selection independent of reproductive isolation. The reason behind this is that strong divergent selection produces few highly differentiated regions with large blocks of haplotypes (Andrew and Rieseberg, 2013; Poelstra et al., 2014).

Nowadays, genome-wide polymorphism data of a wide variety of taxa are becoming increasingly available. Thus, genome scans provide a practical way of understanding the genomic basis of reproductive isolation through identification of genomic differentiations (Rajora, 2019). A major advantage of genome scans is that it is not essential if phenotypic divergence is initially characterized. In other words, genome scan can characterize loci that underlie new divergently selected traits. Furthermore, this approach is used for identifying divergently selected regions between easily interbreeding species (lineages) based on the genome signatures created by both gene flow and selection. Nevertheless, disentangling the underlying forces shaping genomic islands of divergence (i.e. divergent selection associated with reproductive isolation versus linked selection) is still a complicated task, as these processes can leave highly similar signatures and

also occur simultaneously (Yeaman et al., 2016). Although the genome scan detects several islands of differentiation, the candidate regions of the genome can be limited by integrating trait information with the genome scan. This method has been seldom performed for traits that were not initially genetically mapped or well characterized. Using an outlier approach, a study on warblers attempted to detect anonymous outlier loci and compare them to phenotypic variation among populations, yet even these phenotypes had well differentiated colours (Toews et al., 2016). Even though the genome scan analysis is obviously a powerful method for detecting genomic patterns of divergence, if used solely, it cannot characterize causal genes for reproductive isolation (Rajora, 2019).

Patterns of host specificity

Host specificity is a crucial factor in shaping the distribution and diversity patterns of parasite species. Parasite species that infect a single or a narrow range of hosts often exhibit a higher specificity, leading to unique adaptations and co-evolutionary dynamics with their host (Dobson et al., 2008). Furthermore, host specificity may be either associated with the ability of a parasite species to attach and adapt to a host species or the limitations of a parasite species in finding a host species (i.e. adaptive vs non-adaptive factors) (Dobson et al., 2008). In addition, host specificity may be affected by other host-parasite interactions, such as when phoresis (parasite species being mechanically transported by host species) favours host-switching (Harbison and Clayton, 2011). When dealing with different host species, the adaptive cause for host specificity can be interpreted as differences in fitness. In practice, this has rarely been experimentally studied, therefore our current knowledge of host ranges basically comes from fragmentary presence data of wild-living parasite species and their related hosts (Dunn et al., 2009). However, more careful examinations, such as taxonomic re-description of specialist parasite species, reveal that host species can be more wide-ranging than expected (Clayton and Price, 1999).

Based on molecular studies, different cryptic species can be found within morphospecies of parasites (Miura et al., 2005), some of which may show higher host specificity than the wide host range previously suggested for the original morphospecies (Jousson et al., 2000; Poulin and Keeney, 2008; Pouyaud et al., 2006). Such erroneous estimates of host specificity may also stem from other factors, such as parasites that occasionally infect certain hosts, or those that infect hosts but do not reproduce or feed on them (Dunn et al., 2009). In addition, results of experimental

studies suggest that host ranges tend to be broader than what would be obtained from field studies (King & Cable, 2007; Poulin & Keeney, 2008; King et al., 2009). Moreover, under non-experimental conditions (outside library), host-switching can be triggered if host species are in captive environments (Thoney and Hargis Jr, 1991). For instance, if conditions are not optimal in aquaculture systems, certain hosts may become more vulnerable to atypical parasite species, or when different species are kept together, species switches may happen (Justine, Dupoux and Cribb, 2009; Kaneko et al., 1988).

Studies have suggested that host specificity indices not only should examine the number of host species a parasite infects, but also should investigate host phylogenies. Moreover, a non-specificity index has been proposed by Desdevises et al. (2002) which, apart from specialist and generalist classes of parasites, includes also two semi-quantitative classes. Later, an index of host specificity was described from a phylogenetic perspective based on the average phylogenetic or taxonomic distance between pairs of host species (Poulin and Mouillot, 2003).

Besides host phylogeny, other factors also affect parasite transfer. To understand the evolution and ecology of host specificity, Poulin, (2011) proposed an index of functional host specificity (obtained from quantification of similarities in functional traits among host species). Host specificity may also induce co-speciation. Although co-speciation plays an important role in parasite speciation, it does not act as a barrier to host-switching (Brooks and McLennan, 1993; Desdevises et al., 2002; Poulin, 1992).

Host switching

Host-switching or species jumps can result in parasite diversification, too (Munday et al., 2004; Sorenson et al., 2003; Zietara and Lumme, 2002). If a parasite species can successfully shift to a new host, a sudden surge in speciation may be stimulated (Zietara and Lumme, 2002). Host-switching events may also provide information about evolutionary sympatric phases between host species (Barson et al., 2010).

Evaluating the frequency and possibility of host switching in the evolutionary relationships of hosts and parasites may contribute to a deeper understanding of the changing nature of current infectious diseases and future potentially emerging diseases (Cooper et al., 2012; Streicker et al., 2010). In the same vein, as proposed by Poulin *et al*. (2011), to predict the possibility of a newly-

introduced parasite to switch to a new host in a new region, a phylogenetic host specificity index would be useful (Cooper et al., 2012).

Phylogenies, host specificity and host switching

Biological characteristics of hosts and parasites can be mapped onto phylogenetic trees and evolutionary reconstructions and inferences about the evolution of host use and specificity can be made (Desdevises et al., 2002). Dating estimates of specific evolutionary events can also be made based on molecular phylogenies (Barraclough and Nee, 2001). Statistical phylogenetic comparison of host-parasite relationships can reconstruct the ancestral host of certain parasite species. Host-parasite phylogenies are resolved in co-phylogenetic studies by evaluating the contribution of host switching, co-speciation, within-host speciation, and species sorting events ('missing the boat' or extinction of a parasite lineage) (Morand et al., 2015; Page and Charleston, 1998). Among these events, strict co-speciation results in complete congruence between host-parasite phylogenies, whereas the remaining three, coupled with incomplete taxon sampling, will blur such phylogenies (Domingues and Boeger, 2005).

Understanding the evolution of host specificity requires knowledge on covariations in biological traits and environmental factors of hosts and parasites in relation to levels of host specificity. To acquire such knowledge, phylogenetic comparative studies (Cooper et al., 2012; Desdevises et al., 2002) or Bayesian statistics (Faria et al., 2013) can be used. Cooper et al. (2012) applied phylogenetic comparative analyses to examine the underlying factors in parasite host switching among primate species. Their meta-analysis consisted of 2867 host–parasite combinations (including 128 primate and 437 parasite species) and compared evolutionary models with different levels of host gain (host switching) or loss (parasite extinction). A Bayesian statistical framework was used by Faria et al. (2013) to provide a reconstruction of host switching and transmission of the rabies virus in host bats. They also assessed the contribution of several host-parasite traits and environmental predictors. Cooper et al. (2012) and Faria et al. (2013) both found geographical range overlap and genetic relatedness as the most significant predictors of host switching, whereas host ecology was additionally found as an important factor by Cooper et al. (2012).

Choice of model systems

Two biologically rather different parasite systems (*Ligula* tapeworms and human associated *Cimex* bedbugs) were chosen as models for the thesis. Despite their differences, the two systems represent ideal models to study the processes of ecological differentiation – they are both globally distributed and subjected to strong host-related selection. In the case of *Ligula*, the selection pressure is associated with the large numbers of host species it is capable of infecting (Wafa Bouzid et al., 2008; Štefka et al., 2009). In the case of bedbugs, it is represented by the evolutionary recent host switch from bats to humans and by the very recent selection for resistance against insecticides (Kilpinen et al., 2011; Roth et al., 2019; Tawatsin et al., 2011). Each of these processes has the potential to leave imprints in the parasites' genomes. Peculiarities of the biology and life histories of the two parasite systems are provided below.

Ligula intestinalis

The tapeworm *L. intestinalis* (Cestoda: Diphyllobothriidea) is a permanent endoparasite with a complex life cycle (Figure 1), which begins with the release of eggs into freshwater environment. The eggs are produced by adult tapeworms living in the intestines of their definitive hosts, typically piscivorous birds, and passed out with the bird's faeces. In freshwater, the eggs eventually hatch into free-swimming larvae known as coracidia, which are then ingested by the first intermediate host, small crustaceans, mainly cyclopoid and calanoid copepods. Inside the copepod, the coracidia pass through the digestive canal and penetrate into the body, transforming into procercoid larvae. This stage shows significant morphological changes, preparing the parasite for its next stage of development.

Fish, most often cyprinids, serve as the second intermediate host. When a fish consumes an infected copepod, the procercoid larvae are released into the fish's digestive tract. The larvae then migrate through the gut into the body cavity of the fish, where they develop into plerocercoid larvae. This stage is marked by a significant increase in size and the development of a more complex structure, including the formation of a scolex (head) that will eventually attach to the intestine of the definitive host. The life cycle of *L. intestinalis* reaches its final stage when an infected fish is eaten by a piscivorous bird. Inside the bird's digestive system, the plerocercoid larvae attach to the intestinal wall with their scolex. Here, they mature into adult tapeworms, capable of reproducing and releasing eggs into the environment (Dubinina, 1980).

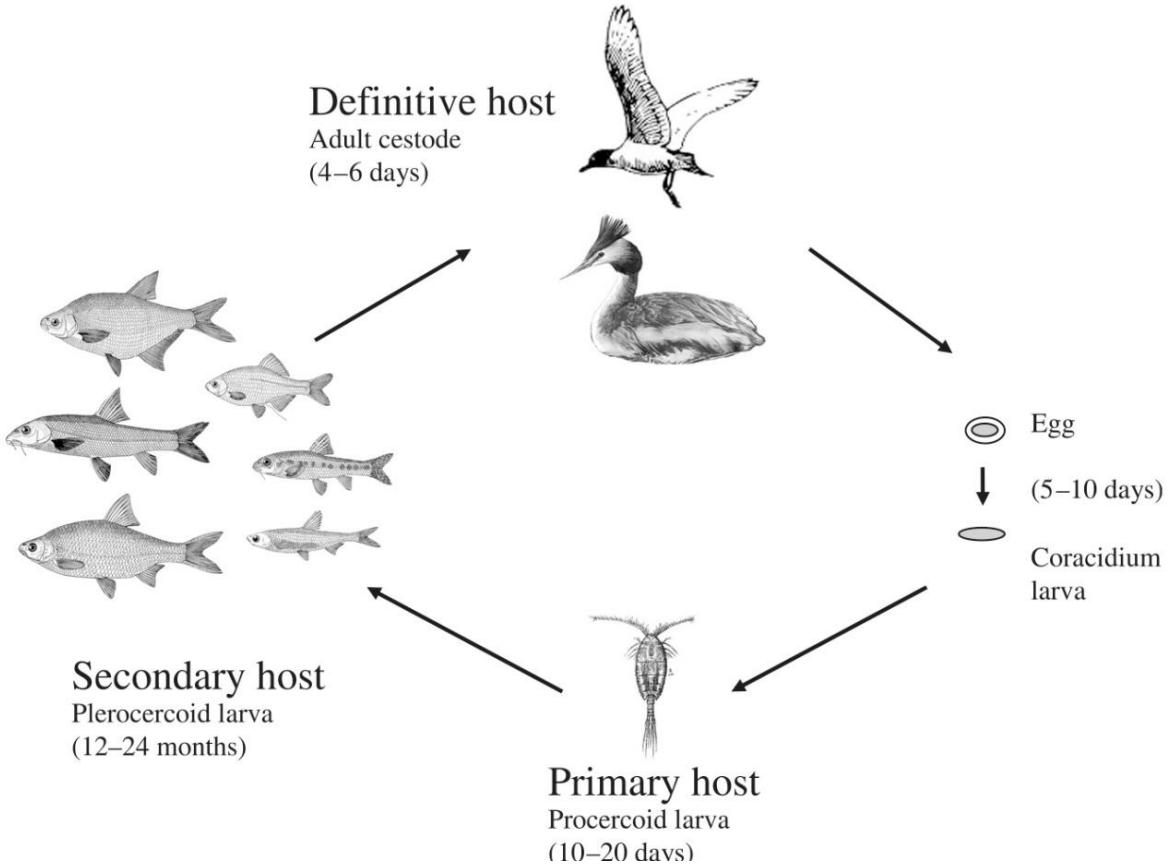


Figure 1. The life cycle of *L. intestinals* (from figure 1 in Stefka *et al* 2009).

Since the beginning of the 21st century, several studies have used molecular data to investigate speciation in ligulid cestodes (see Table 1, which summarizes recent ecological and evolutionary studies carried out on the tapeworm), which are represented by several lineages, sometimes considered as cryptic taxa. For example, Li and Liao, (2003) analysed sequences of mitochondrial ND1, CO1, ITS1, and 28S rRNA and found low genetic differentiation in these four genes in *Ligula intestinalis* and its relative *Digramma interrupta* (now *Ligula alterans*). According to their results, the two parasites in the genus *Ligula* should be considered distinct species. In an earlier study using CO1 and 28S rRNA sequences of *Ligula* from Russia, England, and the Qinghai-Tibet Plateau, the same authors suggested that *Ligula* from China belonged to the same species as the European *Ligula*. Based on their study, host habitat, host affinity, and geographic location could not be considered as fit and proper criteria for the classification of *Ligula* (Li et al., 2000).

Based on rDNA molecular data (ITS2 spacer and 28S), *Ligula* parasites in roach and gudgeon fish from Northern Ireland were found to represent distinct strains considering their effect on the gonadal development of their hosts (Olson et al., 2002). Also, *Ligula* in minnow (*Phoxinus phoxinus*) from Wales were similar to *Ligula* from roach fish in Lough Neagh. Olson et al. study proposed that the introduction of roach fish and a subsequent increase in primary host populations (i.e. great crested grebe) was likely the cause for the presence of separate strains in Lough Neagh.

Another study suggested that *L. intestinalis* from Turkey were genetically separate from Chinese and European *L. intestinalis* (Logan et al., 2004). Bouzid et al. () studied the genetic diversity of *L. intestinalis* from different fish species distributed across China, Russia, UK, Germany, France, Poland, Ukraine, Czech Republic, Estonia, Australia, Canada, Mexico, Tunisia, Algeria, and Ethiopia using two mitochondrial markers, CO1 and cyt *b*, and one nuclear sequence of ITS2. In their study, the evolutionary patterns were estimated at the global and local aspects. At the global level, genetically distinct clusters were obtained, while at the local scale, the migratory primary bird host was found to play a crucial role in preventing the establishment of genetic barriers. Moreover, *Ligula* diverged into two lineages, termed as A and B. Lineage A included *Ligula* from Europe and Tunisia that were collected from 'derived cyprinid fish' (*Phoxinus*, *Abramis*, *Rutilus, Scardinius*, and *Alburnus*), whereas lineage B contained samples from Europe, Algeria, China and Australasia obtained from 'basal cyprinids' (*Barbus*, *Gobio* and *Rhodeus*). In another study on the genetic diversity of *L. intestinalis*, Bouzid et al. () highlighted that inter-simple sequence repeat markers (ISSR) provided a quick and affordable approach to identify reliable markers for assessments of genetic diversity.

Then, Stefka et al. (2009) applied 15 microsatellite loci to assess the genetic diversity among *L. intestinalis* populations across Africa, Asia, Australasia, Europe, and North America. Their results indicated strong genetic structure and considerable levels of polymorphism in *Ligula* and suggested that the parasites subdivided between Tunisia and Europe under the influence of the Mediterranean Sea. Despite the fact that the sea was not a barrier to migration of the primary bird hosts, it is noteworthy that the adult tapeworm only spend a couple of days in the body of the birds (Dubinina, 1980) and therefore could not disperse between the two areas. Furthermore, the authors proposed that the fish immune response may also play an essential role in the occurrence of host-specificity in the genotypic clades of *Ligula*.

So far, the majority of studies on immune responses to the last larval form, i.e. the infective stage, have been done only in roach. These studies revealed a strong cellular response to the parasite, involving several adaptive and innate humoral immune components (Hoole and Arme, 1982, 1983, 1986, 1988; Taylor and Hoole, 1989, 1993, 1994, 1995; Williams and Hoole, 1992, 1995). Nevertheless, in the face of the host response, significant differences to *Ligula* infection can be seen in gudgeon and roach fishes (Arme, 1997), also suggesting possible genetic diversity of *Ligula*. Overall, since Smyth's experiments on *in vitro* maturation of *L. intestinalis* in 1947, this metazoan parasite has proven a valuable model for in vitro studies of tapeworm development and has been extensively used in parasitological, endocrinological, ecological, genetic, immunological and pollution studies.

Table 1 Summary of ecological and evolutionary studies on *L. intestinalis*

Research questions	Material and methods	Results	References
Determining the parameters of parasite-host relation in *Ligula* to age classes, month sampled and gender of tench in Mogan Lake, Turkey.	272 Cyprinid fishes were checked	A sharp decrease in age class 2 July has the highest rate of infections There is no significant rate of prevalence based on gender	(Ergonul and Altindag, 2005)
Occurrence of *L. intestinalis* in British fishes regarding month and fish size.	A large group of fishes were tested from three lakes in the United Kingdom	*Rutilus rutilus* was the most common and widespread host for the parasite August showed the highest rate of fish infection The most infected fish was smaller than 13 cm	(Arme and Owen, 1968)
The impact of the tapeworm *L. intestinalis* on behaviour of its intermediate fish hos.t	The statistical and analytical tools were combined to generate GLMs[2] models	Infected fish uses different microbiome (surface water) and alters morphology The size of parasite significantly correlates to disproportionate impact on its host	(Brown et al., 2001)
Does *L.intestinalis* affect its host (roach) behavior under natural condition?	Parasite occurrence, habitat preference and fish host age were estimated for one year using GLM	Three-years old roach were significantly parasitized The competitive ability of roach is reduced by the parasite Both parasite occurrence and abundance occur close to the shoreline from July to December	(Loot et al., 2001)

[2] Generalized Liner Model

The co-existence of separate species of *Ligula* in Northern Ireland.	ITS-1, 5.8S and ITS-2 were sequenced and compared in the parasite from difference host	The co-existence of separate species of *Ligula* resulted from the introduction of *R. rutlius* and correlated with an increasing number *Podiceps cristatus*	(Olson et al., 2002)
Phylogenetic systematics of Diphyllobothriid tapeworms.	20 ITS-2 RNA sequences from different host and geographical regions were considered	Taxonomic revision is needed for the family of Diphyllobothriidae *Digramma* is an invalid genus in the family Diphyllobothriidae	(Logan et al., 2004)
Examine the phylogenetic structure of *L intestinalis* on macroevolutionary scale and exploring the possibly cryptic species, and spatial distribution of genetic variation within populations.	109 tapeworms from 13 fish host in 18 localities were for Cytb and ITS-2 regions	Reproductive isolation (ecological speciation) was discovered between lineages in sympatric region The genetic structure of *Ligula* is affected by geographic structure and host specificity	(Bouzid *et al.*, 2008)
Describing the ultra-structure of spermiogenesis and spermatological of *L. intestinalis* in order to distinguish Diphyllobothriidae from Bothriocephalidae.	Transmission electron microscopy was used to describe spermiogenesis and spermatozoon	These two orders are different in the presence or absence of a ring of electron-dense cortical microtubes in the anterior extremity of the spermatozoon	(Levron et al., 2009)

Clarifying diversification processes within *L. intestinalis*	15 microsatellite loci and COI gene were analysed for 338 plerocercoids from 24 localities	The genetic structure of *L. intestinalis* complex is affected by geographic distance SSR[3] marker proved a divergence between European and Tunisian populations Mediterranean Sea was considered as a barrier between north Africa and European populations	Stefka et al., 2009
Characterize the complete mitogenomes of *Digramma interrupta* and *L. intestinalis* to investigate whether these two diphyllobothriids belong to the same genus or not?	Whole mitochondrial genome was sequenced for these two species	The study proved that *D. interrupta* and is a congeneric species with *L.intestinals* The most mitochondrial variable genes in these two species are nad2, nad4, nad5, nad6 and cox	(Li et al., 2019)
Taxonomic position and phylogenetic relationship between *L. intestinalis* and *Digramma interrupta*	1031 ciprinid fishesh were checked in northwestern of Iran and phylogeny analysis were carried out with 23 ITS sequences	*Alburnoides bipunctatus* showed the highest rate of infection ITS-2 is a suitable marker for distinguishing *L. intestinalis* from *D. interrupta* *D. interrupta* might be another species of the *Ligula* genus	(Ahmadiara et al., 2018)
Molecular identification of *Ligula* on the South Island, New Zealand.	Three ITS1 sequences of *Ligula* sp from the *Gobiomorphus cotidianus* and *Oncorhynchus tshawytscha* were considered for the study	The results confirmed that these samples belong to *Ligula* genus. The tape worm was reported after 30 years in the New Zealand	(Lagrue et al., 2018)

[3] Simple sequence repeat

Population genomic study on *L. intestinalis*

In a recent master's thesis study, a preliminary genomics dataset was generated using 4000+ Double digest restriction-site associated DNA sequencing (ddRADseq) loci obtained from 85 *L. intestinalis* individuals (Kočová, 2018). The results indicated that the genomic structure of the populations showed six major clusters for *L. intestinalis* across the globe (Figure 2). Contrary to the microsatellite analyses, SNP data suggested emerging genetic structure also among European populations arising from different host species (Figure 3). Both FST and PCA-based outlier tests on a sympatric population of *Ligula* from bream and roach+bleak fishes in northeast of the Czech Republic revealed that the genetic diversity could at least be partially associated with host specific adaptations (Figure 4).

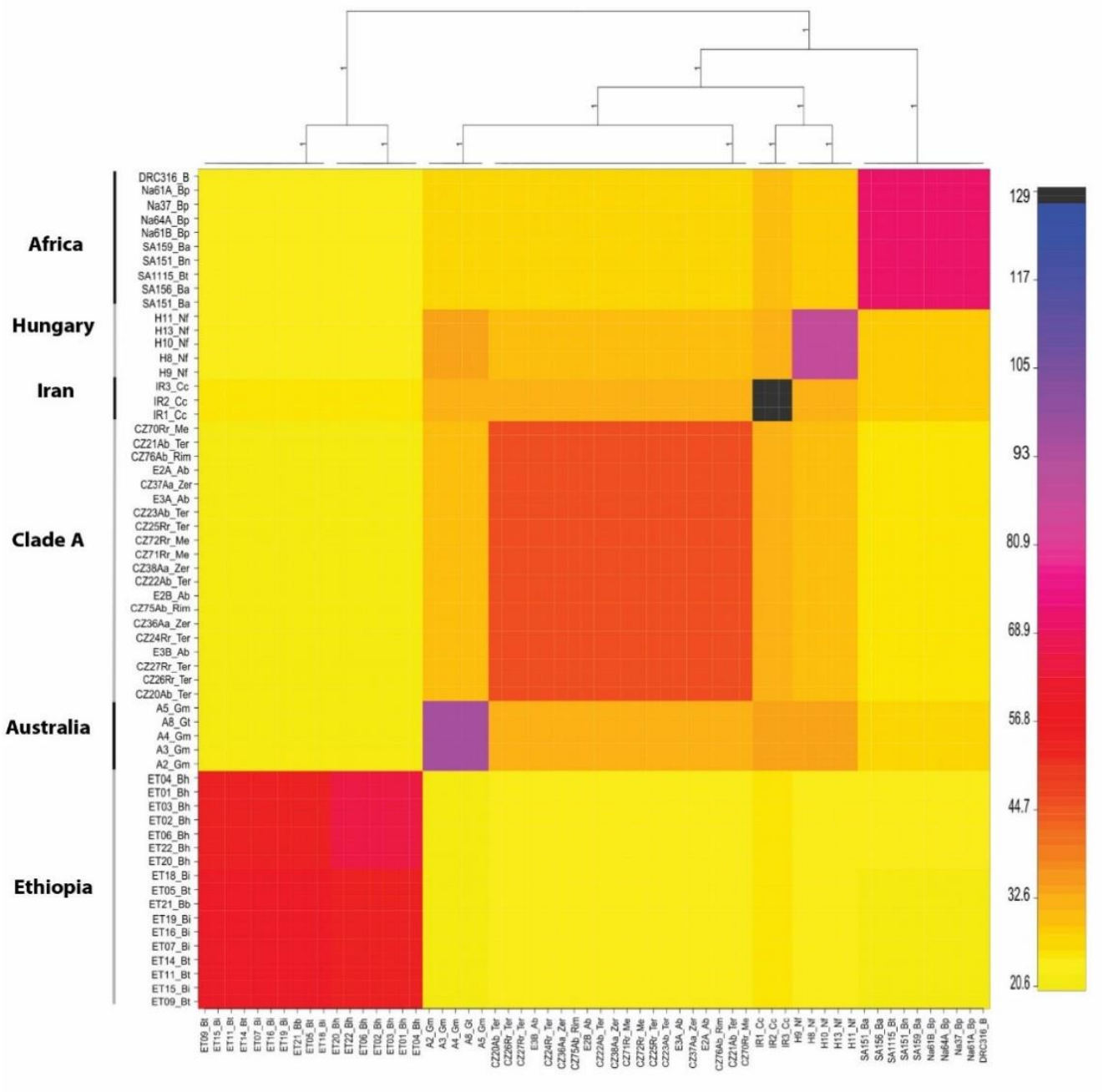


Figure 2. Co-ancestry heatplot of *L. intestinalis* populations across the world. The color scale indicates the degree of sharing identical loci between individuals. Gray and black lines on the left indicate the geographical origin of the samples (adapted from Kočová, 2018)

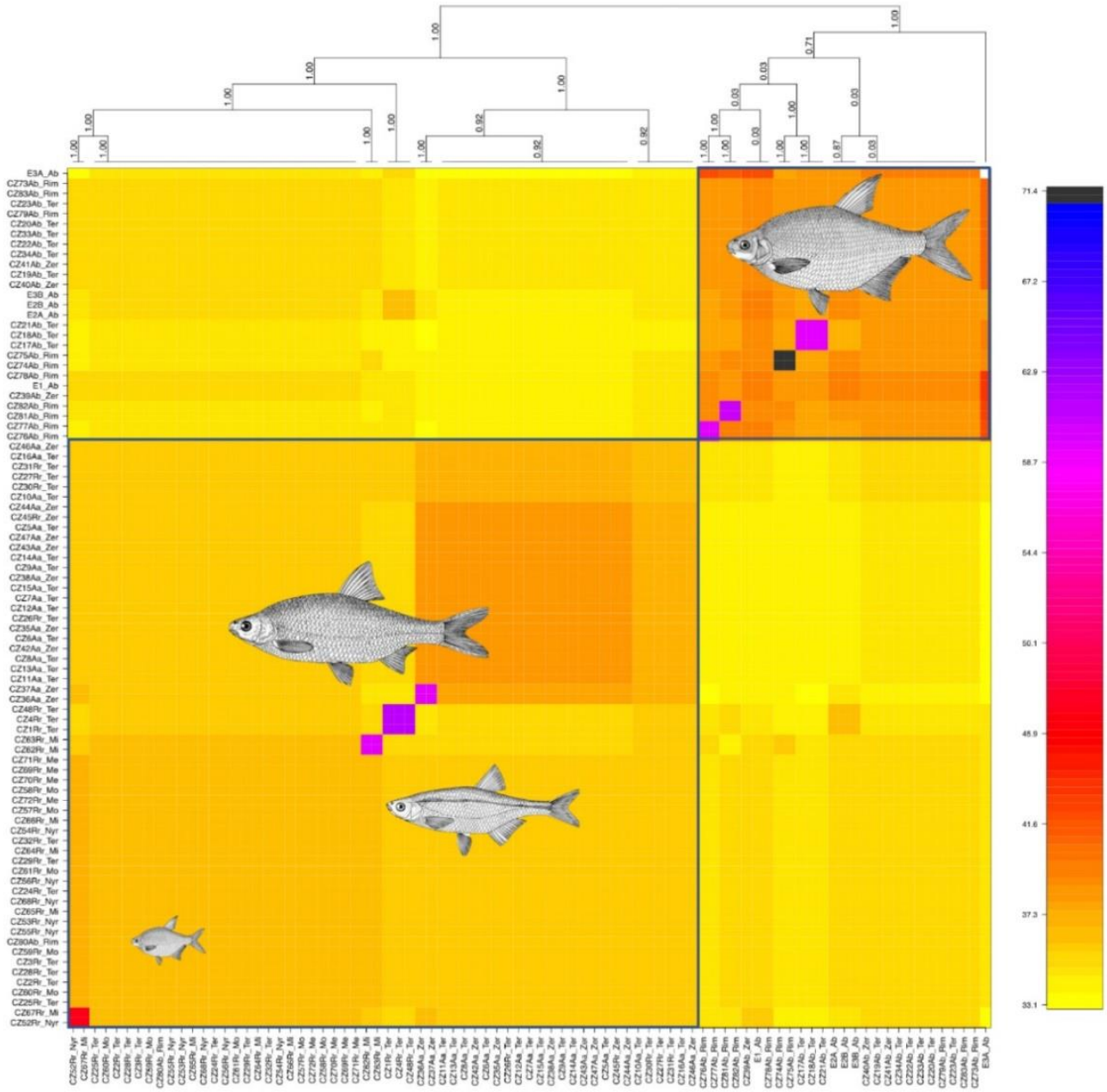


Figure 3. Genomic differentiation in European populations of *Ligula intestinalis* from two host groups. Co-ancestry matrix (heatmap) of 85 individuals distinguished a cluster from bream (top right), and a cluster from roach and bleak (bottom left). Only a single specimen (CZ80Ab_Rim) showed switching between the hosts (marked by a small bream logo on the bottom left) (adapted from Kočová, 2018).

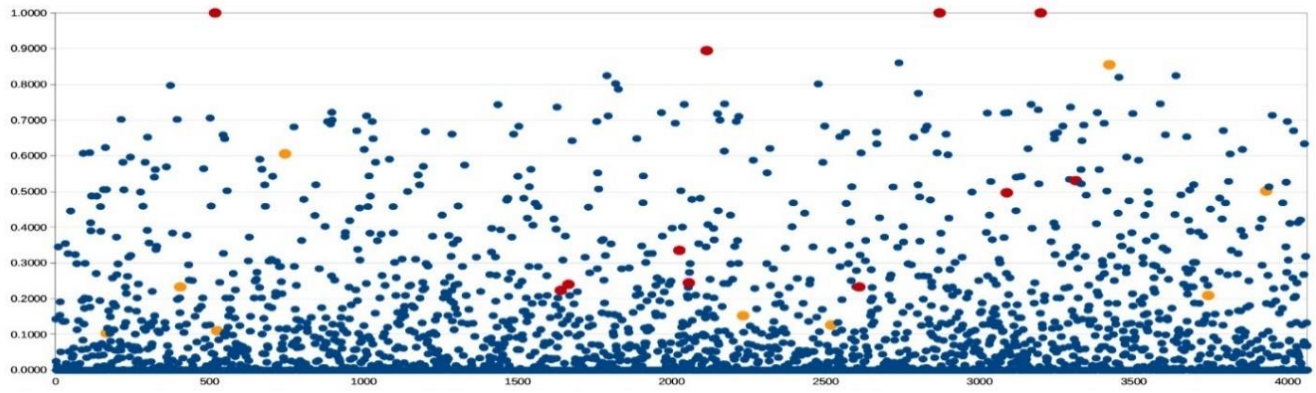


Figure 4. Manhattan plot of F_{ST} differentiation between roach+bleak vs. bream isolated samples of *L. intestinalis* (N=47, 4000+ loci) from a sympatric population in the NE of the Czech Republic. Red dots mark loci positive in both F_{ST} (BayeScan) and PCA (PCAdapt) tests, loci tested positive only in PCA are in yellow (strict, 10% q-value, FDR control was set for both tests) (adapted from Kočová, 2018).

Cimex lectularius* and *C. hemipterus

Bedbugs belong to the order Hemiptera (insects), family Cimicidae, and genus *Cimex*. These hematophagous arthropods feed on bats or birds as their definitive hosts. Globally, they are regarded as one of the major cosmopolitan pests parasitizing humans. Two species of bedbugs whose primary hosts are mostly humans are temporary blood-eating ectoparasites, the common bed bug (*Cimex lectularius* Linnaeus 1758) and the tropical bed bug (*C. hemipterus*). *C. lectularius* is found in temperate regions, while *C. hemipterus* is predominantly found in tropical regions (Usinger, 1966). These parasites of humans have emerged as a serious public health issue, due to their blood-eating feeding habits during the night, causing psychological and physical discomfort (Delaunay, 2012; Rossi and Jennings, 2010; Usinger, 1966). During infestation, the main health problems they may create include systemic involvement, skin reactions and nuisance biting. So far, no evidence of transmission of human diseases by bedbugs has been detected. Bedbugs affect all people from different economic and social levels, and these parasites have been documented in almost every man-made structure, e.g. apartments, college dormitories, hotels, motels, shelters, hospitals, office buildings, and military camps (Potter, 2011). Following World War II, organophosphate insecticides were extensively used to eliminate bed bug infestations, after which these parasites were only observed in rare cases from industrialized countries (Booth et al., 2012; Cooper, 2006).

In recent decades, the resurgence of bedbugs has been significantly driven by the emergence of insecticide resistance, posing substantial challenges to pest control efforts worldwide. The extensive use of insecticides, particularly pyrethroids, has led to the development of resistance in bed bug populations through various mechanisms, including behavioural changes, enhanced detoxification enzymes, and target site insensitivity (Davies et al., 2012; Romero et al., 2007). For instance, studies have identified widespread resistance to pyrethroids in bedbugs across the United States, attributed to mutations in the voltage-gated sodium channel gene (Zhu et al., 2010). Similar resistance patterns have been reported globally, complicating control measures (Dang et al., 2017; Kilpinen et al., 2011).

This resistance has facilitated the spread of bedbugs, exacerbating their status as major public health pests. Bedbugs have adapted to a variety of environments, including homes, hotels, public transportation, and healthcare facilities, leading to increased infestations and significant economic

burdens (Potter et al., 2008). Consequently, controlling resistant bed bug populations requires integrated pest management strategies that combine chemical treatments with non-chemical approaches, such as heat treatments, vacuuming, and bed bug-proof encasements (Romero et al., 2010). These integrated strategies are essential to effectively manage and mitigate the spread of these persistent pests.

Population genetics studies on Bedbugs

A recent fossil-dated phylogenetic approach revealed the evolutionary history and host relationships of bedbugs (Cimicidae) (Roth et al., 2019), suggesting that bedbugs originated around 115 million years ago, predating their assumed bat hosts by over 30 million years. Contrary to previous beliefs, which posited that bedbugs co-evolved directly with their hosts, particularly bats, and ancient humans (Usinger, 1966), the new evidence shows that bedbugs did not co-speciate with bats or ancient humans but rather independently colonized bats multiple times. The study also found that common and tropical bedbugs (*C. lectularius* and *C. hemipterus*) diverged approximately 47 million years ago, indicating that their association with humans occurred through host extension rather than co-speciation (Booth, 2019; Roth et al., 2019). Further findings revealed that bedbugs evolved as host specialists before diversifying to exploit various hosts, including birds and mammals. This pattern of host switching highlights the evolutionary plasticity of bedbugs and their ability to adapt to different hosts over millions of years. Additionally, the phylogenetic analysis provided robust resolutions of debated relationships within Cimicidae, including the paraphyly of groups parasitizing swallows and the monophyly of *Primicimex* and *Bucimex*. These results suggest that bedbugs' host associations are not strictly bound by co-evolutionary events but are shaped by ecological opportunities and host availability, challenging the simplistic view of direct co-speciation and emphasizing a more dynamic evolutionary history (Booth, 2019).

The human-associated *C. lectularius* has been a long-term pest of humans, originally evolving from the ectoparasites of their ancestral hosts, i.e. cave-dwelling bats (Usinger, 1966; Balvin, Roth and Vilimova, 2015). Recent population genetic studies of bedbugs address a wide range of questions (Table 2). A genetic study revealed that approximately 245,000 years ago, one or more lineages of *C. lectularius* diverged from their bat hosts and switched to hominids (Balvín et al., 2012). *C. lectularius* populations of the Old World still live in close association with their early

original bat hosts (Usinger, 1966; Povolný and Usinger, 1966; Balvín et al., 2012a). Recent studies have revealed genetic and reproductive isolation between *C. lectularius* lineages associated with humans and bats within homes (Balvín et al., 2012; Booth et al., 2015). Due to failure of hybridization experiments between these two lineages in producing viable offspring, it could be argued that post-mating barriers led to the divergence of the lineages from each other (Wawrocka et al., 2015). However, contrary to previous reports, recent studies have found that under laboratory conditions, these lineages are reproductively compatible, indicating that other factors such as ecological separation and behavioural barriers, rather than genetic incompatibility alone, might maintain their divergence (DeVries et al., 2017)

Mitochondrial DNA has been widely helpful in understanding propagule pressure (introduction effort or colonization pressure) of invasive populations, because the existence of several mtDNA lineages indicates the introduction of a genetically diverse propagule or multiple distinct introduction events (Fitzpatrick et al., 2012). In a study using sequences of 16S rDNA obtained from 22 geographically distinct infestations in Australia, Canada, and the United States, multiple haplotypes were observed within many of the sampled locations (Szalanski et al., 2014). Therefore, it was concluded that gene flow was commonly present among these populations, and bedbugs used alternative hosts in these regions that acted as sources of further infection before their recent re-emergence.

Furthermore, Microsatellite DNA from the nuclear genome yielded contrasting results on possible ways of bed bug spreading between infestations. In the first scenario, each microsatellite locus contains four or fewer alleles within an infestation, on the condition that the founding propagule was made up of a single sexual pair, a group of closely related individuals, or one singly-mated female (Scenario A). On the contrary, in scenario B, each locus contains more than four alleles at two or more loci, suggesting multiple introductions from more than one source population, or a single introduction from a source population with high genetic diversity. It is noteworthy that genetic diversity provides information about the number of introduction events that are genetically distinct but not about the number of introductions that can potentially occur from that source population. In addition, frequent use of insecticides for the control of bedbugs could eliminate haplotypes related to vulnerability to insecticides. As a result, the history of unique introductions will remain hidden (Dang et al., 2015). So far, little within-infestation genetic variation has been

revealed by microsatellite studies in human-associated populations (Booth et al., 2012; Fountain et al., 2014; Raab et al., 2016; Saenz et al., 2012). As most infestations contain four or fewer alleles within themselves (Raab *et al*., 2016), it appears that introductions follow scenario A, which is contrary to Szalanski et al., (2007), who observed multiple haplotypes within bed bug infestations using mitochondrial DNA markers.

In contrast, in the bat-associated lineage, observing multiple infestations supports scenario B (Booth *et al*., 2015). This result implied that populations of the bat-associated lineage were rarely exposed to insecticides, had higher prevalence in bat roosts, and also had greater host density, allowing more frequent passive dispersal events (Balvín et al., 2012; Bartonička and Růžičková, 2013). Thus, compared to the human-associated lineage, the bat-associated lineage is more likely to be stable, permitting more frequent introductions.

Next generation sequencing studies

Since the recent sequencing of the genome of *C. lectularius* (Rosenfeld et al., 2016; Benoit et al., 2016), the common bed bug has become an invaluable model for evolutionary studies. However, our knowledge is still rudimentary in some areas, such as 1) understanding the spatial and temporal frequency of mutations associated with knockdown resistance (*kdr*) to insecticides and identifying the genetic relationships between globally distributed populations, 2) the effect of increasing international human traffic on the extent of gene flow in the parasite, or 3) the loci involved in successful switch from bat to human hosts. For future studies, data from genome-wide loci (e.g. SNPs) and the sequenced genome are required in order to determine sources of resurgent populations and evolution of knockdown resistance and to understand which genes are responsible for host-associated lineage differentiation.

Table 2 Recent population genetic/genomic studies on bedbugs

Research Questions	Methods	Results	References
The identity of new C. *hemipterus* populations in Honolulu	Morphology and phylogeny methods were used for species identification	The results proved the presence of C. *hemipterus* populations in Honolulu where is the outside of species distribution They reported the presence of kdr-associated mutations in all samples	(Lewis et al., 2020)
The evolution of host associations and sex-reversal of reproductive trait diversification	Four genes region from both mtDNA and nucDNA were amplified for 34 species of Cimicidae	Bedbugs evolved before their primary bat hosts Male mating behaviour may have affected female genitalia heterotopy Rejected the simultaneous divergence between two bedbugs species with *Homo sapiens* and *H. erectus.*	(Roth *et al.*, 2019)
Co-evolution between bedbug and *Wolbachia*	Surface protein gene and another protein-coding locus were considered to track the relationship between bedbugs and *Wolbachia*	Showed co-speciation between *Wolbachia* and their bedbug hosts	(Balvín et al., 2018)
Scans for positive selection on salivary glands in bedbugs	CO1 gene and the nuclear Elongation Factor 1α (EF1α) gene were sequenced for bedbugs	The result depicted that a selective pressure on the salivary apyrase gene in species of *Cimex* A positive selection was found at five codons in all bat associated species	(Talbot et al., 2017)

Testing heteroplasmy in different bedbug populations	Bedbugs were sampled from 24 different infected structures and COI was sequenced	Showed that heteroplasmy is common however, no evidence for nuclear mitochondrial DNA sequences was detected	(Robison et al., 2015)
Taxonomy assessment of four groups of bedbugs (*Cimex. pipistrelli, C. pilosellus, C. hemipterus and C. adjunctus*)	Comparing morphological and molecular results of the four groups	Confirmed the existence of four traditional morphologically defined species in *Cimex* genus. *C. hemipterus* group showed a well-supported relationship to the *C. pipistrelli* group	(Balvin, Roth and Vilimova, 2015)
Estimating contemporary gene flow and genetic differentiation of *C. lectularius* populations associated with bats and humans	756 samples were used for both mtDNA (COI) and microsatellite	The results did not prove any contemporary gene flow among populations human-associated bed bug showed low genetic diversity in comparison to bat associated bedbugs Demonstrated that *C. lectularius* is currently undergoing lineage differentiation	(Booth et al., 2015)
Population genetic structure of C. *lectularius* in the US, Canada, and Australia	16S rRNA and ITS1 genes were sequenced for 136 individuals represented 22 populations	Showed that bedbugs populations did not experience any genetic bottleneck Lack genetic differentiation and high gene flow among populations	(Szalanski et al., 2007)
Spatial genetic structure and gene flow of C. *lectularius* populations in France	183 samples were genotyping using six polymorphic microsatellite loci	Showed a significant genetic differentiation among populations, however could not prove a significant relationship between genetic distance and geographic distance	(Akhoundi et al., 2015)

Phylogenetic relationship of tropical bedbugs in Malaysia	Sampling was carried out for 22 populations in Malaysia, and the mitochondrial (COI) gene was used as a marker.	The results depicted one monophyletic group for tropical bedbugs that separated in two lineages	(Seri Masran and Ab Majid, 2017)
Phylogeny of cimicid bugs of bats in central-eastern Europe, Vietnam and South Africa	216 cimicid bugs (*C. lectularius, C. pipistrelli, Cacodmus ignotus and C. sparsilis*) were collected from these regions, and two genes (COI and ITS) were sequenced	*C. pipistrelli* showed a high genetic homogeneity, however *C. lectularius* posed two haplotypes which significantly diverged from each other	(Hornok et al., 2017)
Identify potential genes involved in pesticide resistance	A total of 216,419 reads with 79,596,412 bp of bedbug transcriptome was surveyed on thin study	Revealed putative members of several detoxification pathways involved in pesticide resistance A significant number of snps and microsatellite loci were found	(Bai et al., 2011)
The mitogenome of the bed bug *C. lectularius* was surveyed	The phylogeny relationship was reconstructed based on the 15,217 bp mitochondrial genome	*C. lectularius* is a member of a paraphyletic Cimicomorpha lineage	(Kolokotronis et al., 2016)
Analyses of the C. *lectularius* sequenced genome to reveal evolutionary adaptations linked with the lifestyle of bedbugs	Genome assembly comprised 650.47Mb and 14,220 predicted protein-coding genes was used for three females and three adult male	Showed the presence of multiple putative lateral gene transfer events. Traumatic insemination has evolved multiple times	(Benoit et al., 2016)
Clarify the functional genomic profile of the bed bug regarding blood meals.	697.9-Mb C *lectularius* genome and transcriptome sequences including 36,985 coding and noncoding genes were used for both male and female in seven different stages of bedbugs life cycle.	Increased gene expression occurs after feeding on human blood Various methods of resistance to insecticides such as sodium channel blockers, enzyme detoxification pathways and thicker chitin layers were proved	(Rosenfeld et al., 2016)

Aims of the thesis

Diversification in parasites typically involves the evolution of new species or subspecies from a common ancestor, a process driven primarily by ecological pressures such as host specificity and geographic isolation (Huyse et al., 2005). This thesis aims to explore various aspects of ecological and genetic dynamics in parasitic species, focusing primarily on the tapeworm *L. intestinalis* and the bedbugs of species *C. lectularius and C. hemipterus*. The research is structured into several chapters, each targeting specific objectives to contribute to the overall understanding of these parasites' evolutionary biology, genetics, and adaptation mechanisms. My primary research goal is to explore the diversification processes in two parasites with biologically distinct systems (endoparasitic *Ligula* tapeworms infecting wildlife and ectoparasitic human-associated *Cimex* bedbugs) and to explore how population genomics can elucidate these mechanisms. Despite their differences, these two systems serve as ideal models for studying ecological differentiation, thanks to their global distribution and strong host-related selection pressures.

In Chapter I, I aim to elucidate the prevalence and population genetic structure of *L. intestinalis* plerocercoids in various cyprinoid fish species across different Czech water bodies. This research seeks to understand the prevalence of *L. intestinalis* plerocercoids in various freshwater lakes and reservoirs in the Czech Republic, and to compare the phylogenetic relationships of these parasite populations across different cyprinoid fish with specimens analysed in previous genetic studies. Additionally, using a subsample of the collected specimens, I aim to test for parasite genetic structure associated with host specificity on a smaller spatial scale. **In Chapter II**, I switch to a more global scale to investigate the historical dispersal and host-switching events that shaped the evolutionary history of the *L. intestinalis* species complex. This chapter focuses on understanding the relative importance of host association and biogeography in the parasite's diversification. Key research questions include the identification of major evolutionary lineages and their host specificity, the timing of major diversification events, and the influence of historical biogeographic and demographic events on lineage diversity. SNP genotyping and mitochondrial sequencing of 139 specimens from 18 fish-host genera across 21 countries, combined with phylogenetic and population genetic analyses, historical biogeography reconstructions, and demographic history modelling was

utilised to elucidate the factors and events that have driven the evolution and diversification of *L. intestinalis* over time.

In Chapter III, I aim to provide a highly contiguous genome assembly for *L. intestinalis* and to investigate RNA transcription patterns to understand the regulation of biological functions during its life cycle. This chapter seeks to establish a comprehensive reference transcriptome for both the plerocercoid and adult stages of *L. intestinalis*. Moreover, I aim to compare the genomic features of the new *Ligula* genome assembly with other available tapeworms in the order Diphyllobothriidea and six genomes representing the Cyclophyllidea. Furthermore, I seek to identify the upregulated and downregulated genes that play critical roles in the plerocercoid and adult stages of the *L. intestinalis* life cycle. Then in the next chapter (**Chapter IV**) the new genome assembly is employed to elucidate the mechanisms underlying possible ecological speciation in *L. intestinalis*, particularly focusing on Lineage A populations in areas where all hosts coexist in sympatry. I intend to examine the potential divergent selection pressures and ecological isolation mechanisms that might influence parasite adaptations and diversity in sympatric settings. By utilizing a comprehensive approach that involves the analysis of genome-wide SNPs and transcriptome data, this chapter aims to ascertain whether host specificity in Lineage A can potentially influence the population structure of parasites in the absence of geographical separation. Additionally, the goal is to identify potential genomic signatures indicative of host specialization within identified *Ligula* populations through selection analyses and to study RNA transcription patterns to identify differentially expressed genes related to host specialization in parasite populations.

In Chapter V, the focus is shifted to the establishment of the tropical bedbug, *Cimex hemipterus*, in Central Europe. This chapter investigates the recent local establishment of *C. hemipterus* through extensive sampling of bedbug populations across Europe from 2002 to 2020. The research utilizes morphological and genetic identification methods, including sequencing of the cytochrome oxidase subunit I (COI) gene and analysis of kdr-associated mutations in the sodium channel gene. Findings indicate recently established infestations, limited mtDNA variability and the presence of pyrethroid resistance mutations, emphasizing the importance of species identification in pest management due to potential differences in control strategies and resistance mechanisms. Finally, in **Chapter VI** I investigate the genetic and transcriptomic divergence in *Cimex lectularius* populations associated with human and bat hosts. This chapter hypothesizes that host switching through ecological fitting enables parasites

to exploit different host species using their existing traits, potentially leading to speciation. The study addresses how host switching influences genetic diversity and population structure in bedbugs, the genetic and transcriptomic adaptations associated with this host switching, and the potential for host specificity to drive speciation in bedbug populations.

References

Agosta, S.J., Brooks, D.R., 2020. The Stockholm Paradigm, in: The Major Metaphors of Evolution: Darwinism Then and Now. Springer, pp. 219–242. https://doi.org/10.1007/978-3-030-52086-1_10

Agosta, S.J., Klemens, J.A., 2008. Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. Ecol. Lett. 11, 1123–1134.

Ahmadiara, E., Hosseini, S.H., Jalousian, F., Ebrahimzadeh Mousavi, H.A., 2018. The Taxonomic Position and Phylogenetic Relationship between Digramma interrupta and Ligula intestinalis Based on Morphological and Molecular Diagnosis. Iran. J. Parasitol. 13, 648–654.

Akhoundi, M., Kengne, P., Cannet, A., Brengues, C., Berenger, J.-M., Izri, A., Marty, P., Simard, F., Fontenille, D., Delaunay, P., 2015. Spatial genetic structure and restricted gene flow in bed bugs (Cimex lectularius) populations in France. Infect. Genet. Evol. 34, 236–243.

Althoff, D.M., Segraves, K.A., Johnson, M.T.J., 2014. Testing for coevolutionary diversification: linking pattern with process. Trends Ecol. Evol. 29, 82–89.

Althoff, D.M., Thompson, J.N., 1999. Comparative geographic structures of two parasitoid-host interactions. Evolution (N. Y). 53, 818–825.

Andrew, R.L., Rieseberg, L.H., 2013. Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. Evolution (N. Y). 67, 2468–2482.

Araujo, S.B.L., Braga, M.P., Brooks, D.R., Agosta, S.J., Hoberg, E.P., von Hartenthal, F.W., Boeger, W.A., 2015. Understanding Host-Switching by Ecological Fitting. PLoS One 10, e0139225. https://doi.org/10.1371/journal.pone.0139225

Arme, C., 1997. Ligula intestinalis: interactions with the pituitary-gonadal axis of its fish host. J. Helminthol. 71, 83–84.

Arme, C., Owen, R.W., 1968. Occurrence and Pathology of Ligula intestinalis Infections in British Fishes. J. Parasitol. 54, 272. https://doi.org/10.2307/3276934

Ascunce, M.S., Toups, M.A., Kassu, G., Fane, J., Scholl, K., Reed, D.L., 2013. Nuclear genetic diversity in human lice (Pediculus humanus) reveals continental differences and high inbreeding among worldwide populations. PLoS One 8, e57619.

Asikainen, K., Hänninen, T., Henttonen, H., Niemimaa, J., Laakkonen, J., Andersen, H.K., Bille, N., Leirs, H., Vaheri, A., Plyusnin, A., 2000. Molecular evolution of Puumala hantavirus in Fennoscandia: phylogenetic analysis of strains from two recolonization routes, Karelia and Denmark. J. Gen. Virol. 81, 2833–2841.

Bai, X., Mamidala, P., Rajarapu, S.P., Jones, S.C., Mittapalli, O., 2011. Transcriptomics of the bed bug (Cimex lectularius). PLoS One 6, e16336. https://doi.org/10.1371/journal.pone.0016336

Balvín, O., Munclinger, P., Kratochvíl, L., Vilímová, J., 2012. Mitochondrial DNA and morphology show independent evolutionary histories of bedbug Cimex lectularius (Heteroptera: Cimicidae) on bats and humans. Parasitol. Res. 111, 457–469. https://doi.org/10.1007/s00436-012-2862-5

Balvín, O., Roth, S., Talbot, B., Reinhardt, K., 2018. Co-speciation in bedbug Wolbachia parallel the pattern in nematode hosts. Sci. Rep. 8, 1–9.

Balvín, O., Roth, S., Vilímová, J., 2015. Molecular evidence places the swallow bug genus Oeciacus Stål within the bat and bed bug genus Cimex Linnaeus (Heteroptera: Cimicidae). Syst. Entomol. 40, 652–665. https://doi.org/10.1111/syen.12127

Barraclough, T.G., Nee, S., 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16, 391–399.

Barson, M., Prikrylova, I., Vanhove, M.P.M., Huyse, T., 2010. Parasite hybridization in

African Macrogyrodactylus spp.(Monogenea, Platyhelminthes) signals historical host distribution. Parasitology 137, 1585.
Bartonička, T., Růžičková, L., 2013. Recolonization of bat roost by bat bugs (Cimex pipistrelli): could parasite load be a cause of bat roost switching? Parasitol. Res. 112, 1615–1622.
Benoit, J.B., Adelman, Z.N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E.C., Szuter, E.M., Hagan, R.W., Gujar, H., Shukla, J.N., Zhu, F., Mohan, M., Nelson, D.R., Rosendale, A.J., Derst, C., Resnik, V., Wernig, S., Menegazzi, P., Wegener, C., Peschel, N., Hendershot, J.M., Blenau, W., Predel, R., Johnston, P.R., Ioannidis, P., Waterhouse, R.M., Nauen, R., Schorn, C., Ott, M.C., Maiwald, F., Johnston, J.S., Gondhalekar, A.D., Scharf, M.E., Peterson, B.F., Raje, K.R., Hottel, B.A., Armisén, D., Crumière, A.J.J., Refki, P.N., Santos, M.E., Sghaier, E., Viala, S., Khila, A., Ahn, S.J., Childers, C., Lee, C.Y., Lin, H., Hughes, D.S.T., Duncan, E.J., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Wheeler, D., Panfilio, K.A., Vargas Jentzsch, I.M., Vargo, E.L., Booth, W., Friedrich, M., Weirauch, M.T., Anderson, M.A.E., Jones, J.W., Mittapalli, O., Zhao, C., Zhou, J.J., Evans, J.D., Attardo, G.M., Robertson, H.M., Zdobnov, E.M., Ribeiro, J.M.C., Gibbs, R.A., Werren, J.H., Palli, S.R., Schal, C., Richards, S., 2016. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat. Commun. 7, 1–10. https://doi.org/10.1038/ncomms10165
Benovics, M., Vukić, J., Šanda, R., Rahmouni, I., Šimková, A., 2020. Disentangling the evolutionary history of peri-Mediterranean cyprinids using host-specific gill monogeneans. Int. J. Parasitol.
Bermingham, E., Moritz, C., 1998. Comparative phylogeography: concepts and applications. Mol. Ecol. 7, 367–369.
Booth, W., 2019. Evolution: Bedbugs Evolved before Their Assumed Ancestral Host. Curr. Biol. 29, R413–R415. https://doi.org/10.1016/j.cub.2019.04.051
Booth, W., Balvín, O., Vargo, E.L., Vilímová, J., Schal, C., 2015. Host association drives genetic divergence in the bed bug, Cimex lectularius. Mol. Ecol. 24, 980–992. https://doi.org/10.1111/mec.13086
Booth, W., Saenz, V.L., Santangelo, R.G., Wang, C., Schal, C., Vargo, E.L., 2012. Molecular markers reveal infestation dynamics of the bed bug (Hemiptera: Cimicidae) within apartment buildings. J. Med. Entomol. 49, 535–546.
Bouzid, W, Lek, S., Mace, M., Ben Hassine, O., Etienne, R., Legal, L., Loot, G., 2008. Genetic diversity of Ligula intestinalis (Cestoda: Diphyllobothriidea) based on analysis of inter-simple sequence repeat markers. J. Zool. Syst. Evol. Res. 46, 289–296.
Bouzid, Wafa, Štefka, J., Hypša, V., Lek, S., Scholz, T., Legal, L., Hassine, O.K. Ben, Loot, G., 2008. Geography and host specificity: Two forces behind the genetic structure of the freshwater fish parasite Ligula intestinalis (Cestoda: Diphyllobothriidae). Int. J. Parasitol. 38, 1465–1479. https://doi.org/10.1016/j.ijpara.2008.03.008
Brooks, D.R., Hoberg, E.P., Boeger, W.A., 2019. The Stockholm Paradigm Climate Change and Emerging Disease. University of Chicago Press. https://doi.org/10.7208/chicago/9780226632582.001.0001
Brooks, D.R., McLennan, D.A., 1993. Comparative study of adaptive radiations with an example using parasitic flatworms (Platyhelminthes: Cercomeria). Am. Nat. 142, 755–778.
Brown, J.M., Leebens-Mack, J.H., Thompson, J.N., Pellmyr, O., Harrison, R.G., 1997. Phylogeography and host association in a pollinating seed parasite Greya politella (Lepidoptera: Prodoxidae). Mol. Ecol. 6, 215–224.
Brown, S.P., Loot, G., Grenfell, B.T., Guégan, J.F., 2001. Host manipulation by Ligula

intestinalis: accident or adaptation? Parasitology 123, 519.
Bueter, C., Weckstein, J., Johnson, K.P., Bates, J.M., Gordon, C.E., 2009. Comparative phylogenetic histories of two louse genera found on Catharus thrushes and other birds. J. Parasitol. 95, 295–307.
Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L.Z., 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 25, 1656–1665.
Caira, J.N., Jensen, K., 2001. An investigation of the co-evolutionary relationships between onchobothriid tapeworms and their elasmobranch hosts. Int. J. Parasitol. 31, 960–975.
Cháves-González, L.E., Morales-Calvo, F., Mora, J., Solano-Barquero, A., Verocai, G.G., Rojas, A., 2022. What lies behind the curtain: Cryptic diversity in helminth parasites of human and veterinary importance. Curr. Res. Parasitol. Vector-borne Dis. 2, 100094.
Clay, K., Kover, P.X., 1996. The red queen hypothesis and plant/pathogen interactions. Annu. Rev. Phytopathol. 34, 29–50. https://doi.org/10.1146/annurev.phyto.34.1.29
Clayton, D.H., Price, R.D., 1999. Taxonomy of New World Columbicola (Phthiraptera: Philopteridae) from the Columbiformes (Aves), with descriptions of five new species. Ann. Entomol. Soc. Am. 92, 675–685.
Cooper, N., Griffin, R., Franz, M., Omotayo, M., Nunn, C.L., 2012. Phylogenetic host specificity and understanding parasite sharing in primates. Ecol. Lett. 15, 1370–1377.
Cooper, R., 2006. Bed bugs--still more questions than answers: A need for research and public awareness. Am. Entomol. 52, 111–112.
Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates Sunderland, MA.
Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23, 3133–3157. https://doi.org/10.1111/mec.12796
Cutter, A.D., Payseur, B.A., 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14, 262–274.
Dallas, T., Holian, L.A., Foster, G., 2020. What determines parasite species richness across host species? J. Anim. Ecol. 89, 1750–1753. https://doi.org/10.1111/1365-2656.13276
Dang, K., Doggett, S.L., Veera Singham, G., Lee, C.-Y., 2017. Insecticide resistance and resistance mechanisms in bed bugs, Cimex spp. (Hemiptera: Cimicidae). Parasit. Vectors 10, 318. https://doi.org/10.1186/s13071-017-2232-3
Dang, K., Lilly, D.G., Bu, W., Doggett, S.L., 2015. Simple, rapid and cost-effective technique for the detection of pyrethroid resistance in bed bugs, Cimex spp. (Hemiptera: Cimicidae). Austral Entomol. 54, 191–196. https://doi.org/10.1111/aen.12109
Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12, 499–510.
Davies, T.G.E., Field, L.M., Williamson, M.S., 2012. The re-emergence of the bed bug as a nuisance pest: Implications of resistance to the pyrethroid insecticides. Med. Vet. Entomol. 26, 241–254. https://doi.org/10.1111/j.1365-2915.2011.01006.x
De Kort, H., Baguette, M., Prunier, J.G., Tessier, M., Monsimet, J., Turlure, C., Stevens, V., 2018. Genetic costructure in a meta-community under threat of habitat fragmentation. Mol. Ecol. 27, 2193–2203.
De Meeûs, T., Renaud, F., 2002. Parasites within the new phylogeny of eukaryotes. Trends Parasitol. 18, 247–251.
De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.
Dekonenko, A., Yakimenko, V., Ivanov, A., Morozov, V., Nikitin, P., Khasanova, S., Dzagurova, T., Tkachenko, E., Schmaljohn, C., 2003. Genetic similarity of Puumala

viruses found in Finland and western Siberia and of the mitochondrial DNA of their rodent hosts suggests a common evolutionary origin. Infect. Genet. Evol. 3, 245–257.
Delaunay, P., 2012. Human travel and traveling bedbugs. J. Travel Med. 19, 373–379.
Desdevises, Y., Morand, S., Jousson, O., Legendre, P., 2002. Coevolution between Lamellodiscus (Monogenea: Diplectanidae) and Sparidae (Teleostei): the study of a complex host-parasite system. Evolution (N. Y). 56, 2459–2471.
DeVries, Z., Mick, R., Balvín, O., Schal, C., 2017. Aggregation behavior and reproductive compatibility in the family Cimicidae. Sci. Rep. 7, 1–10.
Dobson, K.S., Hollon, S.D., Dimidjian, S., Schmaling, K.B., Kohlenberg, R.J., Gallop, R.J., Rizvi, S.L., Gollan, J.K., Dunner, D.L., Jacobson, N.S., 2008. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. J. Consult. Clin. Psychol. 76, 468.
Dobzhansky, T.H., 1936. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. Genetics 21, 113.
Domingues, M. V, Boeger, W.A., 2005. Neotropical Monogenoidea. 47. Phylogeny and coevolution of species of Rhinoxenus (Monogenoidea, Ancyrocephalinae) and their Characiformes hosts (Teleostei, Ostariophysi) with description of four new species. Zoosystema 27, 441–467.
Dubinina, M.N., 1980. Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR. Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR.
Dunn, R.R., Harris, N.C., Colwell, R.K., Koh, L.P., Sodhi, N.S., 2009. The sixth mass coextinction: are most endangered species parasites and mutualists? Proc. R. Soc. B Biol. Sci. 276, 3037–3045.
Durrer, S., 1996. Parasite load and assemblages of bumblebee species.
Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature 491, 756–760.
Emerson, B.C., Hewitt, G.M., 2005. Phylogeography. Curr. Biol. 15, R367–R371.
Engelbrecht, A., Matthee, S., Du Toit, N., Matthee, C.A., 2016. Limited dispersal in an ectoparasitic mite, Laelaps giganteus, contributes to significant phylogeographic congruence with the rodent host, Rhabdomys. Mol. Ecol. 25, 1006–1021.
Ergonul, M.B., Altindag, A., 2005. The occurrence and dynamics of Ligula intestinalis in its cyprinid fish host, tench, Tinca tinca, in Mogan Lake (Ankara, Turkey). Vet. Med. 50, 537.
Faria, N.R., Suchard, M.A., Rambaut, A., Streicker, D.G., Lemey, P., 2013. Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. Philos. Trans. R. Soc. B Biol. Sci. 368, 20120196.
Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28, 342–350.
Fenwick, A., 2012. The global burden of neglected tropical diseases. Public Health 126, 233–236. https://doi.org/10.1016/j.puhe.2011.11.015
Feulner, P.G.D., Chain, F.J.J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., Lenz, T.L., Samonte, I.E., Stoll, M., Bornberg-Bauer, E., 2015. Genomics of divergence along a continuum of parapatric population differentiation. PLoS Genet 11, e1004966.
Fountain, T., Duvaux, L., Horsburgh, G., Reinhardt, K., Butlin, R.K., 2014. Human-facilitated metapopulation dynamics in an emerging pest species, C imex lectularius. Mol. Ecol. 23, 1071–1084.
Garner, B.A., Hand, B.K., Amish, S.J., Bernatchez, L., Foster, J.T., Miller, K.M., Morin, P.A., Narum, S.R., O'Brien, S.J., Roffler, G., 2016. Genomics in conservation: case studies and bridging the gap between data and application. Trends Ecol. Evol. 31, 81–83.

Harbison, C.W., Clayton, D.H., 2011. Community interactions govern host-switching with implications for host–parasite coevolutionary history. Proc. Natl. Acad. Sci. 108, 9525–9529.
Hechinger, R.F., Lafferty, K.D., 2005. Host diversity begets parasite diversity: bird final hosts and trematodes in snail intermediate hosts. Proc. R. Soc. B Biol. Sci. 272, 1059–1066.
Hoberg, E.P., Adams, A., 2000. Phylogeny, history and biodiversity: Understanding faunal structure and biogeography in the marine realm. Bull. Scand. Soc. Parasitol. 10, 19–37.
Hoberg, E.P., Adams, A.M., 1992. Phylogeny, historical biogeography, and ecology of Anophryocephalus spp.(Eucestoda: Tetrabothriidae) among pinnipeds of the Holarctic during the late Tertiary and Pleistocene. Can. J. Zool. 70, 703–719.
Hoberg, E.P., Brooks, D.R., 2013. 14 Episodic processes, invasion and faunal mosaics in evolutionary and ecological time. Balanc. Nat. Hum. impact 199.
Hoberg, E.P., Brooks, D.R., 2008. A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host–parasite systems. J. Biogeogr. 35, 1533–1550.
Hoffman, A., 1991. Testing the Red Queen Hypothesis. J. Evol. Biol. 4, 1–7. https://doi.org/10.1046/j.1420-9101.1991.4010001.x
Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A., 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6, e1000862.
Hohenlohe, P.A., Hand, B.K., Andrews, K.R., Luikart, G., 2018. Population genomics provides key insights in ecology and evolution, in: Population Genomics. Springer, pp. 483–510.
Hornett, E.A., Wheat, C.W., 2012. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. BMC Genomics 13, 361.
Hornok, S., Szőke, K., Boldogh, S.A., Sándor, A.D., Kontschán, J., Tu, V.T., Halajian, A., Takács, N., Görföl, T., Estók, P., 2017. Phylogenetic analyses of bat-associated bugs (Hemiptera: Cimicidae: Cimicinae and Cacodminae) indicate two new species close to Cimex lectularius. Parasit. Vectors 10, 439.
House, N., Miró-Herrans, A.T., Ashfaq, M., Braig, H., 2023. Genome-wide analysis of the human head louse (Pediculus humanus capitis) reveals geographically structured genetic populations 1–22.
Hunter, M.E., Hoban, S.M., Bruford, M.W., Segelbacher, G., Bernatchez, L., 2018. Next-generation conservation genetics and biodiversity monitoring. Evol. Appl. 11, 1029–1034.
Huyse, T., Poulin, R., Théron, A., Theron, A., 2005. Speciation in parasites: a population genetics approach. Trends Parasitol. 21, 469–475.
Johnson, K.P., Kennedy, M., McCracken, K.G., 2006. Reinterpreting the origins of flamingo lice: cospeciation or host-switching? Biol. Lett. 2, 275–278.
Joseph, L., Wilke, T., Alpers, D., 2002. Reconciling genetic expectations from host specificity with historical population dynamics in an avian brood parasite, Horsfield's Bronze-Cuckoo Chalcites basalis of Australia. Mol. Ecol. 11, 829–837.
Jousson, O., Pawlowski, J., Zaninetti, L., Zechman, F.W., Dini, F., Di Guiseppe, G., Woodfield, R., Millar, A., Meinesz, A., 2000. Invasive alga reaches California. Nature 408, 157–158.
Juhásová, L., Králova-Hromadova, I., Bazsalovicsová, E., Minárik, G., Štefka, J., Mikulíček, P., Pálková, L., Pybus, M., 2016. Population structure and dispersal routes of an invasive parasite, Fascioloides magna, in North America and Europe. Parasites and Vectors 9, 1–19. https://doi.org/10.1186/s13071-016-1811-z
Justine, J.-L., Dupoux, C., Cribb, T., 2009. Resolution of the discrepant host-specificity of Pseudorhabdosynochus species (Monogenea, Diplectanidae) from serranid fishes in the

tropical Indo-Pacific. Acta Parasitol. 54, 119–130.
Kamiya, T., O'Dwyer, K., Nakagawa, S., Poulin, R., 2014. Host diversity drives parasite diversity: meta-analytical insights into patterns and causal mechanisms. Ecography (Cop.). 37, 689–697. https://doi.org/10.1111/j.1600-0587.2013.00571.x
Kilpinen, O., Kristensen, M., Vagn Jensen, K.M., 2011. Resistance differences between chlorpyrifos and synthetic pyrethroids in Cimex lectularius population from Denmark. Parasitol. Res. 109, 1461–1464. https://doi.org/10.1007/s00436-011-2423-3
Kolokotronis, S.-O., Foox, J., Rosenfeld, J.A., Brugler, M.R., Reeves, D., Benoit, J.B., Booth, W., Robison, G., Steffen, M., Sakas, Z., 2016. The mitogenome of the bed bug Cimex lectularius (Hemiptera: Cimicidae). Mitochondrial DNA Part B 1, 425–427.
Korhonen, P.K., Kinkar, L., Young, N.D., Cai, H., Lightowlers, M.W., Gauci, C., Jabbar, A., Chang, B.C.H.H., Wang, T., Hofmann, A., Koehler, A. V., Li, J.J., Li, J.J., Wang, D., Yin, J., Yang, H., Jenkins, D.J., Saarma, U., Laurimäe, T., Rostami-Nejad, M., Irshadullah, M., Mirhendi, H., Sharbatkhori, M., Ponce-Gordo, F., Simsek, S., Casulli, A., Zait, H., Atoyan, H., de la Rue, M.L., Romig, T., Wassermann, M., Aghayan, S.A., Gevorgyan, H., Yang, B., Gasser, R.B., 2022. Chromosome-scale Echinococcus granulosus (genotype G1) genome reveals the Eg 95 gene family and conservation of the EG95-vaccine molecule. Commun. Biol. 5, 199. https://doi.org/10.1038/s42003-022-03125-1
Králová-Hromadová, I., Bazsalovicsová, E., Štefka, J., Špakulová, M., Vávrová, S., Szemes, T., Tkach, V., Trudgett, A., Pybus, M., 2011. Multiple origins of European populations of the giant liver fluke Fascioloides magna (Trematoda: Fasciolidae), a liver parasite of ruminants. Int. J. Parasitol. 41, 373–383. https://doi.org/10.1016/j.ijpara.2010.10.010
Krasnov, B.R., Shenbrot, G.I., Khokhlova, I.S., Allan Degen, A., 2004. Relationship between host diversity and parasite diversity: flea assemblages on small mammals. J. Biogeogr. 31, 1857–1866.
Lagrue, C., Presswell, B., Dunckley, N., Poulin, R., 2018. The invasive cestode parasite Ligula from salmonids and bullies on the South Island, New Zealand. Parasitol. Res. 117, 151–156.
Levron, C., Sitko, J., Scholz, T., 2009. Spermiogenesis and Spermatozoon of the Tapeworm Ligula intestinalis (Diphyllobothriidea): Phylogenetic Implications. J. Parasitol. 95, 1–9. https://doi.org/10.1645/GE-1646.1
Lewis, C.D., Levine, B.A., Vargo, E.L., Schal, C., Booth, W., 2020. Recent Detection of Multiple Populations of the Tropical Bed Bug (Hemiptera: Cimicidae) Exhibiting kdr-Associated Mutations in Hawaii. J. Med. Entomol.
Li, J., Liao, X., 2003. The taxonomic status of Digramma (Pseudophyllidea: Ligulidae) inferred from DNA sequences. J. Parasitol. 89, 792–799.
Li, J., Liao, X., Yang, H., 2000. Molecular characterization of a parasitic tapeworm (Ligula) based on DNA sequences from formalin-fixed specimens. Biochem. Genet. 38, 309–322.
Li, W.-H., Yang, Y., Zhang, N.-Z., Wang, J.-K., Liu, Y.-J., Li, L., Yan, H.-B., Jia, W.-Z., Fu, B., 2021. Comparative transcriptome analyses of the developmental stages of Taenia multiceps. Front. Vet. Sci. 8, 677045. https://doi.org/10.3389/fvets.2021.677045
Liow, L.H., Van Valen, L., Stenseth, N.C., 2011. Red Queen: from populations to taxa and communities. Trends Ecol. Evol. 26, 349–358. https://doi.org/10.1016/j.tree.2011.03.016
Logan, F.J., Horák, A., Štefka, J., Aydogdu, A., Scholz, T., 2004. The phylogeny of diphyllobothriid tapeworms (Cestoda: Pseudophyllidea) based on ITS-2 rDNA sequences. Parasitol. Res. 94, 10–15.
Loot, G., Lek, S., Brown, S.P., Guégan, J.-F., 2001. Phenotypic modification of roach (Rutilus rutilus L.) infected with Ligula intestinalis L.(Cestoda: Pseudophyllidea). J. Parasitol. 87, 1002–1010. https://doi.org/10.1645/0022-3395(2001)087[1002:PMORRR]2.0.CO;2
Luikart, G., England, P.R., Tallmon, D., Jordan, S., Taberlet, P., 2003. The power and promise

of population genomics: from genotyping to genome typing. Nat. Rev. Genet. 4, 981–994.
Maia Da Silva, F., Junqueira, A.C. V, Campaner, M., Rodrigues, A.C., Crisante, G., Ramirez, L.E., Caballero, Z.C.E., Monteiro, F.A., Coura, J.R., Añez, N., 2007. Comparative phylogeography of Trypanosoma rangeli and Rhodnius (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. Mol. Ecol. 16, 3361–3373.
Martínez-Aquino, A., Ceccarelli, F.S., Eguiarte, L.E., Vázquez-Domínguez, E., de León, G.P.-P., 2014. Do the historical biogeography and evolutionary history of the digenean Margotrema spp. across central Mexico mirror those of their freshwater fish hosts (Goodeinae)? PLoS One 9, e101700.
Martinů, J., Hypša, V., Štefka, J., Martinu, J., Hypsa, V., Stefka, J., 2018. Host specificity driving genetic structure and diversity in ectoparasite populations: Coevolutionary patterns in Apodemus mice and their lice. Ecol. Evol. 8, 10008–10022. https://doi.org/10.1002/ece3.4424
Matthee, C.A., Engelbrecht, A., Matthee, S., 2018. Comparative phylogeography of parasitic Laelaps mites contribute new insights into the specialist-generalist variation hypothesis (SGVH). BMC Evol. Biol. 18, 131.
McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66, 526–538.
McDonald, B.A., Linde, C., 2002. Pathogen population genetics, evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40, 349–379.
Miura, O., Kuris, A.M., Torchin, M.E., Hechinger, R.F., Dunham, E.J., Chiba, S., 2005. Molecular-genetic analyses reveal cryptic species of trematodes in the intertidal gastropod, Batillaria cumingi (Crosse). Int. J. Parasitol. 35, 793–801.
Morand, S., Krasnov, B.R., 2010. The biogeography of host-parasite interactions. Oxford University Press.
Morand, S., Krasnov, B.R., Littlewood, D.T.J., 2015. Parasite diversity and diversification. Cambridge University Press.
Muller, H.J., 1940. An analysis of the process of structural change in chromosomes ofDrosophila. J. Genet. 40, 1.
Munday, P.L., van Herwerden, L., Dudgeon, C.L., 2004. Evidence for sympatric speciation by host shift in the sea. Curr. Biol. 14, 1498–1504.
Nachman, M.W., Payseur, B.A., 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos. Trans. R. Soc. B Biol. Sci. 367, 409–421.
Nadeau, N.J., Martin, S.H., Kozak, K.M., Salazar, C., Dasmahapatra, K.K., Davey, J.W., Baxter, S.W., Blaxter, M.L., Mallet, J., Jiggins, C.D., 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol. Ecol. 22, 814–826.
Nadeau, N.J., Whibley, A., Jones, R.T., Davey, J.W., Dasmahapatra, K.K., Baxter, S.W., Quail, M.A., Joron, M., Ffrench-Constant, R.H., Blaxter, M.L., 2012. Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. Philos. Trans. R. Soc. B Biol. Sci. 367, 343–353.
Nadler, S.A., De León, G.P.-P., 2011. Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology. Parasitology 138, 1688–1709. https://doi.org/10.1017/S003118201000168X
Nash, D.R., Als, T.D., Maile, R., Jones, G.R., Boomsma, J.J., 2008. A mosaic of chemical coevolution in a large blue butterfly. Science (80-.). 319, 88–90.
Nazarizadeh, M., Nováková, M., Drábková, M., Catchen, J., Olson, P.D., Štefka, J., 2024. Highly resolved genome assembly and comparative transcriptome profiling reveal genes

related to developmental stages of tapeworm Ligula intestinalis. Proc. R. Soc. B Biol. Sci. 291, 2007–2023. https://doi.org/10.1098/rspb.2023.2563
Nieberding, C., Libois, R., Douady, C.J., Morand, S., Michaux, J.R., 2005. Phylogeography of a nematode (Heligmosomoides polygyrus) in the western Palearctic region: persistence of northern cryptic populations during ice ages? Mol. Ecol. 14, 765–779.
Nieberding, C., Morand, S., Libois, R., Michaux, J.R., 2004. A parasite reveals cryptic phylogeographic history of its host. Proc. R. Soc. London. Ser. B Biol. Sci. 271, 2559–2568.
Nieberding, C.M., Durette-Desset, M.-C., Vanderpoorten, A., Casanova, J.C., Ribas, A., Deffontaine, V., Feliu, C., Morand, S., Libois, R., Michaux, J.R., 2008. Geography and host biogeography matter for understanding the phylogeography of a parasite. Mol. Phylogenet. Evol. 47, 538–554.
Nieberding, C.M., Morand, S., 2006. Comparative phylogeography: The use of parasites for insights into host history, in: Micromammals and Macroparasites. Springer, pp. 277–293.
Noor, M.A.F., Bennett, S.M., 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity (Edinb). 103, 439–444.
Nosil, P., 2012. Ecological Speciation, Ecological Speciation. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001
Nosil, P., Feder, J.L., 2012. Genomic divergence during speciation: causes and consequences.
Nosil, P., Feder, J.L., Flaxman, S.M., Gompert, Z., 2017. Tipping points in the dynamics of speciation. Nat. Ecol. Evol. 1, 1–8.
Olson, P.D., Littlewood, D.T.J., Griffiths, D., Kennedy, C.R., Arme, C., 2002. Evidence for the co-existence of separate strains or species of Ligula in Lough Neagh, Northern Ireland. J. Helminthol. 76, 171.
Olson, P.D., Tracey, A., Baillie, A., James, K., Doyle, S.R., Buddenborg, S.K., Rodgers, F.H., Holroyd, N., Berriman, M., 2020. Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. BMC Biol. 18, 1–16. https://doi.org/10.1186/s12915-020-00899-w
Ondrejicka, D.A., Locke, S.A., Morey, K., Borisenko, A. V, Hanner, R.H., 2014. Status and prospects of DNA barcoding in medically important parasites and vectors. Trends Parasitol. 30, 582–591.
Page, R.D.M., 2003. Tangled trees: phylogeny, cospeciation, and coevolution. University of Chicago Press.
Page, R.D.M., Charleston, M.A., 1998. Trees within trees: phylogeny and historical associations. Trends Ecol. Evol. 13, 356–359.
Payseur, B.A., Rieseberg, L.H., 2016. A genomic perspective on hybridization and speciation. Mol. Ecol. 25, 2337–2360.
Pellmyr, O., Leebens-Mack, J., Thompson, J.N., 1998. Herbivores and molecular clocks as tools in plant biogeography. Biol. J. Linn. Soc. 63, 367–378.
Pérez-Ponce de León, G., Choudhury, A., 2005. Biogeography of helminth parasites of freshwater fishes in Mexico: the search for patterns and processes. J. Biogeogr. 32, 645–659.
Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science (80-.). 344, 1410–1414.
Potter, M.F., 2011. The history of bed bug management. Am. Entomol 57, 14–25.
Potter, M.F., Romero, A., Haynes, K.F., 2008. Battling bed bugs in the USA. Proc. Sixth Int. Conf. Urban Pests 401–406.
Poulin, R., 2011. Evolutionary ecology of parasites, Evolutionary ecology of parasites.

Princeton university press.
Poulin, R., 1992. Determinants of host-specificity in parasites of freshwater fishes. Int. J. Parasitol. 22, 753–758.
Poulin, R., Keeney, D.B., 2008. Host specificity under molecular and experimental scrutiny. Trends Parasitol. 24, 24–28.
Poulin, R., Morand, S., 2000. The diversity of parasites. Q. Rev. Biol. 75, 277–293.
Poulin, R., Mouillot, D., 2003. Parasite specialization from a phylogenetic perspective: a new index of host specificity. Parasitology 126, 473.
Pouyaud, L., Desmarais, E., Deveney, M., Pariselle, A., 2006. Phylogenetic relationships among monogenean gill parasites (Dactylogyridea, Ancyrocephalidae) infesting tilapiine hosts (Cichlidae): systematic and evolutionary implications. Mol. Phylogenet. Evol. 38, 241–249.
Provan, J., Bennett, K.D., 2008. Phylogeographic insights into cryptic glacial refugia. Trends Ecol. Evol. 23, 564–571.
Raab, R.W., Moore, J.E., Vargo, E.L., Rose, L., Raab, J., Culbreth, M., Burzumato, G., Koyee, A., McCarthy, B., Raffaele, J., 2016. New introductions, spread of existing matrilines, and high rates of pyrethroid resistance result in chronic infestations of bed bugs (Cimex lectularius L.) in lower-income housing. PLoS One 11, e0117805.
Rajora, O.P., 2019. Population genomics: concepts, approaches and applications. Springer.
Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., Bowers, J.E., Burke, J.M., Rieseberg, L.H., 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Commun. 4, 1–8.
Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A.A., Farkas, T.E., Lucek, K., Hellen, E., Soria-Carrasco, V., Dennis, S.R., 2017. Transitions between phases of genomic differentiation during stick-insect speciation. Nat. Ecol. Evol. 1, 1–13.
Robison, G.A., Balvin, O., Schal, C., Vargo, E.L., Booth, W., 2015. Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the bed bug (Hemiptera: Cimicidae). J. Med. Entomol. 52, 734–738.
Romero, A., Potter, M.F., Haynes, K.F., 2010. Evaluation of chlorfenapyr for control of the bed bug, Cimex lectularius L. Pest Manag. Sci. 66, 1243–1248. https://doi.org/10.1002/ps.2002
Romero, A., Potter, M.F., Potter, D.A., Haynes, K.F., 2007. Insecticide Resistance in the Bed Bug: A Factor in the Pest's Sudden Resurgence? J. Med. Entomol. 44, 175–178. https://doi.org/10.1603/0022-2585(2007)44[175:iritbb]2.0.co;2
Ronquist, F., Sanmartín, I., 2011. Phylogenetic methods in biogeography. Annu. Rev. Ecol. Evol. Syst. 42.
Rosenfeld, J.A., Reeves, D., Brugler, M.R., Narechania, A., Simon, S., Durrett, R., Foox, J., Shianna, K., Schatz, M.C., Gandara, J., Afshinnekoo, E., Lam, E.T., Hastie, A.R., Chan, S., Cao, H., Saghbini, M., Kentsis, A., Planet, P.J., Kholodovych, V., Tessler, M., Baker, R., DeSalle, R., Sorkin, L.N., Kolokotronis, S.O., Siddall, M.E., Amato, G., Mason, C.E., 2016. Genome assembly and geospatial phylogenomics of the bed bug Cimex lectularius. Nat. Commun. 7, 1–10. https://doi.org/10.1038/ncomms10164
Rossi, L., Jennings, S., 2010. Bed bugs: a public health problem in need of a collaborative solution. J. Environ. Health 72.
Roth, S., Balvín, O., Iorio, O. Di, Siva-Jothy, M.T., Petr, B., Calva, O., Faundez, E.I., McFadzen, M., Lehnert, M.P., Khan, F.A.A., Naylor, R., Simov, N., Morrow, E., Willassen, E., Reinhardt, K., Di Iorio, O., Siva-Jothy, M.T., Benda, P., Calva, O., Faundez, E.I., McFadzen, M., Lehnert, M.P., Khan, F.A.A., 2018. A molecular phylogeny of bedbugs elucidates the evolution of host associations and sex-reversal of reproductive trait diversification. bioRxiv 367425. https://doi.org/10.1101/367425

Roth, S., Balvín, O., Siva-Jothy, M.T., Di Iorio, O., Benda, P., Calva, O., Faundez, E.I., Khan, F.A.A., McFadzen, M., Lehnert, M.P., 2019. Bedbugs evolved before their bat hosts and did not co-speciate with ancient humans. Curr. Biol. 29, 1847–1853.

Rundle, H.D., Nosil, P., 2005. Ecological speciation. Ecol. Lett. 8, 336–352.

Saenz, V.L., Booth, W., Schal, C., Vargo, E.L., 2012. Genetic analysis of bed bug populations reveals small propagule size within individual infestations but high genetic diversity across infestations from the eastern United States. J. Med. Entomol. 49, 865–875.

Safran, R.J., Scordato, E.S.C., Wilkins, M.R., Hubbard, J.K., Jenkins, B.R., Albrecht, T., Flaxman, S.M., Karaardıç, H., Vortman, Y., Lotem, A., Nosil, P., Pap, P., Shen, S., Chan, S.F., Parchman, T.L., Kane, N.C., 2016. Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. Mol. Ecol. 25, 3865–3883. https://doi.org/10.1111/mec.13740

Sapegina, V.F., 1988. Fleas of small mammals and birds in the forest-park area of the city of Novosibirsk. Parazitologiia 22, 132–136.

Schmid-Hempel, P., 2001. On the evolutionary ecology of host–parasite interactions: addressing the question with regard to bumblebees and their parasites. Naturwissenschaften 88, 147–158.

Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.-P., Bank, C., Brännström, Å., 2014. Genomics and the origin of species. Nat. Rev. Genet. 15, 176–192.

Seri Masran, S.N.A., Ab Majid, A.H., 2017. Genetic diversity and phylogenetic relationships of cytochrome C oxidase subunit I in Cimex hemipterus (Hemiptera: Cimicidae) populations in Malaysia. J. Med. Entomol. 54, 974–979.

Sorenson, M.D., Sefc, K.M., Payne, R.B., 2003. Speciation by host switch in brood parasitic indigobirds. Nature 424, 928–931.

Štefka, J., Hypša, V., Scholz, T., 2009. Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: Microsatellite study of Ligula intestinalis (Cestoda). Mol. Ecol. 18, 1187–1206. https://doi.org/10.1111/j.1365-294X.2008.04074.x

Stern, D.L., Orgogozo, V., 2008. The loci of evolution: how predictable is genetic evolution? Evol. Int. J. Org. Evol. 62, 2155–2177.

Streicker, D.G., Turmelle, A.S., Vonhof, M.J., Kuzmin, I. V, McCracken, G.F., Rupprecht, C.E., 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science (80-.). 329, 676–679.

Szalanski, A.L., Austin, J.W., McKern, J.A., Miller, D., McCoy, T., Gold, R.E., 2007. Population genetics of the bed bug Cimex lectularius (Heteroptera: Cimicidae). J. Med. Entomol.

Szalanski, A.L., Austin, J.W., McKern, J.A., Steelman, C.D., Gold, R.E., 2014. Mitochondrial and ribosomal internal transcribed spacer 1 diversity of Cimex lectularius (Hemiptera: Cimicidae). J. Med. Entomol. 45, 229–236.

Talbot, B., Balvín, O., Vonhof, M.J., Broders, H.G., Fenton, B., Keyghobadi, N., 2017. Host association and selection on salivary protein genes in bed bugs and related blood-feeding ectoparasites. R. Soc. Open Sci. 4, 170446. https://doi.org/10.1098/rsos.170446

Tawatsin, A., Thavara, U., Chompoosri, J., Phusup, Y., Jonjang, N., Khumsawads, C., Bhakdeenuan, P., Sawanpanyalert, P., Asavadachanukorn, P., Mulla, M.S., Siriyasatien, P., Debboun, M., 2011. Insecticide Resistance in Bedbugs in Thailand and Laboratory Evaluation of Insecticides for the Control of Cimex hemipterus and Cimex lectularius (Hemiptera: Cimicidae) . J. Med. Entomol. 48, 1023–1030. https://doi.org/10.1603/me11003

Thieltges, D.W., Hof, C., Dehling, D.M., Brändle, M., Brandl, R., Poulin, R., 2011. Host

diversity and latitude drive trematode diversity patterns in the European freshwater fauna. Glob. Ecol. Biogeogr. 20, 675–682.
Thoney, D.A., Hargis Jr, W.J., 1991. Monogenea (Platyhelminthes) as hazards for fish in confinement. Annu. Rev. Fish Dis. 1, 133–153.
Toews, D.P.L., Taylor, S.A., Vallender, R., Brelsford, A., Butcher, B.G., Messer, P.W., Lovette, I.J., 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr. Biol. 26, 2313–2318.
Turner, J.G., Hughes, L.F., Caspary, D.M., 2005. Affects of aging on receptive fields in rat primary auditory cortex layer V neurons. J. Neurophysiol. 94, 2738–2747.
Usinger, R.L., 1966a. Monograph of Cimicidae (Hemiptera, Heteroptera).
Usinger, R.L., 1966b. Monograph of Cimicidae (Hemiptera-Heteroptera), in: Monograph of Cimicidae (Hemiptera - Heteroptera). Entomological Society of America, p. 582. https://doi.org/10.4182/BQCN5049.1966.i
Valdés, A., Ehrlén, J., 2017. Caterpillar seed predators mediate shifts in selection on flowering phenology in their host plant. Ecology 98, 228–238.
Vaughn, C.C., Taylor, C.M., 2000. Macroecology of a host-parasite relationship. Ecography (Cop.). 23, 11–20.
Via, S., 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philos. Trans. R. Soc. B Biol. Sci. 367, 451–460.
Via, S., Conte, G., Mason-Foley, C., Mills, K., 2012. Localizing FST outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow. Mol. Ecol. 21, 5546–5560.
Vijay, R., Ravichandran, C.S., 2016. Optimal placement and sizing of distributed power sources in microgrid for power loss minimization using bat motivated optimization algorithm. Asian J. Res. Soc. Sci. Humanit. 6, 252–266.
Wasimuddin, Bryja, J., Ribas, A., Baird, S.J.E., Piálek, J., Goüy de Bellocq, J., 2016. Testing parasite 'intimacy': the whipworm T richuris muris in the E uropean house mouse hybrid zone. Ecol. Evol. 6, 2688–2701. https://doi.org/10.1002/ece3.2022
Wawrocka, K., Balvín, O., Bartonička, T., 2015. Reproduction barrier between two lineages of bed bug (Cimex lectularius)(Heteroptera: Cimicidae). Parasitol. Res. 114, 3019–3025.
Wickström, L.M., Haukisalmi, V., Varis, S., Hantula, J., Fedorov, V.B., Henttonen, H., 2003. Phylogeography of the circumpolar Paranoplocephala arctica species complex (Cestoda: Anoplocephalidae) parasitizing collared lemmings (Dicrostonyx spp.). Mol. Ecol. 12, 3359–3371.
Wolf, J.B.W., Ellegren, H., 2017. Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. 18, 87.
Wood, C.L., Johnson, P.T.J., 2016. How Does Space Influence the Relationship Between Host and Parasite Diversity? J. Parasitol. 102, 485–494. https://doi.org/10.1645/15-920
Wu, C., 2001. The genic view of the process of speciation. J. Evol. Biol. 14, 851–865.
Yeaman, S., Hodgins, K.A., Lotterhos, K.E., Suren, H., Nadeau, S., Degner, J.C., Nurkowski, K.A., Smets, P., Wang, T., Gray, L.K., 2016. Convergent local adaptation to climate in distantly related conifers. Science (80-.). 353, 1431–1433.
Yunzhi, Z., Zhengda, G., Xiguang, F., Xingde, D., Houyong, W., Xue, W., Yuan, L., 2002. Study on the relationship between fleas and hosts in Mt. Baicaoling, Yunnan Province, China. Di Fang Bing Tong bao= Endem. Dis. Bull. 17, 22–24.
Zardoya, R., Meyer, a, 1996. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13, 933–942. https://doi.org/10.1093/oxfordjournals.molbev.a025661
Zhu, F., Wigginton, J., Romero, A., Moore, A., Ferguson, K., Palli, R., Potter, M.F., Haynes, K.F., Palli, S.R., 2010. Widespread distribution of knockdown resistance mutations in the

bed bug, Cimex lectularius (Hemiptera: Cimicidae), populations in the United States. Arch. Insect Biochem. Physiol. 73, 245–257. https://doi.org/10.1002/arch.20355
Zietara, M.S., Lumme, J., 2002. Speciation by host switch and adaptive radiation in a fish parasite genus Gyrodactylus (Monogenea, Gyrodactylidae). Evolution (N. Y). 56, 2445–2458.

Chapter I

Different hosts in different lakes: prevalence and population genetic structure of *Ligula intestinalis* plerocercoids in Czech water bodies

Folia Parasitologica
DOI: 10.14411/fp.2022.018

Institute of Parasitology, Biology Centre CAS
Folia Parasitologica 2022, 69: 018
doi: 10.14411/fp.2022.018

Folia Parasitologica
http://folia.paru.cas.cz

Research Article OPEN ACCESS

Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies

Masoud Nazarizadeh[1,2], **Jiří Peterka**[3], **Jan Kubečka**[3], **Mojmír Vašek**[3], **Tomáš Jůza**[3], **Karlos Ribeiro de Moraes**[3], **Martin Čech**[3], **Michaela Holubová**[3], **Allan T. Souza**[3], **Petr Blabolil**[2,3], **Milan Muška**[3], **Lobsang Tsering**[3], **Daniel Bartoň**[3], **Milan Říha**[3], **Marek Šmejkal**[3], **Michal Tušer**[3], **Lukáš Vejřík**[3], **Jaroslava Frouzová**[3], **Ivan Jarić**[3], **Marie Prchalová**[3], **Ivana Vejříková**[3] and **Jan Štefka**[1,2*]

[1] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic;

[2] Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic;

[3] Institute of Hydrobiology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

Abstract: *Ligula intestinalis* (Linnaeus, 1758) is a tapeworm parasite with a worldwide distribution that uses a wide variety of fish species as its second intermediate host. In the present study, we investigated the prevalence and population genetic structure of plerocercoids of *L. intestinalis* in five common cyprinoid species, roach *Rutilus rutilus* (Linnaeus), freshwater bream *Abramis brama* (Linnaeus), white bream *Blicca bjoerkna* (Linnaeus), bleak *Alburnus alburnus* (Linnaeus), and rudd *Scardinius erythrophthalmus* (Linnaeus), collected in six water bodies of the Czech Republic (Milada, Most, Medard, Jordán, Římov and Lipno). Of the six study sites, the highest frequency of parasitism was recorded in Lake Medard (15%). The overall prevalence rate among the species was as follows: roach > rudd ≥ freshwater bream > bleak > white bream. Two mitochondrial genes (*cyt*b and *COI*) were used to compare the population genetic structure of parasite populations using selected samples from the five fish species. The results of the phylogenetic analysis indicated that all populations of *L. intestinalis* were placed in Clade A, previously identified as the most common in Europe. At a finer scale, haplotype network and PCoA analyses indicated the possible emergence of host specificity of several mtDNA haplotypes to the freshwater bream. Moreover, pairwise Fixation indices (F_{ST}) revealed a significant genetic structure between the parasite population in freshwater bream and other host species. Parasite populations in roach not only showed the highest rate of prevalence but also depicted a maximum number of shared haplotypes with populations from bleak and rudd. Our results suggest that recent ecological differentiation might have influenced tapeworm populations at a fine evolutionary scale. Thus, the differences in prevalence between fish host species in different lakes might be influenced not only by the parasite's ecology, but also by its genetic diversity.

Keywords: tapeworm, Czech Republic, host specificity, freshwater, fish parasite

This article contains supporting information (Tables S1–S3, Figs. 1–4) online at http://folia.paru.cas.cz/suppl/2022-69-018.pdf

Host specificity is a crucial factor shaping the distribution and diversity patterns of parasite species. Understanding host specificity is essential in order to characterise the diversity and study the biogeography of parasites (Štefka et al. 2009, Martinů et al. 2018, Bernard et al. 2019, Wells and Clark 2019). Given that the global species richness of parasites is higher than the diversity of non-parasitic species (Windsor 1998) host specificity is of utmost importance for the studies of co-evolution (Blatrix and Herbers 2003), but also co-extinction, i.e. the extinction of host species resulting in the extinction of their associated parasite species (Stork and Lyal 1993). Moreover, host specificity can act as an isolation barrier restricting gene flow among parasite populations (Schirrmann and Leuchtmann 2015). Accordingly, the emergence of host specificity can be associated with the earliest stages of speciation in parasites (Meinilä et al. 2004).

Host specificity may be either associated with the ability of a parasite species to attach and adapt to a given host species or with the limitations of a parasite species in finding a host species (i.e., adaptive *vs* non-adaptive factors) (Nosil 2015). Consequently, estimation of the source of adaptive evolutionary change in parasites requires population genetic studies of host specificity (Huyse et al. 2005). On the contrary, phylogenetic analyses can uncover morphologically indistinguishable but genetically distinct species, i.e., cryptic species, differing in their biology

Address for correspondence: Jan Štefka, Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, České Budějovice; E-mail: jan.stefka@gmail.com

Table 1. Sampling locations in the Czech Republic

Reservoirs	Area	Maximum depth	Locality
Lipno	48.70 km^2	25 m	48.7000N, 14.0666E
Římov	2.10 km^2	43 m	48.8341N, 14.4814E
Jordán	0.51 km^2	12.5 m	49.4213N, 14.6650E
Most	3.11 km^2	71 m	50.5396N, 13.6460E
Medard	4.93 km^2	50 m	50.1794N, 12.5958E
Milada	2.50 km^2	22 m	50.6536N, 13.9444E

(Nazarizadeh et al. 2022). Complexes of cryptic species or lineages have been revealed multiple times in helminths (e.g., Blouin 2002, Bouzid et al. 2008a, Ristau et al. 2013). Lastly, studying patterns of population genetic structures in parasitic helminths can shed light on the phylogeographic history of parasites and their host species, and more importantly, identify the ecological drivers of speciation patterns in parasites (Cole and Viney 2019).

The diphyllobothriidean cestode *Ligula intestinalis* (Linnaeus, 1758) and its related (cryptic) species have become a favourite object of phylogenetic and population genetic studies (e.g., Bouzid et al. 2008b, Štefka et al. 2009) due to their wide host spectrum and a geographical distribution covering the Holarctic (Dubinina 1980), Paleotropical (Britton et al. 2009) and Australasian realms (Morgan 2003, Chapman et al. 2006, Lagrue et al. 2018). *Ligula intestinalis* is an endoparasite demonstrating a life cycle with two intermediate hosts: a diaptomid or cyclopoid copepod as the first host and a planktivorous fish as the second intermediate host. Cyprinoid fishes in natural and artificial reservoirs are frequently exposed to infection with *L. intestinalis* (Arme and Owen 1968, Kennedy 1974). In addition to its common cyprinoid fish hosts, *L. intestinalis* and other species of the genus also infect other fish taxa, including species of the families Salmonidae, Galaxiidae and Catostomidae (Dubinina 1980, Bean and Winfield 1991, Baruš and Prokeš 1995, Groves and Shields 2001). This tapeworm spends a maximum of five days in a fish-eating bird as the final host (Dubinina 1980), during which it reaches sexual maturity and begins to produce eggs that are released into the water.

The fish stage (plerocercoid) plays a significant role in the behaviour, fecundity and health of the fish which may consequently lead to considerable aquaculture damage (Sweeting 1977, Wyatt and Kennedy 1989, Carter et al. 2005). Moreover, the parasite can significantly affect fish growth, stress resistance and susceptibility to predators (Longshaw et al. 2010, Iwanowicz 2011). Mortality of fish infected with *L. intestinalis* may occur either directly from their inability to survive through winter (Wyatt and Kennedy 1989) or indirectly through increased predation risk by birds and by other fish (Van Dobben 1952, Holmes and Bethel 1972, Sweeting 1977, Palm et al. 2018). Therefore, the prevalence of *L. intestinalis* in its fish hosts has been monitored by multiple studies (Dence 1958, Ergonul and Altindag 2005, Ivankov et al. 2020). Furthermore, variation in the rate of prevalence between different host species can indicate the emergence of host specificity and thus it is an important tool in ecological parasitology (Britton et al. 2009).

Most population genetic studies on *L. intestinalis* have concentrated on analysing evolutionary lineages at a large scale (global or continental). For instance, using mtDNA sequences Bouzid et al. (2008b) revealed the existence of two evolutionary lineages of *Ligula* (Clades A and B) in Europe and the Mediterranean, which was later confirmed by a microsatellite study (Štefka et al. 2009). Clade A samples were specific to fishes of the Leuciscide subfamilies Leuciscinae and Alburninae, whereas the Clade B was specific to other, less abundant, cyprinoid species, such as gudgeon *Gobio gobio* (Linnaeus), bitterlings (*Rhodeus* sp.) and barbels (*Barbus* sp.). Despite their genetic differentiation, both lineages shared the definitive hosts, great crested grebe *Podiceps cristatus* (Linnaeus) and goosander *Mergus merganser* (Linnaeus). Therefore, differences in fish-host specificity were proved for the Euro-Mediterranean clades.

These results confirmed findings of an earlier population genetic study performed at a local scale in Lough Neagh (Northern Ireland) (Olson et al. 2002), in which *Ligula* parasites from roach *Rutilus rutilus* (Linnaeus) and gudgeon were found to represent distinct strains, considering their effect on the gonadal development of their hosts. Furthermore, *Ligula* from minnow *Phoxinus phoxinus* (Linnaeus) from Wales was found genetically similar to *Ligula* from roach in Lough Neagh. Olson et al. (2002) proposed that the introduction of roach and a subsequent increase in final host populations (*P. cristatus*) was likely the cause for the presence of separate strains in Lough Neagh.

Here, we aimed to monitor the occurrence and genetic structure of plerocercoids of *L. intestinalis* among the five most abundant cyprinoid species in the Czech Republic. We attempted to (a) compare the prevalence of plerocercoids of *L. intestinalis* in six freshwater lakes and reservoirs in the Czech Republic, and (b) compare the phylogenetic relationship of *Ligula* populations from the five host species to specimens analysed in previous genetic studies. Using a subsample of the collected specimens, we also aimed to (c) test for parasite genetic structure associated with host specificity at a smaller spatial scale than in the majority of previous studies.

MATERIALS AND METHODS

Description of the study area

There are almost 24,000 fishponds and reservoirs on the territory of the Czech Republic, with a total area of 52,000 hectares (Ministry of Agriculture of the Czech Republic 2019). In the present study, we focused on six artificial lakes and reservoirs covering the western half of the Czech Republic (Bohemia), including Milada and Most lakes in the north, Medard lake in the west, and the Lipno, Římov and Jordán reservoirs in the south (Fig. 1, Table 1).

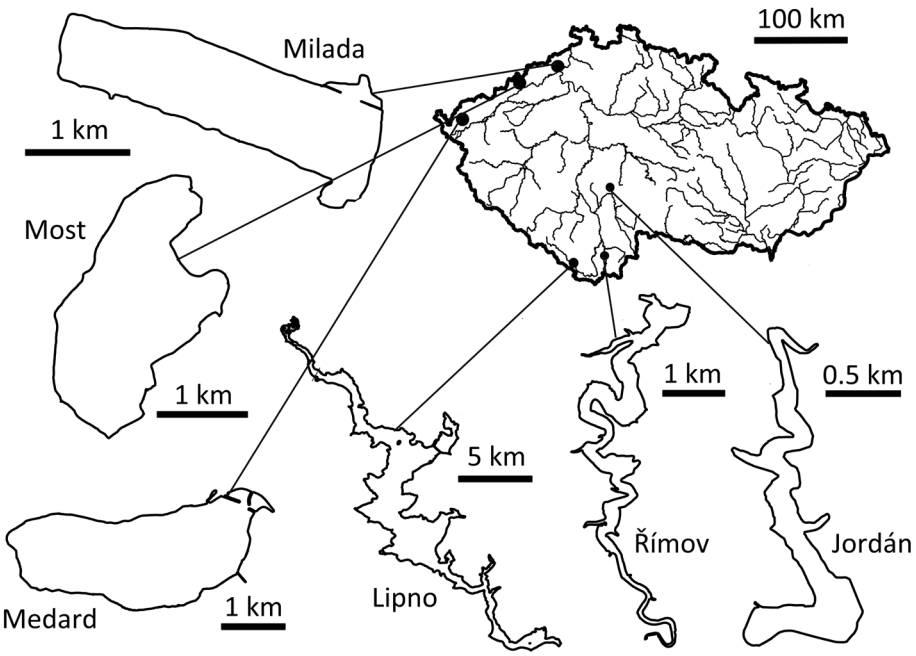


Fig. 1. Map of the six lakes and reservoirs studied.

Sampling

Permits for sampling were issued by the state enterprise Povodí Vltavy for the Lipno and Římov reservoirs, by the municipality Tábor for the Jordán Reservoir, by the state enterprise Palivový kombinát Ústí for the Milada and Most lakes, and by the joint company Sokolovská uhelná for the Medard Lake. Between July and September 2020, a total of 6,630 fish belonging to five species of the family Leuciscidae were collected: freshwater bream, *Blicca bjoerkna* (Linnaeus), white bream, *Abramis brama* (Linnaeus), bleak, *Alburnus alburnus* (Linnaeus), roach, *Rutilus rutilus* (Linnaeus), and rudd, *Scardinius erythrophthalmus* Linnaeus. They were captured using benthic and pelagic CEN multi-mesh gillnets with stretched mesh sizes ranging from 5 to 55 mm (knot-to-knot), supplemented by benthic and pelagic large-mesh gillnets with stretched mesh sizes ranging from 70 to 130 mm (knot-to-knot) (Blabolil et al. 2017). To capture the fish, the gillnets were set overnight. The fish catch was standardised as the number of fish per area of installed gillnets (individuals per 1,000 m^2).

After extraction from the nets, the fish individuals were immediately killed and their body cavities were examined for the presence of plerocercoids of *Ligula intestinalis*. In particular, the anterior part of the body has a rounded shape, and no external segmentation of the strobila is apparent. The genital complexes extend in an irregular longitudinal row in the strobila of plerocercoids and the ovaries, testes, genital ducts and vitelline follicles are located in the transverse segments (Bykhovskaya-Pavlovskaya 1964). The plerocercoids were fixed in 96% ethanol and stored at freezing temperatures.

Estimating the rate of prevalence

The prevalence was calculated as the percentage of infected fish individuals. The significance of the differences in prevalence in different host species and reservoirs was tested with the non-parametric Kruskal-Wallis H test and pairwise Wilcox test (Zar 1999). In addition, fish abundance was estimated for the five host species based on catch from the multi-mesh gillnet method (Appelberg et al. 1995). Then, a linear regression test was carried out between the fish abundances (average between pelagic and benthic standardised abundance) and the prevalence of infection to test any possible connection between the abundance and rate of infection. All statistical analyses were performed using the stats package in R version 4.0.1 software (R Core Team 2021).

DNA isolation, amplification and PCR

In total, 30 *Ligula* specimens covering the five host species from Lipno, Medard, Římov and Most were selected to verify their assumed mtDNA clade membership (Clade A; Bouzid et al. 2008b) and to reveal possible internal population structure. DNA extraction was performed using the DNeasy Blood and Tissue kit (Qiagen). Then, DNA was eluted into 60 µl of the AE buffer. Two fragments of the mtDNA containing 914 bp of the mitochondrial cytochrome b gene (*Cyt*b) and 396 bp of the mitochondrial cytochrome oxidase subunit I gene (*COI*) loci were amplified using four pairs of primers (F2Dnihcob: 5′-GTT TTA CTG ATA GGT TAT TTA AAC-3′, R2Dnihcob_mod: 5′- CAG TTT AAA AAT CGA GTT AAA GAT-3′) (Wicht et al. 2010) and COIA2: 5′-CAT ATG TTT TGA TTT TTT GG-3′ and COIB2: 5′- AKA ACA TAA TGA AAA TGA GC-3′ (Li et al. 2000, Bouzid et al. 2008b, Štefka et al. 2009). PCRs were carried out in a 12 µl volume using 1 µl of extracted DNA, 10 pM of each primer, 6.25 µl 2× concentrated PPP Master Mix (Top-Bio, Vestec, Czech Republic). The amplification protocol consisted of one denaturation step at 94 °C for 15 min, then 30 cycles of denaturation at 94 °C for 30 seconds, annealing at 50 °C for 45 seconds and an extension step at 72 °C for 45 seconds, followed by the last elongation step at 72 °C for 10 min. PCR products were enzymatically cleaned up with VWR ExoCleanUp FAST PCR reagent (VWR, USA) following manufacturer's protocol. Purified PCR products were sequenced using the PCR primers in a commercial laboratory (SeqMe, Dobříš, Czech Republic).

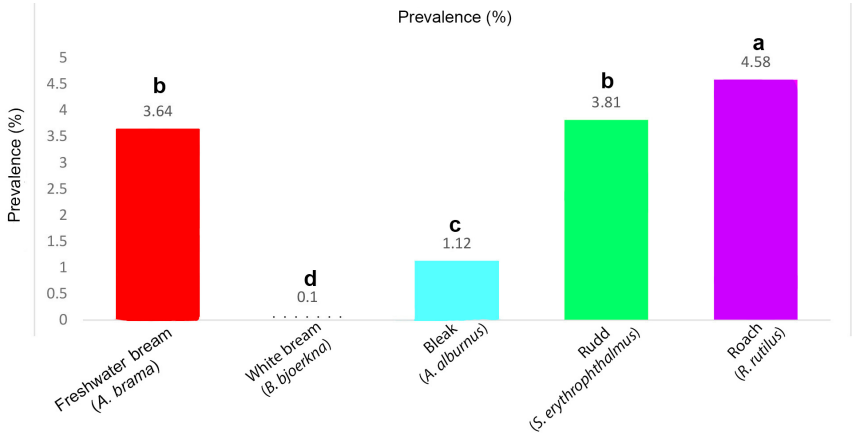


Fig. 2. Prevalence of plerocercoids of *Ligula intestinalis* (Linnaeus, 1758) in five cyprinid species across all localities. Shared letters denote non-significant results, while different letters denote significant differences (Kruskal-Wallis test, $p \leq 0.05$), all localities (values above columns).

Sequence analysis

The Geneious Prime software (https://www.geneious.com; version 20.1.1) was applied to edit the raw sequences. A multiple sequence alignment was generated using ClustalW in MEGA version 5 (Tamura et al. 2011). Then, all sequences were translated into protein-coding sequences based on the invertebrate mitochondrial genetic code to examine any possible stop codons. A substitution saturation test was carried out with the plotting of transversions and transitions against genetic distance and estimating the index of substitution saturation (Iss) and its critical value (Iss.c) in DAMBE Version 6.0.4 (Xia and Xie 2001, Xia and Lemey 2009). DnaSP version 6.0 (Librado and Rozas 2009) was used to estimate nucleotide diversity, haplotype diversity and sequence polymorphisms. Genetic distances among the parasite populations were determined using the uncorrected pairwise genetic distances with 1,000 bootstraps.

Phylogenetic relationships

We generated a concatenated dataset of two fragments of mtDNA (*Cyt*b and *COI*) comprising 1,310 bp in length for the 30 selected samples of *L. intestinalis* and aligned it with 82 sequences from two previous studies (Bouzid et al. 2008b, 2013; Table S1). Bayesian Inference (BI) and Maximum Likelihood (ML) approaches were carried out to reconstruct the phylogenetic relationships of the newly sampled plerocercoids. Using PartitionFinder version 2.1.1 (Lanfear et al. 2017), we identified the appropriate substitution nucleotide model for genes and codons based on the Bayesian Information Criterion (BIC). Bayesian phylogenetic analyses were performed via MrBayes version 3.2.2 (Ronquist and Huelsenbeck 2003) using the selected model of sequence evolution (Table S2). We ran one cold and three heated chains for 40 million generations, with trees being sampled every 1,000 generations. We discarded the first 25% of the runs as burn-in. Then, by combining the post-burn-in trees, a 50 percent majority rule consensus tree was constructed. Bayesian posterior probabilities (PP) were computed to estimate support of the Bayesian tree. Using the best-fitting model, a maximum likelihood analysis was applied to construct an ML tree in IQTREE version 2.1.2, (Minh et al. 2020). Branch support of the ML tree was computed using 1,000 rapid bootstrap replicates (Hoang et al. 2018). *Dibothriocephalus latus* (Linnaeus, 1758), *D. nihonkaiensis* (Yamane, Kamo, Bylund et Wikgren, 1986) and *Diphyllobothrium stemmacephalum* Cobbold, 1858 (GenBank accession number AB269325, AB268585 and AP017648, respectively) were used as outgroups based on data provided by Waeschenbach et al. (2017).

Population genetic structure

Population genetic relationships among populations of *L. intestinalis* in the Czech water bodies were reconstructed using the concatenated mtDNA dataset by a maximum parsimony algorithm in PopArt (Leigh and Bryant 2015) and by the Principal Coordinates Analysis (PCoA) in adegenet R package (Jombart 2008) in R 4.0.1 software. Moreover, to compare the newly studied populations with the Clade A haplotypes from the previous study (Bouzid et al. 2008b, Štefka et al. 2009) a haplotype network containing a total of 80 sequences (801 bp in length) was constructed. Pairwise F-statistics for the newly sampled Czech populations were executed with 10,000 permutations in Arlequin version 3.5 and parasite populations were categorised into five groups based on their hosts.

RESULTS

Fish abundance and rate of infection

Roach was the most abundant fish species in Milada (208 individuals/1,000 m^2), Most (90 individuals/1,000 m^2) and Jordán (310 individuals/1,000 m^2) water bodies. Bleak was numerically the most dominant species in Lipno (236 individuals/1,000 m^2) and Římov (148 individuals/1,000 m^2) reservoirs. In Lake Medard, rudd had the highest fish abundance (26 individuals/1,000 m^2; Table S3). Analysis of prevalence for five cyprinoid species, calculated cumulatively for all six water bodies, is presented in Fig. 2. Prevalence differed significantly among all species pairs ($p \leq 0.05$), except for freshwater bream and rudd. For example, roach (4.6%) and white bream (0.1%) had the highest and lowest significant prevalence, respectively. The prevalence for rudd (3.8%) was significantly higher than the prevalence for bleak (1.1%) or white bream (0.1%).

Table 2. Prevalence of infection of five cyprinoid fishes with *Ligula intestinalis* (Linnaeus, 1758) in six reservoirs (N = total number samples, n = total number of infected fish, P = prevalence in %). Numbers followed by a letter in column P mark the level of significance in the prevalence rate of roach between pairs of different reservoirs in a Kruskal-Wallis test. Shared letters [a–c] denote non-significant results, while different letters denote significant differences ($p \leq 0.05$).

	Medard			Jordán			Milada			Most			Lipno			Římov		
	N	n	P	N	n	P	N	n	P	N	n	P	N	n	P	N	n	P
Freshwater bream	6	0	**0**	39	0	**0**	0	0	**0**	0	0	**0**	60	4	**6.66**	169	6	**3.55**
White bream	8	0	**0**	1617	2	**0.12**	0	0	**0**	0	0	**0**	236	0	**0**	45	0	**0**
Bleak	18	0	**0**	224	0	**0**	0	0	**0**	0	0	**0**	1001	4	**0.39**	589	18	**3.05**
Rudd	37	2	**5.4**	64	0	**0**	33	0	**0**	128	8	**6.25**	0	0	**0**	0	0	**0**
Roach	60	17	**28.3** [a]	176	0	**0**	412	22	**5.39** [b]	483	52	**10.7** [c]	636	0	**0**	434	10	**2.30** [b]
Total	129	19	**14.72**	2120	2	**0.094**	455	22	**4.49**	611	60	**9.81**	1933	8	**0.41**	1237	34	**2.74**

Table 3. Genetic characteristics of the concatenated dataset (1310 bp) estimated for samples of *Ligula intestinalis* (Linnaeus, 1758) from the five host species (samples comprising more than three individuals) including number of individuals (N), nucleotide diversity (Pi), number of haplotypes (h), haplotype diversity (Hd), (K) the total number of mutations.

Parasite populations	N	h	hd	pi	k
Freshwater bream	5	5	1.00	0.0021	7
White bream	2	2	-	-	-
Bleak	4	1	0.00	0.0000	0
Rudd	4	4	1.00	0.0016	4
Roach	15	11	0.93	0.0035	25

The prevalence recorded for each species in each freshwater reservoir is given in Table 2. According to our observations, roach was either the most abundant or the second most abundant cyprinoid species in lakes Medard, Milada and Most (gillnet catches; Table S3), where it also had the highest prevalence (28.3%, 5.3% and 10.7%, respectively). In contrast, freshwater bream in most lakes and reservoirs (except Římov) had the lowest abundance of all species studied. However, in two reservoirs (Lipno and Římov) the prevalence of *Ligula intestinalis* was higher in freshwater bream than in other host species (Table 2, Fig. S1). Overall, Medard and Jordán demonstrated the highest and lowest prevalence of *L. intestinalis*, respectively. Roach was the only species sampled frequently enough to allow comparison of prevalence among reservoirs. Its prevalence rate was significantly higher in the Medard than in Římov, Milada and Most ($p \leq 0.05$; Table 2). No correlation was found between fish abundance and prevalence of infection with *L. intestinalis* ($p > 0.05$; Fig. S2).

Sequences analyses

Our concatenated dataset (*Cyt*b and *COI*) containing 30 sequences comprised 21 haplotypes, showing 1,274 invariable sites, 22 parsimony informative sites, and 14 singleton sites, out of the 1,310 bp aligned positions. Neither stop codons nor insertions/deletions were observed in the datasets. Table 3 demonstrates genetic characteristics for the parasite populations comprising more than three sampled individuals. Minimum and maximum nucleotide diversities were calculated to be 0.000 and 0.00348 (the samples from roach and bleak, respectively). Also, haplotype diversities ranged between 0.0 and 1 in bleak and rudd, and freshwater bream, respectively (Table 3).

Phylogenetic relationships

The substitution saturation analysis showed that the Iss was significantly smaller than Iss.c ($p < 0.005$), suggesting the suitability of the concatenated dataset for phylogenetic analyses. The pattern of transitions and transversions plotted against the genetic distance also confirmed that the dataset retains sufficient phylogenetic signals (Fig. S3).

Both phylogenetic trees (ML and BI) recovered identical topologies. Consequently, only the BI tree is demonstrated (Fig. 3). The phylogenetic analysis disclosed six distinct lineages within the monophyletic *L. intestinalis* complex which was consistent with the previous studies (Bouzid et al. 2008b, Štefka et al. 2009). Our phylogenetic results depicted that all samples of plerocercoids from the five host species in the Czech water bodies were placed within Clade A as expected.

Population genetic structure

The pattern of population structure revealed by haplotype network analysis showed haplotype sharing among the parasite samples in rudd, bleak and roach. In contrast, samples from the freshwater bream were separated from the other haplotypes by three mutational steps (Fig. 4). Similarly, the haplotype network also containing samples from the previous studies, based on the 405 bp of *Cyt*b and 396 bp *COI* (801 bp), revealed no shared haplotypes between *L. intestinalis* samples from the freshwater bream and other host species whereas common haplotypes were discovered between rudd and roach hosts (Fig S4). However, despite the different status of the bream related haplotypes, they did not create a single cluster in this larger sample containing other European populations. The PCoA analysis of the current samples revealed a similar clustering pattern, including the assignment of three parasite populations (rudd, bleak and roach) to one cluster and clear separation of the freshwater bream population from the rest. The first and the second axes in the PCoA explained 43.9% and 16.3% of the total variations, respectively (Fig. 5).

The maximum and minimum values of pairwise Fixation indices (F_{ST}) were estimated to be 0.71 and 0.02, respectively (Table 4). In accordance with the haplotype network and PCoA results, the parasites from freshwater bream demonstrated a significant ($P < 0.05$) genetic differentiation in comparison to other samples. Similarly, the highest and the lowest genetic distances (0.06% and 0.01%) were calculated between the freshwater bream and the roach, bleak and rudd, respectively.

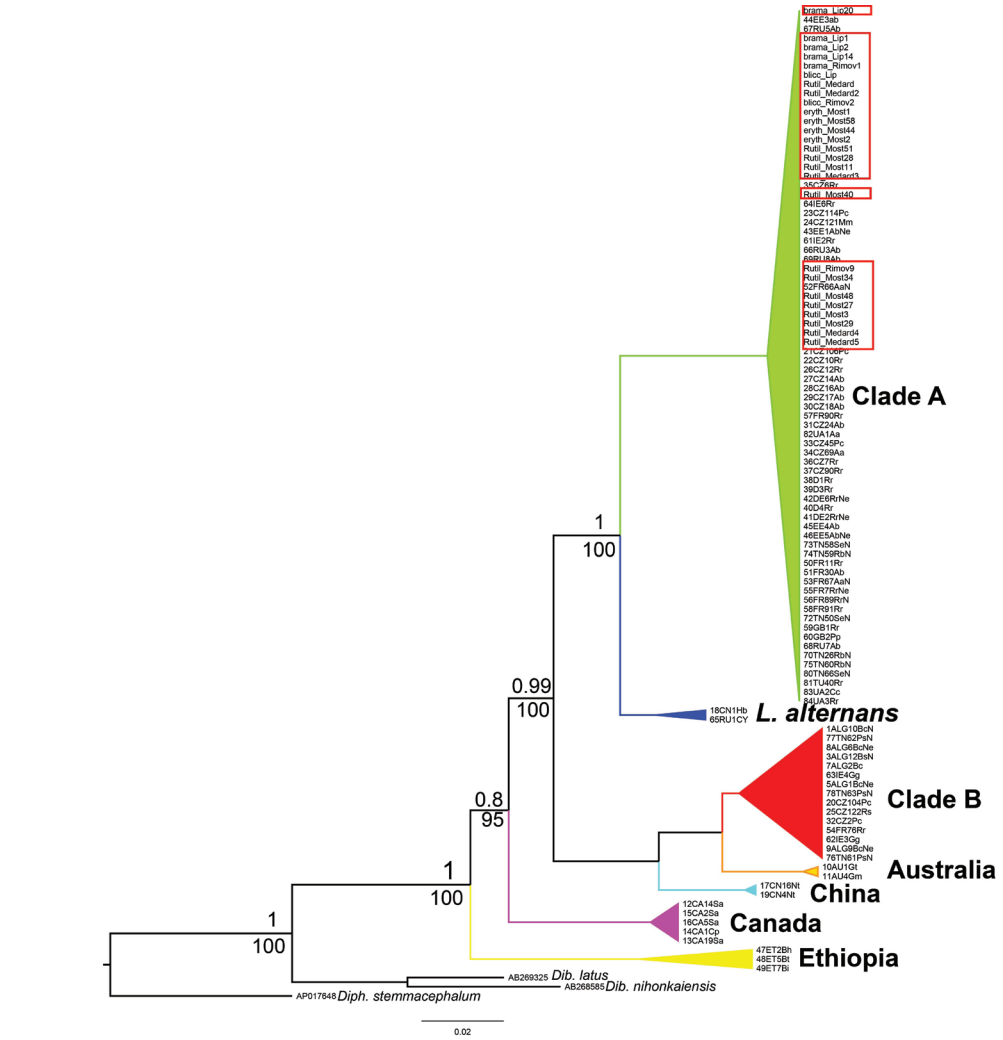


Fig. 3. Bayesian phylogenetic tree reconstructed for the *Ligula intestinalis* (Linnaeus, 1758) species complex using a concatenated mtDNA dataset (*Cyt*b and *COI*, 1,310 bp). For each node, nodal supports indicate Bayesian Inferrence (BI, top) and Maximum Likelihood (ML, base) support values. Newly obtained sequences from the Czech Republic are highlighted by red rectangles.

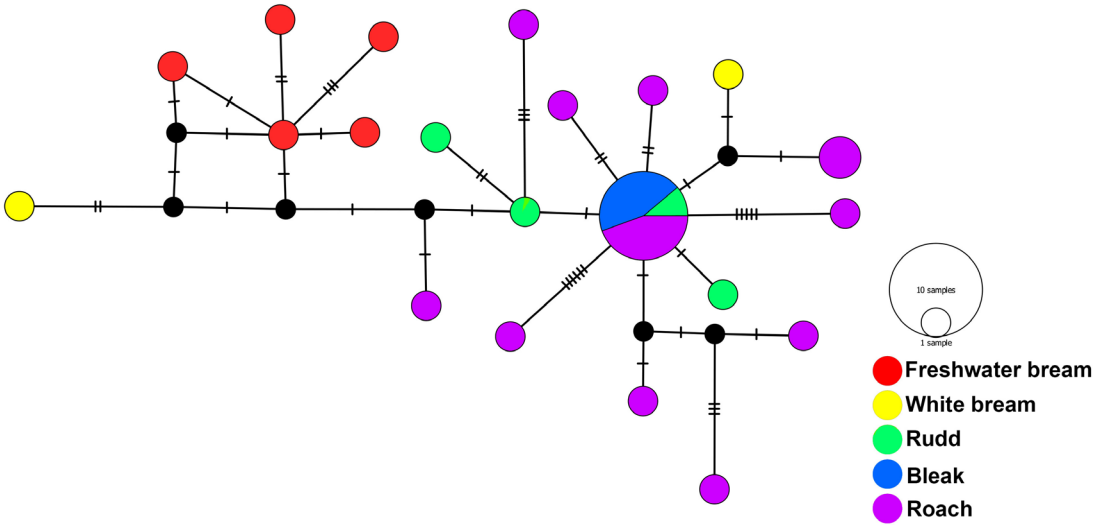


Fig. 4. Haplotype network of 30 samples of plerocercoids of *Ligula intestinalis* (Linnaeus, 1758) selected to represent hosts from the Czech Republic. The network was generated using a concatenated mtDNA dataset (1,310 bp). Individual haplotypes were coloured based on their respective fish hosts (using the same colours as in Fig. 1). The size of the circles is proportional to haplotype frequencies. Putative unsampled haplotypes are shown by black circles and dash symbols indicate mutational steps.

Table 4. Pairwise values of Fst (below diagonal) and genetic distance (above diagonal) for samples of *Ligula intestinalis* (Linnaeus, 1758) from the five host species. P-values lower than 0.05 are flagged with an asterisk (*).

	N	Freshwater bream	White bream	Bleak	Rudd	Roach
Freshwater bream	5	0	0.05%	0.04%	0.04%	0.06%
White bream	2	0.31*	0	0.04%	0.05%	0.03%
Bleak	4	0.55*	0.14	0	0.01%	0.02%
Rudd	6	0.45*	0.10	0.02	0	0.02%
Roach	15	0.71*	0.38	0.13	0.10	0

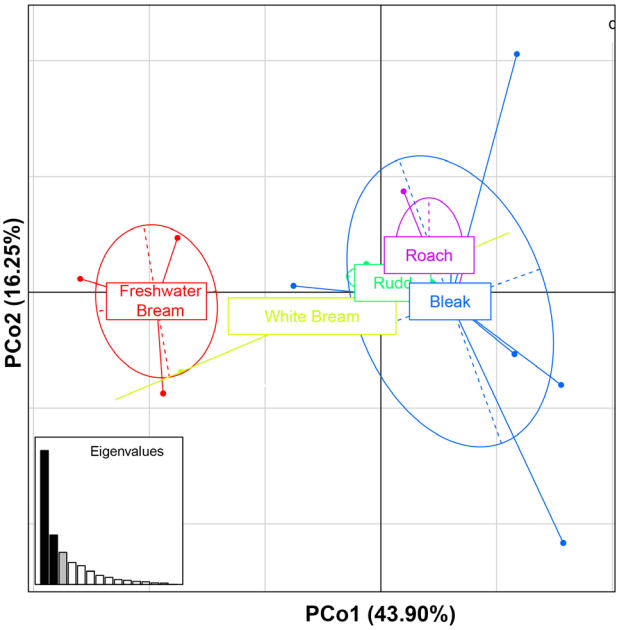


Fig. 5. Principle coordinate analysis (PCoA) based on Euclidean genetic distance for the parasite samples from the five host species, freshwater bream *Abramis brama* (Linnaeus), white bream *Blicca bjoerkna* (Linnaeus), bleak *Alburnus alburnus* (Linnaeus), rudd *Scardinius erythrophthalmus* (Linnaeus) and roach *Rutilus rutilus* (Linnaeus).

DISCUSSION

In the present study, a total number of 6,630 fish of five cyprinoid species from three lakes and three reservoirs (Medard, Milada, Most, Jordán, Římov and Lipno) in the Czech Republic were captured and examined for the presence of plerocercoids of *Ligula intestinalis*. While the presence of *L. intestinalis* in Lipno has already been demonstrated in a population genetic study (Štefka et al. 2009), the presence of the parasite in Medard, Milada, Most, Jordán and Římov has not been previously reported. Each of the host species studied (freshwater bream, white bream, bleak, rudd and roach) was infected in at least one of the lakes. Then, we used population genetic analysis on a selected set of samples to analyse the level of population structure and test for possible occurrence of host specificity in the plerocercoids. Obtained results suggest an emergence of a bream-specific genetic cluster. Below we discuss the genetic results with regard to the obtained rates of prevalence in the Czech water bodies.

Prevalence of *Ligula intestinalis* in different fish and lakes

We attempted to monitor the infection rate of *L. intestinalis* within one summer season (July and September 2020). The results showed that the total prevalence (across all fish species) is less than 15% in all the reservoirs studied in the Czech Republic, whereas roach in Medard had the highest prevalence (28%) when host species were considered separately. Earlier studies show a large variation in prevalence both in the Czech Republic and elsewhere. In a study conducted in the Nové Mlýny reservoirs (South Moravia, Czech Republic), the infection rate of *L. intestinalis* in white bream ranged from 1.5% in April-May to 11.5% in November-December (Baruš and Prokeš 1995). In contrast, in a study of *L. intestinalis* in freshwater bream from the Aras Dam (Iran) the prevalence was 45%, with a significantly higher infection rate in autumn compared to winter (Nezafat et al. 2008). In France, the highest value of prevalence was observed at the end of the summer and in the autumn (Brown et al. 2001). In Zimbabwe, the highest seasonal infection rate (prevalence 21–46%) was reported in the cooler months of July to September (Barson and Marshall 2003). In addition to the season and temperature, which are generally considered to be major factors affecting the speed of development of this parasite, *L. intestinalis* may exhibit also epizootic cycles over multi-seasonal periods, during which it rapidly increases its numbers in the lake resulting in increased fish mortality (Kennedy et al. 2001).

Our single season data did not pick up excessively high prevalence rates suggesting an epizootic peak, although the rate for roach in Medard was relatively high with possible ecological impact. The pronounced infection rate in roach compared to the other four species appears to be consistent with previous reports from the United Kingdom, where infection rates were higher in roach than in gudgeon, rudd, freshwater bream, minnow and dace *Leuciscus leuciscus* (Linnaeus) (Arme and Owen 1968, Kennedy and Burrough 1981). Further research has shown that although *L. intestinalis* prevalence fluctuates in irregular cycles, epizootics of the parasite are associated with rapid increases in roach populations (Kennedy et al. 2001).

However, our research showed that the component of a hidden population structure needs to be considered when interpreting the ecological impact of prevalence rates. For example, roach in Lipno was free of *L. intestinalis* despite the large sample size, while the prevalence in freshwater bream was considerable (6.7%). At the first sight, this result could imply that different phases of the *L. intestinalis* infection cycle are present in the lake, switching between different hosts. However, even though our sequencing analysis confirmed that all Czech samples belong to the same mtDNA Clade A (as in Bouzid et al. 2008b, Bouzid et al. 2013) we also found emerging genetic differentiation between freshwater bream plerocercoids and the rest of population of *L. intestinalis* from the samples. Thus, the ecology of roach and bream is probably highly independent of each other.

For *L. intestinalis* populations to thrive, they must succeed in all three stages of their life cycle: a copepod as the first host, a zooplanktivorous fish as the second intermediate host, and a bird as the final host, in which they reach sexual maturity. The adult worm spends only a maximum of five days in its final host (Kennedy et al. 2001), but the migratory birds play an important role in the life cycle of *L. intestinalis*, as they are the main means of dispersal (Štefka et al. 2009). Among birds in the Czech Republic, the presence of adult *L. intestinalis* was detected in the great crested grebe and great cormorant *Phalacrocorax carbo* (Linnaeus) in southern Moravia (Ryšavý and Sitko 1995, Levron et al. 2009, Moravec and Scholz 2016).

During our fish surveys, we spotted six potential final hosts – great cormorants (*P. carbo*), grebes *Podiceps auritus* (Linnaeus), *Podiceps cristatus*, *P. nigricollis* Brehm, and ducks (*Mergus merganser*, *Aythya ferina* (Linnaeus) at the visited reservoirs and lakes. However, several other species common to the area, such as herons and gulls, are thought to frequent the studied localities. Although the role of some of these species in the introduction and spread of the parasite between lakes seems to be of minor importance because their abundance is low or they rarely forage on fish, cormorants, grebes and herons are widespread and highly mobile species, thus allowing dispersal of *L. intestinalis*. Nevertheless, whilst dispersal rates seem sufficient to maintain gene flow across relatively large distances (Štefka et al. 2009), it is questionable whether they are sufficient to modulate differences in plerocercoid prevalence between the same hosts in different lakes. Whereas prevalence in some fish species was low in most lakes and cannot be directly contrasted, other species (roach and possibly freshwater bream, Table 2) showed marked differences. Although several ecological factors (e.g., fish abundance, primary host density, etc.) contribute to creating lake-specific patterns, we suggest that the propagule size and frequency of introduction of *L. intestinalis* by definitive hosts should also be considered. Our data do not allow for testing all possible factors, but we did not find a link between the fish host abundance and prevalence rate across the studied lakes (Fig. S2).

Phylogeny and population genetic structure of *L. intestinalis*

It has been previously acknowledged that the evolution of *Ligula* tapeworms is defined by the interplay between geography and host specificity at a global scale (Bouzid et al. 2008b). Interestingly, at a local scale, epidemiological differences between freshwater bream and roach infecting plerocercoids are known to exist (Loot et al. 2001), but neither a microsatellite study (Štefka et al. 2009), nor the most densely sampled mtDNA study of the Clade A to date (Bouzid et al. 2008b, Štefka et al. 2009) recovered any genetic differences between the freshwater bream and roach (plus bleak and rudd) samples. Somewhat unexpectedly, the analysis of population genetic structure at a fine scale in our study revealed a high degree of host specificity of several haplotypes of *L. intestinalis* to freshwater bream. Even though our sample size is small, the pattern becomes obvious also in a larger sample, when the new data are analysed with all available Clade A sequences (Fig. S4). Plerocercoids from breams create several clusters, or even if scattered into individual haplotypes they are not shared with other hosts. Such a pattern might indicate an ongoing process of lineage sorting. The results of both the PCoA analysis and the pairwise F_{ST} of the current samples also showed that *L. intestinalis* collected from the freshwater bream diverged from the rest of the samples, while the samples in roach, rudd and bleak demonstrated shared haplotypes and low F_{ST} differentiation. White bream samples were less frequently collected and whilst their haplotypes were not shared with any other species, they also did not create a single cluster.

Our results might indicate an emerging barrier between ecologically differentiated (host-specific) parasite populations. However, our data are too limited, both in the number of sequenced samples and the volume of genetic information, to conclude if the parasite populations have been influenced by an ecological speciation process and whether any reproductive barriers evolved between the host-specific clusters. Nevertheless, our finding represents a promising candidate for testing a hypothesis of speciation due to divergent natural selection in different hosts, as defined by Nosil (2012). Even though ecological speciation is understood as one of the primary mechanisms of sympatric speciation, the process can play a role in all geographical settings, including parapatric and allopatric layouts (Rundle and Nosil 2005, Nosil 2012).

Preliminary population genomic data, a RADseq analysis of plerocercoids performed by Kočová (2018), also suggest a differentiation between bream and the rest of the samples within Clade A. Other processes, such as a historical separation of populations of *L. intestinalis* from different hosts during the quaternary glaciation period followed by admixture, may have contributed to creating the current pattern of population structure. Thus, it is evident that additional epidemiological, genomic and transcriptome data are needed for populations of *L. intestinalis* sampled across different hosts before a firm conclusion could be reached on the ecological speciation routes of this parasite.

Acknowledgements. The authors are grateful to Anna Mácová and Roman Hrdlička for their help in fieldwork and to Tijana Cvetković for help with laboratory work. We would like to thank two anonymous reviewers for their helpful comments and suggestions. The research was supported by the Czech Science Foundation (grant no. GA19-04676S). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the program "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042) is appreciated.

Author contributions. MN concluded the research under the supervision of JŠ. MN and JŠ drafted the text with contributions by JP and PB. All authors provided study samples and oversaw the manuscript.

REFERENCES

Appelberg M., Berger H.-M., Hesthagen T., Kleiven E., Kurkilahti M., Raitaniemi J., Rask M. 1995: Development and intercalibration of methods in Nordic freshwater fish monitoring. Water. Air. Soil Pollut. 85: 401–406.

Arme C., Owen R.W. 1968: Occurrence and pathology of *Ligula intestinalis* infections in British fishes. J. Parasitol. 54: 272–280.

Barson M., Marshall B.E. 2003: The occurrence of the tapeworm *Ligula intestinalis* (L.), in *Barbus paludinosus* from a small dam in Zimbabwe. African J. Aquat. Sci. 28: 175–178.

Baruš V., Prokeš M. 1995: Length-weight relationship of *Ligula intestinalis* plerocercoids in adult silver bream and discussion on estimation of the parasite age. Appl. Parasitol. 36: 192–199.

Bean C.W., Winfield I.J. 1991: Influences of the tapeworm *Ligula intestinalis* (L.) on the spatial distributions of juvenile roach *Rutilus rutilus* (L.) and gudgeon *Gobio gobio* (L.) in Lough Neagh, Northern Ireland. Netherlands J. Zool.42: 419–429.

Bernard M.S., Strittmatter M., Murúa P., Heesch S., Cho G.Y., Leblanc C., Peters A.F. 2019: Diversity, biogeography and host specificity of kelp endophytes with a focus on the genera *Laminarionema* and *Laminariocolax* (Ectocarpales, Phaeophyceae). Eur. J. Phycol. 54: 39–51.

Blabolil P., Boukal D.S., Ricard D., Kubečka J., Říha M., Vašek M., Prchalová M., Čech M., Frouzová J., Jůza T., Muška M., Tušer M., Draštík V., Šmejkal M., Vejřík L., Peterka J. 2017: Optimal gillnet sampling design for the estimation of fish community indicators in heterogeneous freshwater ecosystems. Ecol. Indic. 77: 368–376.

Blatrix R., Herbers J.M. 2003: Coevolution between slave-making ants and their hosts: host specificity and geographical variation. Mol. Ecol. 12: 2809–2816.

Blouin M.S. 2002: Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer. Int. J. Parasitol. 32: 527–531.

Bouzid W., Lek S., Mace M., Ben Hassine O., Etienne R., Legal L., Loot G. 2008a: Genetic diversity of *Ligula intestinalis* (Cestoda: Diphyllobothriidea) based on analysis of inter-simple sequence repeat markers. J. Zool. Syst. Evol. Res. 46: 289–296.

Bouzid W., Štefka J., Bahri-Sfar L., Beerli P., Loot G., Lek S., Haddaoui N., Hypša V., Scholz T., Dkhil-Abbes T. 2013: Pathways of cryptic invasion in a fish parasite traced using coalescent analysis and epidemiological survey. Biol. Invasions 15: 1907–1923.

Bouzid W., Štefka J., Hypša V., Lek S., Scholz T., Legal L., Hassine O.K. Ben, Loot G. 2008b: Geography and host specificity: two forces behind the genetic structure of the freshwater fish parasite *Ligula intestinalis* (Cestoda: Diphyllobothriidae). Int. J. Parasitol. 38: 1465–1479.

Britton J.R., Jackson M.C., Harper D.M. 2009: *Ligula intestinalis* (Cestoda: Diphyllobothriidae) in Kenya: a field investigation into host specificity and behavioural alterations. Parasitology 136: 1367–1373.

Brown S.P., Loot G., Grenfell B.T., Guégan J.F. 2001: Host manipulation by *Ligula intestinalis*: accident or adaptation? Parasitology 123: 519.

Bykhovskaya-Pavlovskaya I.E. 1964: Key to Parasites of Freshwater Fish of the USSR. Israel Program for Scientific Translations, Jerusalem, 919 pp.

Carter B.C., Shubeita G.T., Gross S.P. 2005: Tracking single particles: a user-friendly quantitative evaluation. Phys. Biol. 2: 60.

Chapman A., Hobbs R.P., Morgan D.L., Gill H.S. 2006: Helminth parasitism of *Galaxias maculatus* (Jenyns, 1842) in southwestern Australia. Ecol. Freshw. Fish. 15: 559–564.

Cole R., Viney M. 2019: Correction to: The population genetics of parasitic nematodes of wild animals. Parasit. Vectors. 12: 1.

Dence W.A. 1958: Studies on *Ligula*-infected common shiners (*Notropis cornutus frontalis* Agassiz) in the Adirondacks. J. Parasitol 44: 334–338.

Dubinina M.N. 1980: Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR. Amerind Publishing Company, Delhi, 320 pp.

Ergonul M.B., Altindag A. 2005: The occurrence and dynamics of *Ligula intestinalis* in its cyprinid fish host, tench, *Tinca tinca*, in Mogan Lake (Ankara, Turkey). Vet. Med. 50: 537.

Groves K.L., Shields B.A. 2001: Observations on the plerocercoid stage of the tapeworm *Ligula* in three species of fish from the lower Crooked River of central Oregon. J. Aquat. Anim. Health. 13: 285–289.

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018: UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35: 518–522.

Holmes J.C., Bethel W.M. 1972: Modification of intermediate host behavior by parasites. In: E.U. Canning and C.A. Wright (Eds.), Behavioral Aspects of Parasite Transmission, Academic Press, New York, pp. 128–149.

Huyse T., Poulin R., Théron A., Theron A. 2005: Speciation in parasites: a population genetics approach. Trends Parasitol. 21: 469–475.

Ivankov V.N., Kaplunenko V.A., Bol'shakov S.G., Zheleznova L. V. 2020: First detections of the tapeworm *Ligula intestinalis* (Linnaeus, 1758) (Cestoda: Ligulidae) in the anadromous far eastern redfin *Tribolodon hakonensis* (Gunther, 1880) (Teleostei: Cyprinidae) in Primorye. Russ. J. Mar. Biol. 46: 230–231.

Iwanowicz D.D. 2011: Overview on the effects of parasites on fish health. In: R.C. Cipriano, A.W. Bruckner and I.S. Shchelkunov (Eds.), Proceedings of the Third Bilateral Conference between Russia and the United States, Aquatic Animal Health 2009, Shepherdstown, 12–20 July 2009. Khaled bin Sultan Living Oceans Foundation, Landover, pp. 176–184.

Jombart T. 2008: *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405.

Kennedy C.R. 1974: A checklist of British and Irish freshwater fish parasites with notes on their distribution. J. Fish Biol. 6: 613–644.

Kennedy C.R., Burrough R.J. 1981: The establishment and subsequent history of a population of *Ligula intestinalis* in roach *Rutilis rutilis* (L.). J. Fish Biol. 19: 105–126.

Kennedy C.R., Shears P.C., Shears J.A. 2001: Long-term dynamics of *Ligula intestinalis* and roach *Rutilus rutilus*: a study of three epizootic cycles over thirty-one years. Parasitology 123: 257–269.

Kočová P. 2018: [Population-genomic analysis of adaptation in a parasite with a wide host range – tapeworm *Ligula intestinalis*.] Master thesis, University of South Bohemia in České Budějovice. (In Czech.)

Lagrue C., Presswell B., Dunckley N., Poulin R. 2018: The invasive cestode parasite *Ligula* from salmonids and bullies on the South Island, New Zealand. Parasitol. Res. 117: 151–156.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2017: Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34: 772–773.

Leigh J.W., Bryant D. 2015: POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6: 1110–1116.

Levron C., Sitko J., Scholz T. 2009: Spermiogenesis and spermatozoon of the tapeworm *Ligula intestinalis* (Diphyllobothriidea): phylogenetic implications. J. Parasitol. 95: 1–9.

Li J., Liao X., Yang H. 2000: Molecular characterization of a parasitic tapeworm (*Ligula*) based on DNA sequences from formalin-fixed specimens. Biochem. Genet. 38: 309–322.

Librado P., Rozas J. 2009: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Longshaw M., Frear P.A., Nunn A.D., Cowx I.G., Feist S.W. 2010: The influence of parasitism on fish population success. Fish. Manag. Ecol. 17: 426–434.

Loot G., Francisco P., Santoul F., Lek S., Guégan J.-F. 2001: The three hosts of the *Ligula intestinalis* (Cestoda) life cycle in Lavernose-Lacasse gravel pit, France. Arch. Hydrobiol. 152: 511–525.

Martinů J., Hypša V., Štefka J. 2018: Host specificity driving genetic structure and diversity in ectoparasite populations: coevolutionary patterns in *Apodemus* mice and their lice. Ecol. Evol. 8: 10008–10022.

Meinilä M., Kuusela J., Ziętara M.S., Lumme J. 2004: Initial steps of speciation by geographic isolation and host switch in salmonid pathogen *Gyrodactylus salaris* (Monogenea: Gyrodactylidae). Int. J. Parasitol. 34: 515–526.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020: IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37: 1530–1534.

Ministry of Agriculture of the Czech Republic 2019: Report on Water Management in the Czech Republic in 2019, Accessed April 25, 2022, https://eagri.cz/Report

Moravec F., Scholz T. 2016: Helminth parasites of the lesser great cormorant *Phalacrocorax carbo sinensis* from two nesting regions in the Czech Republic. Folia Parasitol. 63: 022.

Morgan D.L. 2003: Distribution and biology of *Galaxias truttaceus* (Galaxiidae) in south-western Australia, including first evidence of parasitism of fishes in Western Australia by *Ligula intestinalis* (Cestoda). Environ. BioL. Fishes. 66:155–167.

Nazarizadeh M., Martinů J., Nováková M., Stanko M., Štefka J. 2022: Phylogeography of the parasitic mite *Laelaps agilis* in Western Palearctic shows lineages lacking host specificity but possessing different demographic histories. BMC Zool. 7: 15.

Nezafat R.B., Khara H., Satari M. 2008: Parasite infection of bream (*Abramis brama orientalis* Berg, 1949) in Aras Dam lake. J. Biol. Sci. 2: 83–96.

Nosil P. 2012: Ecological Speciation. Oxford University Press, Oxford, 280 pp.

Olson P.D., Littlewood D.T.J., Griffiths D., Kennedy C.R., Arme C. 2002: Evidence for the co-existence of separate strains or species of *Ligula* in Lough Neagh, Northern Ireland. J. Helminthol. 76: 171.

Palm H.W., Theisen S., Pikalov E., Kleinertz S. 2018: An update: manipulation of fish phenotype by parasites. Ref. Mod. Life Sci. 2018: 1–9.

R Core Team 2021: R: a language and environment for statistical computing.

Ristau K., Steinfartz S., Traunspurger W. 2013: First evidence of cryptic species diversity and significant population structure in a widespread freshwater nematode morphospecies (*Tobrilus gracilis*). Mol. Ecol. 22: 4562–4575.

Ronquist F., Huelsenbeck J.P. 2003: MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Rundle H.D., Nosil P. 2005: Ecological speciation. Ecol. Lett. 8: 336–352.

Ryšavý B., Sitko J. 1995: New findings of tapeworms (Cestoda) of birds from Moravia and synopsis of bird Cestodes from Czech Republic. Acta Sci. Nat. Brno 29: 1–66.

Schirrmann M.K., Leuchtmann A. 2015: The role of host-specificity in the reproductive isolation of *Epichloë endophytes* revealed by reciprocal infections. Fungal Ecol. 15: 29–38.

Štefka J., Hypša V., Scholz T. 2009: Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: microsatellite study of *Ligula intestinalis* (Cestoda). Mol. Ecol. 18: 1187–1206.

Stork N.E., Lyal C.H.C. 1993: Extinction or 'co-extinction' rates? Nature 366: 307.

Sweeting R.A. 1977: Studies on *Ligula intestinalis.* Some aspects of the pathology in the second intermediate host. J. Fish Biol. 10: 43–50.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. 2011: MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.

Van Dobben W.H. 1952: The food of the cormorant in the Netherlands. Ardea 55: 1–63.

Waeschenbach A., Brabec J., Scholz T., Littlewood D.T.J., Kuchta R. 2017: The catholic taste of broad tapeworms – multiple routes to human infection. Int. J. Parasitol. 47: 831–843.

Wells K., Clark N.J. 2019: Host specificity in variable environments. Trends Parasitol. 35: 452–465.

Windsor D.A. 1998: Controversies in parasitology. Most of the species on Earth are parasites. Int. J. Parasitol. 28: 1939–1941.

Wyatt R.J., Kennedy C.R. 1989: Host-constrained epidemiology of the fish tapeworm *Ligula intestinalis* (L.). J. Fish Biol. 35: 215–227.

Xia X., Lemey P. 2009: Assessing substitution saturation with DAMBE. In: P. Lemey, M. Salemi, A.-M. Vandamme (Eds.), The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. Second Edition. Cambridge University Press, Cambridge, pp. 615–630.

Xia X., Xie Z. 2001: DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. 92: 371–373.

ZAR J.H. 1999: Biostatistical Analysis. Prentice Hall, New Jersey, 663 pp.

Received 29 October 2021 Accepted 26 April 2022 Published online 15 September 2022

Cite this article as: Nazarizadeh M., Peterka J., Kubečka J., Vašek M., Jůza T., de Moraes K.R., Čech M., Holubová M., Souza A.T., Blabolil P., Muška M., Tsering L., Bartoň D., Říha M., Šmejkal M., Tušer M., Vejřík L., Frouzová J., Jarić I., Prchalová M., Vejříková I., Štefka J. 2022: Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies. Folia Parasitol. 69: 018.

Institute of Parasitology, Biology Centre CAS
Folia Parasitologica 2022, 69: 018: Supplementary files
Supplementary **Tables S1–S3**, **Figs. 1–4**

Folia Parasitologica
http://folia.paru.cas.cz

Research Article OPEN ACCESS

Nazarizadeh M., Peterka J., Kubečka J., Vašek M., Jůza T., de Moraes K.R., Čech M., Holubová M., Souza A.T., Blabolil P., Muška M., Tsering L., Bartoň D., Říha M., Šmejkal M., Tušer M., Vejřík L., Frouzová J., Jarić I., Prchalová M., Vejříková I., Štefka J. 2022: Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies. Folia Parasitol. 69: 018. DOI 10.14411/fp.2022.018

Table S1. List of samples, geographical origin, and accession numbers of the genes (Cytb and COI) of the *Ligula* complex used in phylogenetic and population genetic structure analyses.

ID	Cytb	COI	Accession number Cytb	COI	host	locality	references
brama_Lip20	914 bp	396 bp	OP408033	OP390380	*Abramis brama*	Lipno Reservoir	present study
brama_Lip1	914 bp	396 bp	OP408034	OP390381	*Abramis brama*	Lipno Reservoir	present study
brama_Lip2	914 bp	396 bp	OP408035	OP390382	*Abramis brama*	Lipno Reservoir	present study
brama_Lip14	914 bp	396 bp	OP408036	OP390383	*Abramis brama*	Lipno Reservoir	present study
brama_Rimov1	914 bp	396 bp	OP408037	OP390384	*Abramis brama*	Římov Reservoir	present study
blicc_Lip	914 bp	96 bp	OP408038	OP390385	*Blicca bjoerkna*	Lipno Reservoir	present study
blicc_Rimov2	914 bp	396 bp	OP408039	OP390386	*Blicca bjoerkna*	Římov Reservoir	present study
eryth_Most1	914 bp	396 bp	OP408040	OP390387	*Scardinius erythrophthalmus*	Most Lake	present study
eryth_Most58	914 bp	396 bp	OP408041	OP390388	*Scardinius erythrophthalmus*	Most Lake	present study
eryth_Most44	914 bp	396 bp	OP408042	OP390389	*Scardinius erythrophthalmus*	Most Lake	present study
eryth_Most2	914 bp	396 bp	OP408043	OP390390	*Scardinius erythrophthalmus*	Most Lake	present study
Rutil_Most51	914 bp	396 bp	OP408044	OP390391	*Rutilus rutilus*	Most Lake	present study
Rutil_Medard3	914 bp	396 bp	OP408045	OP390392	*Rutilus rutilus*	Medard Lake	present study
Rutil_Most40	914 bp	396 bp	OP408046	OP390393	*Rutilus rutilus*	Most Lake	present study
Rutil_Rimov9	914 bp	396 bp	OP408047	OP390394	*Rutilus rutilus*	Římov Lake	present study
Rutil_Most34	914 bp	396 bp	OP408048	OP390395	*Rutilus rutilus*	Most Lake	present study
Rutil_Most48	914 bp	396 bp	OP408049	OP390396	*Rutilus rutilus*	Most Lake	present study
Rutil_Most28	914 bp	396 bp	OP408050	OP390397	*Rutilus rutilus*	Most Lake	present study
Rutil_Most11	914 bp	396 bp	OP408051	OP390398	*Rutilus rutilus*	Most Lake	present study
Rutil_Most27	914 bp	396 bp	OP408052	OP390399	*Rutilus rutilus*	Most Lake	present study
Rutil_Most3	914 bp	396 bp	OP408053	OP390400	*Rutilus rutilus*	Most Lake	present study
Albur_LipnO2	914 bp	396 bp	OP408054	OP390401	*Alburnus alburnus*	Lipno Reservoir	present study
Albur1_Lipno	914 bp	396 bp	OP408055	OP390402	*Alburnus alburnus*	Lipno Reservoir	present study
Albur2_Lipno	914 bp	396 bp	OP408056	OP390403	*Alburnus alburnus*	Lipno Reservoir	present study
Albur3_Lipno	914 bp	396 bp	OP408057	OP390404	*Alburnus alburnus*	Lipno Reservoir	present study
Rutil_Most29	914 bp	396 bp	OP408058	OP390405	*Rutilus rutilus*	Most Lake	present study
Rutil_Medard	914 bp	396 bp	OP408059	OP390406	*Rutilus rutilus*	Medard Lake	present study
Rutil_Medard2	914 bp	396 bp	OP408060	OP390407	*Rutilus rutilus*	Medard Lake	present study
Rutil_Medard4	914 bp	396 bp	OP408061	OP390408	*Rutilus rutilus*	Medard Lake	present study
Rutil_Medard5	914 bp	396 bp	OP408062	OP390409	*Rutilus rutilus*	Medard Lake	present study
CZ2Pc	405 bp	396 bp	EU241190	EU241293	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
CZ6Rr	405 bp	396 bp	EU241169	EU241277	*Rutilus rutilus*	Czech Republic	Bouzid et al. 2008
CZ7Rr	405 bp	396 bp	EU241159	EU241278	*Rutilus rutilus*	Czech Republic	Bouzid et al. 2008
CZ10Rr	405 bp	396 bp	EU241165	EU241280	*Rutilus rutilus*	Czech Republic	Bouzid et al. 2008
CZ12Rr	405 bp	396 bp	EU241170	EU241281	*Rutilus rutilus*	Czech Republic	Bouzid et al. 2008
CZ14Ab	405 bp	396 bp	EU241182	EU241263	*Abramis brama*	Czech Republic	Bouzid et al. 2008
CZ16Ab	405 bp	396 bp	EU241179	EU241283	*Abramis brama*	Czech Republic	Bouzid et al. 2008
CZ17Ab	405 bp	396 bp	EU241183	EU241284	*Abramis brama*	Czech Republic	Bouzid et al. 2008
CZ18Ab	405 bp	396 bp	EU241184	EU241285	*Abramis brama*	Czech Republic	Bouzid et al. 2008
CZ24Ab	405 bp	396 bp	EU241180	EU241267	*Abramis brama*	Czech Republic	Bouzid et al. 2008
CZ45Pc	405 bp	396 bp	EU241186	EU241288	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
CZ90Rr	405 bp	396 bp	EU241178	EU241282	*Rutilus rutilus*	Czech Republic	Bouzid et al. 2008
CZ69Aa	405 bp	396 bp	EU241177	EU241246	*Rutilus rutilus*	Czech Republic	Bouzid et al. 2008
CZ104Pc	405 bp	396 bp	EU241191	EU241291	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
CZ107Pc	405 bp	396 bp	EU241168	EU241262	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
CZ106Pc	405 bp	396 bp	EU241167	EU241244	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
CZ109Pc	405 bp	396 bp	EU241164	EU241245	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
CZ121Mm	405 bp	396 bp	EU241193	EU241239	*Mergus merganser*	Czech Republic	Bouzid et al. 2008
CZ114Pc	405 bp	396 bp	EU241187	EU241287	*Podiceps cristatus*	Czech Republic	Bouzid et al. 2008
RU3Ab	405 bp	396 bp	EU241212	EU241252	*Abramis brama*	Russia	Bouzid et al. 2008
RU4Ab	405 bp	396 bp	EU241158	EU241253	*Abramis brama*	Russia	Bouzid et al. 2008
RU5Ab	405 bp	396 bp	EU241210	EU241309	*Abramis brama*	Russia	Bouzid et al. 2008

Address for correspondence: Jan Štefka, Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Branišovská 31, České Budějovice; E-mail: jan.stefka@gmail.com

RU7Rr	405 bp	396 bp	EU241213	EU241255	*Rutilus rutilus*	Russia	Bouzid et al. 2008
RU8Ab	405 bp	396 bp	EU241211	EU241310	*Abramis Brama*	Russia	Bouzid et al. 2008
TN26Rb	405 bp	396 bp	JQ279127	JQ279091	*Sarmarutilus rubilio*	Tunisia	Bouzid et al. 2013
TN50Se	405 bp	396 bp	JQ279130	JQ279093	*Scardinius erythrophthalmus*	Tunisia	Bouzid et al. 2013
TN59Rb	405 bp	396 bp	JQ279135	JQ279098	*Sarmarutilus rubilio*	Tunisia	Bouzid et al. 2013
TN60Rb	405 bp	396 bp	JQ279136	JQ279099	*Sarmarutilus rubilio*	Tunisia	Bouzid et al. 2013
TN66Se	405 bp	396 bp	JQ279141	JQ279103	*Scardinius erythrophthalmus*	Tunisia	Bouzid et al. 2013
TN63Ps	405 bp	396 bp	JQ279139	JQ279102	*Pseudophoxinus callensis*	Tunisia	Bouzid et al. 2013
TN58Se	405 bp	396 bp	JQ279131	JQ279094	*Scardinius erythrophthalmus*	Tunisia	Bouzid et al. 2013
TN5Rr	405 bp	396 bp	EU241215	EU241312	*Rutilus rutilus*	Tunisia	Bouzid et al. 2008
TN49Se	405 bp	396 bp	EU241162	EU241272	*Scardinius erythrophthalmus*	Tunisia	Bouzid et al. 2008
DE1Rr	405 bp	396 bp	EU241203	EU241301	*Rutilus rutilus*	Germany	Bouzid et al. 2008
DE3Rr	405 bp	396 bp	EU241204	EU241302	*Rutilus rutilus*	Germany	Bouzid et al. 2008
DE4Rr	405 bp	396 bp	EU241202	EU241273	*Rutilus rutilus*	Germany	Bouzid et al. 2008
DE5Rr	405 bp	396 bp	EU241185	EU241274	*Rutilus rutilus*	Germany	Bouzid et al. 2008
FR11Rr	405 bp	396 bp	EU241199	EU241258	*Rutilus rutilus*	France	Bouzid et al. 2008
FR30Ab	405 bp	396 bp	EU241201	EU241259	*Abramis brama*	France	Bouzid et al. 2008
FR64Aa	405 bp	396 bp	EU241163	EU241260	*Alburnus alburnus*	France	Bouzid et al. 2008
FR82Rr	405 bp	396 bp	EU241200	EU241261	*Rutilus rutilus*	France	Bouzid et al. 2008
FR90Rr	405 bp	396 bp	EU241172	EU241299	*Rutilus rutilus*	France	Bouzid et al. 2008
FR91Rr	405 bp	396 bp	EU241173	EU241300	*Rutilus rutilus*	France	Bouzid et al. 2008
IE1Rr	405 bp	396 bp	EU241171	EU241248	*Rutilus rutilus*	Ireland	Bouzid et al. 2008
IE2Rr	405 bp	396 bp	EU241206	EU241250	*Rutilus rutilus*	Ireland	Bouzid et al. 2008
IE3Gg	405 bp	396 bp	EU241188	EU241305	*Gobio gobio*	Ireland	Bouzid et al. 2008
IE4Gg	405 bp	396 bp	EU241208	EU241290	*Gobio gobio*	Ireland	Bouzid et al. 2008
IE5Gg	405 bp	396 bp	EU241189	EU241289	*Gobio gobio*	Ireland	Bouzid et al. 2008
IE6Rr	405 bp	396 bp	EU241207	EU241249	*Rutilus rutilus*	Ireland	Bouzid et al. 2008
GB1Rr	405 bp	396 bp	EU241205	EU241303	*Rutilus rutilus*	United Kingdom	Bouzid et al. 2008
ET7Bi	405 bp	396 bp	EU241198	EU241297	*Labeobarbus intermedius*	Ethiopia	Bouzid et al. 2008
ET5Bt	405 bp	396 bp	EU241197	EU241296	*Labeobarbus tsanesis*	Ethiopia	Bouzid et al. 2008
ET2Bh	405 bp	396 bp	EU241196	EU241295	*Enteromius humilis*	Ethiopia	Bouzid et al. 2008
EE2Ab	405 bp	396 bp	EU241192	EU241275	*Abramis brama*	Estonia	Bouzid et al. 2008
EE3Ab	405 bp	396 bp	EU241160	EU241276	*Abramis brama*	Estonia	Bouzid et al. 2008
EE4Ab	405 bp	396 bp	EU241195	EU241294	*Abramis brama*	Estonia	Bouzid et al. 2008
PL8Ra	405 bp	396 bp	EU241194	EU241292	*Rhodeus amarus*	Poland	Bouzid et al. 2008
UA1Aa	405 bp	396 bp	EU241181	EU241316	*Alburnus alburnus*	Ukraine	Bouzid et al. 2008
UA2Cc	405 bp	396 bp	EU241218	EU241238	*Carassius carassius*	Ukraine	Bouzid et al. 2008
UA3Rr	405 bp	396 bp	EU241176	EU241317	*Rutilus rutilus*	Ukraine	Bouzid et al. 2008
GB2Pp	405 bp	396 bp	EU241175	EU241304	*Phoxinus phoxinus*	United Kingdom	Bouzid et al. 2008
GB3Pp	405 bp	396 bp	EU241161	EU241247	*Phoxinus phoxinus*	United Kingdom	Bouzid et al. 2008
CN3Nt	405 bp	396 bp	EU241155	EU241235	*Neosalanx taihuensis*	China	Bouzid et al. 2008
CN4Nt	405 bp	396 bp	EU241157	EU241237	*Neosalanx taihuensis*	China	Bouzid et al. 2008
CN16Nt	405 bp	396 bp	EU241156	EU241236	*Neosalanx taihuensis*	China	Bouzid et al. 2008
CA1Cp	405 bp	396 bp	EU241152	EU241228	*Coulsius plumbeus*	Canada	Bouzid et al. 2008
CA2Sa	405 bp	396 bp	EU241148	EU241225	*Semotilus atromaculatus*	Canada	Bouzid et al. 2008
CA5Sa	405 bp	396 bp	EU241149	EU241226	*Semotilus atromaculatus*	Canada	Bouzid et al. 2008
CA14Sa	405 bp	396 bp	EU241151	EU241227	*Semotilus atromaculatus*	Canada	Bouzid et al. 2008
CA19Sa	405 bp	396 bp	EU241150	EU241224	*Semotilus atromaculatus*	Canada	Bouzid et al. 2008
AU1Gt	405 bp	396 bp	EU241146	EU241222	*Galaxias truttaceus*	Australia	Bouzid et al. 2008
AU4Gm	405 bp	396 bp	EU241147	EU241223	*Galaxias maculatus*	Australia	Bouzid et al. 2008
ALG2Bc	405 bp	396 bp	EU241143	EU241219	*Barbinae* gen. et sp.	Algeria	Bouzid et al. 2008
ALGB3Bc	405 bp	396 bp	EU241144	EU241220	*Barbinae* gen. et sp.	Algeria	Bouzid et al. 2008
ALG4Bc	405 bp	396 bp	EU241145	EU241221	*Barbinae* gen. et sp.	Algeria	Bouzid et al. 2008
CN1Hb	405 bp	396 bp	EU241153	EU241229.1	*Hemiculter bleekeri*	China	Bouzid et al. 2008
CN2Hb	405 bp	396 bp	EU241154	EU241230.1	*Hemiculter bleekeri*	China	Bouzid et al. 2008
RU1Hl	405 bp	396 bp	EU241209	EU241311.1	*Hemiculter lucidus*	Russia	Bouzid et al. 2008

Table S2. Results of BIC model selection conducted in PartitionFinder for different partitions of the dataset.

Partition	BIC
Cytb-, position 1	HKY+I+G
Cytb-, position 2	GTR+I
Cytb-, position 3	GTR+G
Cytb	GTR+I+G
COI-, position 1	HKY+I+G
COI-, position 2	GTR+G
COI-, position 3	GTR+G
COI	GTR+I+G

Table S3. Abundance [inds./1000 m^2] of fish older than young-of-the-year of the five selected cyprinid fish species in gillnet catches in different habitats in investigated lakes and reservoirs. n.s. denotes depths that were not sampled.

Water body:	Milada Lake					Most Lake					Medard Lake			
Fish species:	Roach	Rudd	White bream	Bream	Bleak	Roach	Rudd	White bream	Bream	Bleak	Roach	Rudd	White bream	Bream
Benthic habitats:														
0–3 m	194.4	39.6	0.0	0.0	0.0	50.0	113.9	0.0	0.0	0.0	0.0	55.6	0.0	0.0
3–6 m	227.8	2.8	0.0	0.0	0.0	97.2	91.7	0.0	0.0	0.0	11.1	2.8	2.8	5.6
6–9 m	138.9	0.0	0.0	0.0	0.0	83.3	47.2	0.0	0.0	0.0	52.8	8.3	1.4	13.9
9–12 m	75.0	0.0	0.0	0.0	0.0	94.4	27.8	0.0	0.0	0.0	5.6	0.0	1.4	2.8
Average	159.0	10.6	0.0	0.0	0.0	81.3	70.1	0.0	0.0	0.0	17.4	16.7	1.4	5.6
Pelagic habitats:														
0–3 m	127.8	47.2	0.0	0.0	0.0	3.5	2.7	0.0	0.0	0.0	0.0	36.1	0.0	0.0
3–6 m	38.9	0.0	0.0	0.0	0.0	11.9	10.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6–9 m	30.6	0.0	0.0	0.0	0.0	8.7	2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9–12 m	0.0	0.0	0.0	0.0	0.0	10.2	2.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Average	49.3	11.8	0.0	0.0	0.0	8.6	4.7	0.0	0.0	0.0	0.0	9.0	0.0	0.0
Sum of averages	208.3	22.4	0.0	0.0	0.0	89.8	74.8	0.0	0.0	0.0	17.4	25.7	1.4	5.6

Waterbody:	Jordán Reservoir					Římov Reservoir					Lipno Reservoir				
Fish species:	Bleak	Roach	Rudd	White bream	Bream	Bleak	Roach	Rudd	White bream	Bream	Bleak	Roach	Rudd	White bream	Bream	Bleak
Benthic habitats:																
0–3 m	0.0	313.9	13.9	161.1	50.0	61.1	133.0	0.0	13.3	37.5	15.6	410.0	0.0	65.3	16.7	4.2
3–6 m	0.0	30.6	0.0	133.3	8.3	8.3	60.0	0.0	18.3	46.2	0.0	230.0	0.0	145.0	21.2	1.5
6–9 m	0.0	0.0	0.0	5.6	0.0	0.0	4.2	0.0	0.0	9.4	4.2	4.2	0.0	75.0	15.6	0.0
9–12 m	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Average	0.0	86.1	3.5	75.0	14.6	17.4	49.3	0.0	20.4	23.3	4.9	161.0	0.0	71.3	13.4	1.4
Pelagic habitats:																
0–3 m	0.0	461.1	22.2	41.7	11.1	463.9	27.1	0.0	14.6	11.8	407.0	30.6	2.1	5.6	12.7	675.0
3–6 m	0.0	208.3	0.0	252.8	0.0	8.3	8.3	0.0	0.0	0.0	0.0	6.5	0.0	38.0	4.6	20.4
6–9 m	0.0	2.8	0.0	8.3	0.0	0.0	11.1	0.0	0.0	0.0	22.2	0.0	0.0	0.0	0.0	8.3
9–12 m	0.0	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Average	0.0	224.1	7.4	100.9	3.7	157.4	15.5	0.0	4.9	3.9	143.1	12.4	0.7	14.5	5.8	234.6
Sum of averages	0.0	310.2	10.9	175.9	18.3	174.8	64.8	0.0	25.3	27.2	148.0	173.4	0.7	85.8	19.2	236.0

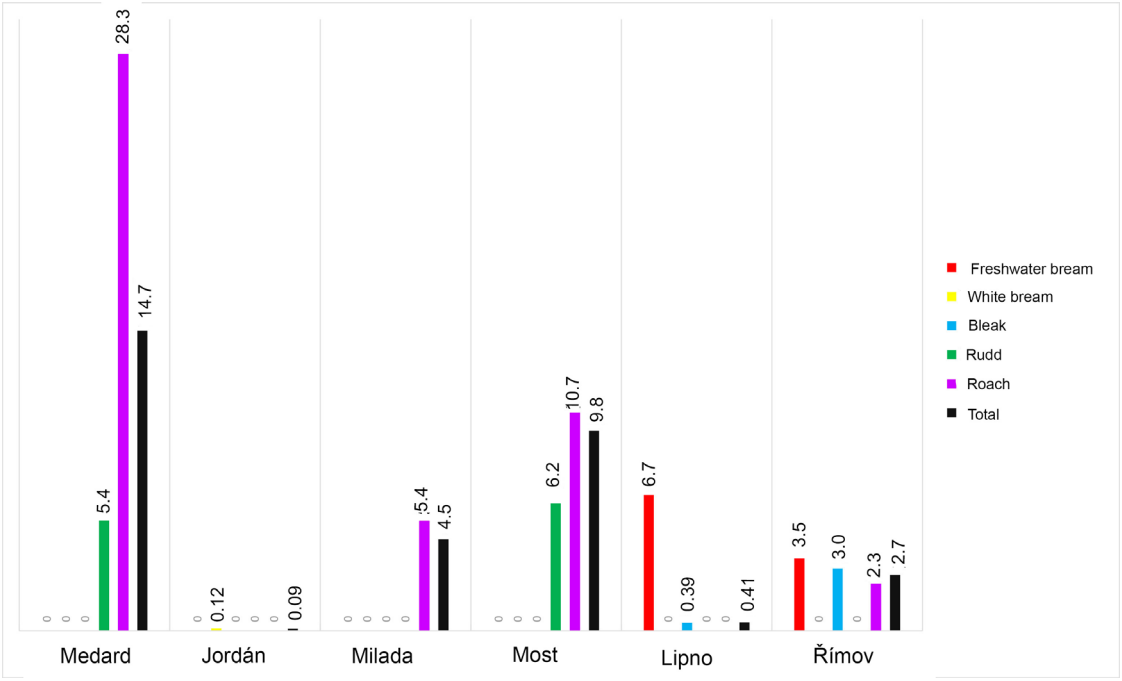


Fig. S1. Prevalence (in %) of *Ligula intestinalis* (Linnaeus, 1758) in six water bodies and five cyprind hosts sampled in the Czech Republic.

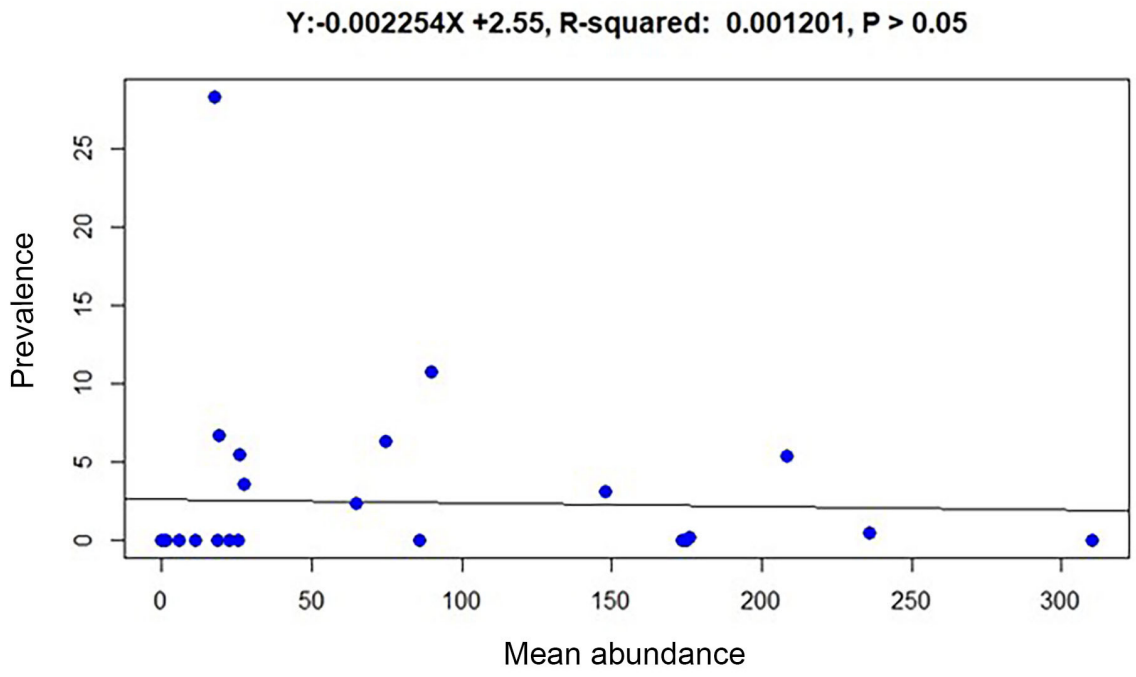


Fig. S2. Plot of the linear regression between prevalence and fish abundance in Czech water bodies.

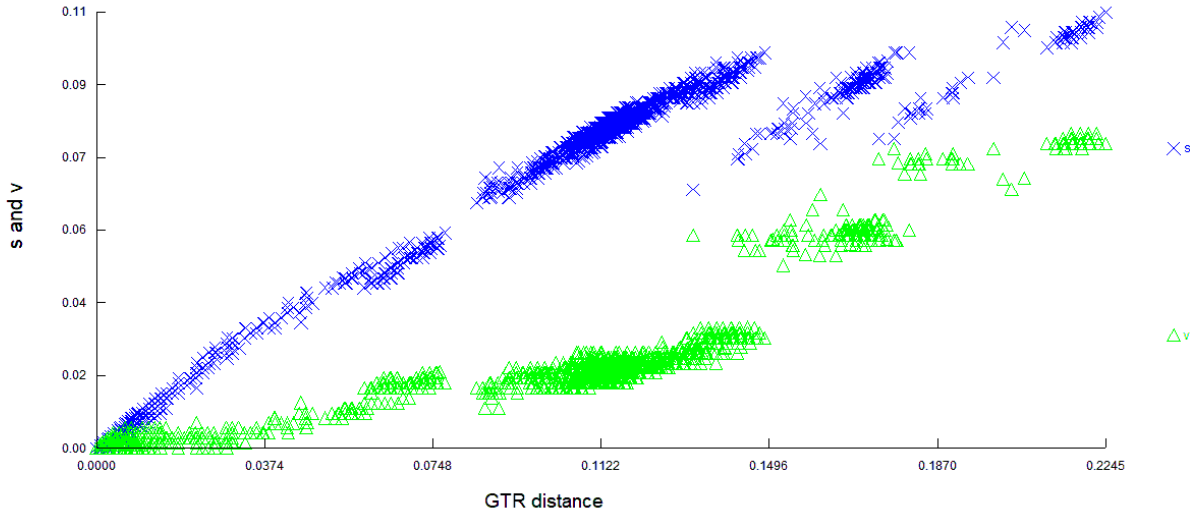


Fig. S3. Plot of the total numbers of transitions (s) and transversions (v) against corrected distances based on the GTR model revealed no trend toward a saturation for transversions.

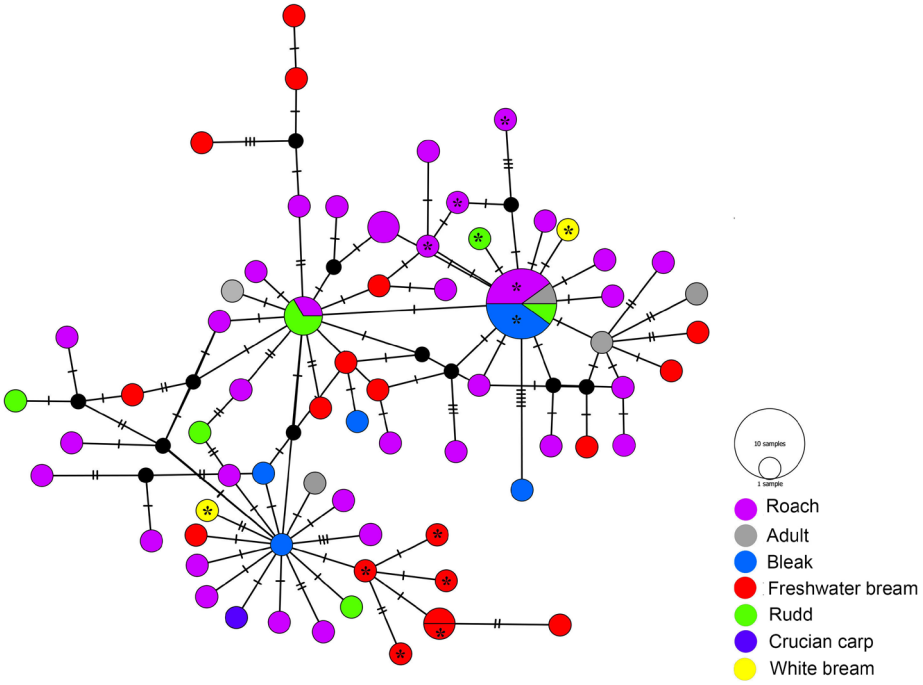


Fig. S4. Haplotype network of 80 *Ligula intestinalis* (Linnaeus, 1758) samples belonging to the Clade A based on 405 bp Cytb and 398 bp COI fragments (801bp) overlapping with previous studies. The size of each circle (haplotype) is proportional to the number of individuals and the short lines represent one mutation step. Asterisks mark haplotypes from the 30 new sequences. Adult tapeworm samples from definitive hosts, either *Podiceps cristatus* (Linnaeus) or *Mergus merganser* (Linnaeus), are represented by grey circles.

REFERENCES

Bouzid W., Štefka J., Bahri-Sfar L., Beerli P., Loot G., Lek S., Haddaoui N., Hypša V., Scholz T., Dkhil-Abbes T. 2013: Pathways of cryptic invasion in a fish parasite traced using coalescent analysis and epidemiological survey. Biol. Invasions 15: 1907–1923.

Bouzid W., Štefka J., Hypša V., Lek S., Scholz T., Legal L., Hassine O.K. Ben, Loot G. 2008: Geography and host specificity: two forces behind the genetic structure of the freshwater fish parasite *Ligula intestinalis* (Cestoda: Diphyllobothriidae). Int. J. Parasitol. 38: 1465–1479.

Chapter II

Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite - the *Ligula intestinalis* species complex

Molecular Phylogenetics and Evolution
DOI:10.1016/j.ympev.2022.107677

Molecular Phylogenetics and Evolution 180 (2023) 107677

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite – The *Ligula intestinalis* species complex

Masoud Nazarizadeh [a,b], Milena Nováková [b], Géraldine Loot [c], Nestory P. Gabagambi [d], Faezeh Fatemizadeh [e], Odipo Osano [f], Bronwen Presswell [g], Robert Poulin [g], Zoltán Vitál [h], Tomáš Scholz [a,b], Ali Halajian [i], Emiliano Trucchi [j], Pavlína Kočová [k], Jan Štefka [a,b,*]

[a] *Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic*
[b] *Institute of Parasitology, Biology Centre CAS, České Budějovice, Czech Republic*
[c] *UMR-5174, EDB (Laboratoire Evolution and Diversité Biologique), CNRS, IRD, Université Toulouse III Paul Sabatier, France*
[d] *Tanzania Fisheries Research Institute, Kyela, Mbeya, Tanzania*
[e] *Department of Environmental Science, Faculty of Natural Resources, University of Tehran, Karaj, Iran*
[f] *School of Environmental Studies, University of Eldoret, Kenya*
[g] *Department of Zoology, University of Otago, Dunedin, New Zealand*
[h] *Research Center for Fisheries and Aquaculture, Institute of Aquaculture and Environmental Safety, Hungarian University of Agriculture and Life Sciences, Szarvas, Hungary*
[i] *Research Administration and Development, and 2-DSI-NRF SARChI Chair (Ecosystem health), Department of Biodiversity, University of Limpopo, South Africa*
[j] *Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy*
[k] *Čakovice, Týnec nad Sázavou, Czech Republic*

ARTICLE INFO

Keywords:
Historical biogeography
Dated phylogeny
Demographic history
Host specificity
Host switching

ABSTRACT

Studies on parasite biogeography and host spectrum provide insights into the processes driving parasite diversification. Global geographical distribution and a multi-host spectrum make the tapeworm *Ligula intestinalis* a promising model for studying both the vicariant and ecological modes of speciation in parasites. To understand the relative importance of host association and biogeography in the evolutionary history of this tapeworm, we analysed mtDNA and reduced-represented genomic SNP data for a total of 139 specimens collected from 18 fish-host genera across a distribution range representing 21 countries. Our results strongly supported the existence of at least 10 evolutionary lineages and estimated the deepest divergence at approximately 4.99–5.05 Mya, which is much younger than the diversification of the fish host genera and orders. Historical biogeography analyses revealed that the ancestor of the parasite diversified following multiple vicariance events and was widespread throughout the Palearctic, Afrotropical, and Nearctic between the late Miocene and early Pliocene. Cyprinoids were inferred as the ancestral hosts for the parasite. Later, from the late Pliocene to Pleistocene, new lineages emerged following a series of biogeographic dispersal and host-switching events. Although only a few of the current *Ligula* lineages show narrow host-specificity (to a single host genus), almost no host genera, even those that live in sympatry, overlapped between different *Ligula* lineages. Our analyses uncovered the impact of historical distribution shifts on host switching and the evolution of host specificity without parallel host-parasite co-speciation. Historical biogeography reconstructions also found that the parasite colonized several areas (Afrotropical and Australasian) much earlier than was suggested by only recent faunistic data.

1. Introduction

Host specificity is a key feature of parasitic and symbiotic organisms. It defines their ability to survive on different hosts (Poulin, 2011), disperse (Mácová et al., 2018) and maintain the genetic diversity of populations (Martinů et al., 2018; Wacker et al., 2019). The evolution of host specificity in multi-host parasites is a prospective candidate for the mechanism of sympatric adaptive speciation (Kalbe et al., 2016). For

* Corresponding author at: Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.
E-mail address: jan.stefka@gmail.com (J. Štefka).

https://doi.org/10.1016/j.ympev.2022.107677
Received 20 October 2022; Received in revised form 6 December 2022; Accepted 12 December 2022
Available online 24 December 2022

long, the evolution of host-specific lineages has been viewed mostly from the co-speciation (or co-diversification) angle, where the evolution of the parasite is directly associated with that of the host(s) (Hafner and Nadler, 1990; Štefka et al., 2011). However, the origin of host-specific lineages is often linked to a series of historical biogeographical events (Ding et al., 2022; Perrot-Minnot et al. 2018). Therefore, incongruence may appear between the parasite and host phylogenies for various reasons, including founder events, bottlenecks and extinctions either in the host or parasite, and, most importantly, due to host-switching that could have occurred over large-scale episodic periods of palaeo-environmental and climatic fluctuations, such as Quaternary glaciations (Hoberg et al., 2012; Johnson et al., 2002; Zarlenga et al., 2006). Under stable environmental conditions, co-divergence between closely associated organisms may occur (Haukisalmi et al., 2016), whereas under changing conditions, hosts possibly move in search of suitable habitats and find themselves in contact with species they did not previously overlap with. These fluctuations provide new opportunities for parasites to switch between host species (Hoberg et al., 2012; Hoberg and Brooks, 2008, 2013).

Following a successful host switch, gene flow may be maintained or ceased between the ancestral and newly established parasite population, depending on the duration of the distribution overlap between the host species (e. g., Techer et al., 2022; Wang et al., 2016). Phylogenomic and historical biogeography methods allow for reconstructing past distribution changes and associated gene flow events in parasite lineages, providing tools to study the emergence and diversification of parasite species, identification of cryptic lineages and prediction of disease transmission (Angst et al., 2022; Doyle et al., 2022). Yet, although metazoan parasites represent 40 % of all living metazoan species, relatively few studies have investigated the historical biogeography of parasitic species (Dobson et al., 2008). Instead, these studies often investigate the host–parasite co-phylogeny. Whilst the co-phylogenetic approach has demonstrated significant co-evolutionary signals in parasites with a direct life cycle (Rahmouni et al., 2022; Šimková et al., 2004, 2022), it often finds weak host–parasite co-evolutionary signals in generalist multi-host parasites or parasites with complex life cycles involving intermediate hosts, possessing more varied dispersal opportunities (Hoberg and Brooks, 2008; Perrot-Minnot et al. 2018). Importantly, the co-phylogenetic framework does not consider the evolution of host-specific lineages in parasites whose origin is much younger than the age of the hosts, where the co-phylogenetic methods simply cannot be applied (Bouzid et al., 2008a, 2008b).

To study the contribution of past processes to the evolution of host-specific parasite lineages, we applied historical biogeography methods to study the tapeworm *Ligula intestinalis* sensu lato (Cestoda: Diphyllobothriidea). Its widespread geographical distribution and multi-host spectrum make it a promising model for studying the vicariant and ecological modes of speciation (Hoole et al., 2010; Nazarizadeh et al., 2022a; Štefka et al., 2009). It is an obligatory endoparasite with a complex life cycle involving two intermediate hosts (crustacean copepod as the first and planktivorous fish as the second intermediate host) and birds feeding on the infected fish as the definitive host (Dubinina, 1980). *L. intestinalis* spends a maximum of only five days in its final host, rapidly entering the sexual phase and dispersing eggs into the water with host faeces, while the secondary larval stage, plerocercoid, represents the most important and conspicuous phase of the parasite's life cycle (Dubinina, 1980). It lasts at least 10 months and occurs in the peritoneal cavity of planktivorous fish, negatively impacting the health, fertility and behaviour of individual fish (Gabagambi et al., 2019; Loot et al., 2002) as well as the ecology of the entire fish population (Kennedy et al., 2001). *L. intestinalis* usually appears in cyprinoid fish [previously the family Cyprinidae], but is capable of invading intermediate hosts from various taxonomic groups such as Salmonidae, Galaxiidae and Catostomidae (Chapman et al., 2006; Dubinina, 1980; Štefka et al., 2009).

Studies exploring the phylogenetic relationships and population structure of the *L. intestinalis* species complex identified at least six major lineages with different levels of host and geographic distribution: (1) clade A (European and north African populations), (2) clade B (European, Chinese, Australian and North African populations), (3) *L. alternans* (Asia) [previously *L. digramma*], (4) Canada, (5) Chinese and (6) Ethiopian populations (Bouzid et al., 2008a, 2008b; Štefka et al., 2009). The studies indicated that both geography and host specificity may determine the genealogical relationships of the parasites by inducing genetic differentiation. However, their sampling lacked several biogeographical areas, and the historical demography of the lineages was not elucidated. Thus, the course of interaction between the historical geographical distribution of the parasite and the evolution of host-specific lineages has remained a contentious issue.

The main objective of the present study is to reveal the contribution of historical geographical distribution and past host usage (specificity) in the formation of contemporary lineage diversity in parasites with high dispersal capabilities. Using a geographically wide sampling of *Ligula* tapeworms as a study system, we characterised it with SNP genotyping and mitochondrial sequencing. We aimed to i) determine the true lineage diversity and levels of host specificity of parasites forming a species complex which lacks distinctive morphological traits; ii) provide dates for major diversification events in the evolutionary history of the parasite, and iii) determine the factors and events influencing lineage diversity in space and time by the analyses of historical areas, demographic history of lineages and past gene flow.

2. Materials and methods

2.1. Sample collection and DNA extraction

We analysed 139 specimens of *Ligula* plerocercoids collected from 18 fish genera over a broad geographic area representing 21 countries (Table S1). Integrity and quantity of newly isolated samples (extracted with DNeasy blood and tissue kit [Qiagen] from specimens preserved in 96 % ethanol) and DNA extracts from previous studies (Bouzid et al., 2008a, 2008b; Štefka et al., 2009) were verified on 0.8 % agarose gel and with Qubit 2.0 Fluorometer (Invitrogen). The final set of mitochondrial (86 new specimens were added to 76 sequences from Bouzid et al., 2008a, 2008b and Štefka et al., 2013 studies) and ddRAD (all new specimens) samples covered the distribution of the major lineages found in previous studies, which was complemented with several new areas to represent five terrestrial biogeographic realms: Nearctic, Palearctic, Afrotropical, Oriental and Australasian (Fig. 1, Table S1a and S1b).

2.2. Mitochondrial DNA sequencing

Amplification of three fragments of the mitochondrial genome, including cytochrome *b* (Cyt *b*), cytochrome oxidase subunit 1 (COI) and NADH dehydrogenase 1 (ND1) genes generated a concatenated data set with a total length of 1691 bp. All genes were amplified by polymerase chain reaction (PCR; see Supplementary Material, section 1) and sequenced using primers specifically designed for this study or previously published primers (Table S2). Sanger sequences were assembled and edited using Geneious Prime v.2022.1 (Gene Codes) and checked for possible stop codons based on the Alternative Flatworm Mitochondrial Code (transl_table 14).

2.3. ddRAD library preparation and reference genome

Libraries of ddRAD were prepared using 350 ng of DNA per sample following a modified version of the protocol proposed by Peterson et al. (2012) (Supplementary Material, section 2). Two samples of *Dibothriocephalus latus* (Diphyllobothriidea) were included as outgroups. Libraries were sequenced at several lanes of Illumina NovaSeq 150 bp PE (Novogene UK) yielding 6.8 million paired-end reads per sample on average. Steps for preparation of multiplexed ddRAD-seq libraries for

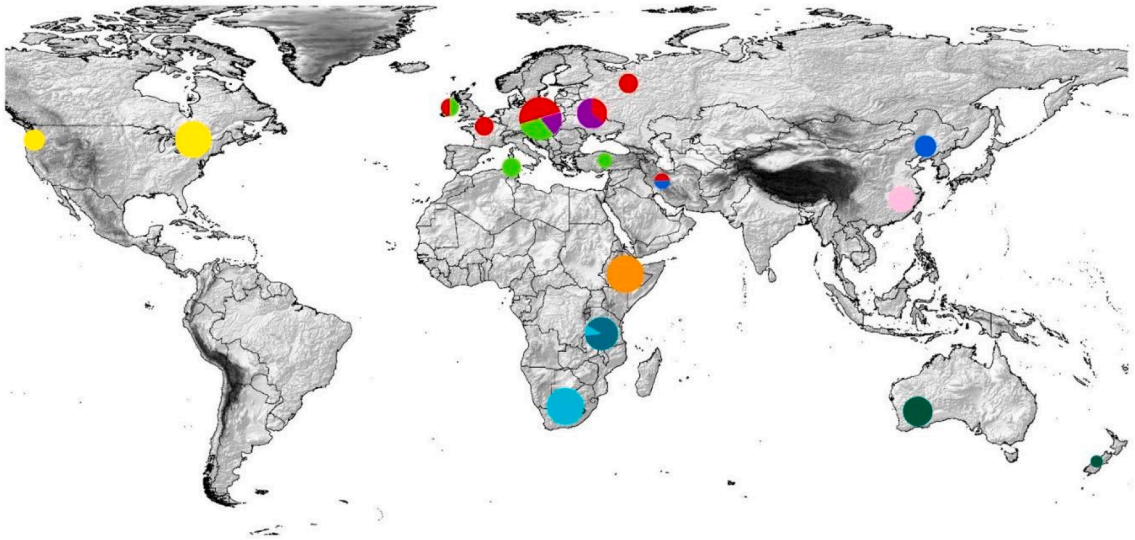


Fig. 1. *Ligula* sampling localities and patterns of genetic structure. Pie charts reflect the proportion of individuals assigned to each phylogenetic lineage (shown in Fig. 2); Lineage A (green), Lineage B (red), *L. alternans* (blue), China (light pink), Australia and New Zealand (dark green), *L. pavlovskii* (magenta), East African Rift (EAR; dark turquoise), Central and South Africa (CSA; light turquoise), Ethiopia (orange), and Nearctic (yellow). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Genetic characteristics of different parasite populations based on the mtDNA and ddRAD data sets. hap: Number of mitochondrial haplotypes, hd: Mitochondrial haplotype diversity, Pi: Mitochondrial nucleotide diversity, He: Expected heterozygosity, MLH: Multi-locus heterozygosity, Ho: Observed heterozygosity, Fis: Inbreeding coefficient.

Parasite lineages	Host order / genera	Geographic distribution	N (mtDNA/ ddRAD)	mtDNA (Cytb + COI)			ddRAD					
				hap	hd	Pi	He	MLH	Pi (SD)	Ho	Fis (SD)	Private alleles
Lineage A	***Cypriniformes*** / *Abramis, Blicca, Scardinus, Exos, Squalius, Rutilus, Phoxinus, Alburnus, Carassius*	Czech Republic, France, Germany, Ireland, Italy, Russia, United Kingdom, Ukraine, and Tunisia (introduced)	(69/65)	63	0.99	0.0095	0.079	0.045 (0.02)	0.057 (0.03)	0.050	0.15 (0.20)	8.11
Lineage B	***Cypriniformes*** / *Rhodeus, Gobio, Pseudophoxinus, Barbus*	Algeria,Czech Republic, Ireland, Poland, Tunisia, Turkey	(18/7)	16	0.98	0.015	0.092	0.059 (0.02)	0.071 (0.03)	0.060	0.09 (0.15)	5.86
L. alternans	***Cypriniformes*** / *Hemiculter, Squalius*	China, Iran, and Russia	(6/3)	6	1	0.015	0.0064	0.0046 (0.02)	0.0057 (0.02)	0.0005	−0.25 (0.14)	5.81
China	***Osmeriformes*** / *Neosalanx*	China	(5/7)	4	0.90	0.0025	0.038	0.027 (0.004)	0.028 (0.001)	0.027	−0.01 (0.14)	6.73
Australia and New Zealand	***Galaxiiformes*** / *Galaxias* ***Gobiiiformes*** / *Gobiomorphus*	Australia and New Zealand	(11/6)	5	0.61	0.0196	0.0023	0.002 (0.02)	0.002 (0.04)	0.002	−0.08 (0.22)	5.80
L. pavlovskii	***Gobiiiformes*** / *Neogobius, Pomatoschistus*	Hungary, Germany, Poland, Ukraine	(10/17)	9	0.97	0.0029	0.104	0.043 (0.01)	0.089 (0.03)	0.044 (0.01)	0.49 (0.24)	9.8
East African Rift (EAR)	***Cypriniformes*** / *Engraulicypris, Rastrineobola*	Kenya and Tanzania	(15/7)	9	0.91	0.0047	0.019	0.0145 (0.02)	0.015 (0.03)	0.014 (0.02)	−0.005 (0.24)	6
Nearctic	***Cypriniformes*** / *Semotilus, Rhinichthys/ Couesius*	Canada, USA (Oregon)	(11/10)	8	0.92	0.0042	0.020	0.014 (0.003)	0.017 (0.001)	0.014	0.11 (0.25)	8.9
Central and South Africa (CSA)	***Cypriniformes*** / *Barbus, Enteromius*	Democratic Republic of the Congo, Namibia and South Africa	(9/5)	6	0.83	0.0055	0.011	0.007 (0.019)	0.008 (0.02)	0.007 (0.001)	0.02 (0.16)	5.80
Ethiopia	***Cypriniformes*** / *Labeobarbus, Enteromius*	Ethiopia (Lake Tana)	(6/12)	6	1	0.0185	0.030	0.015 (0.02)	0.022 (0.03)	0.013 (0.004)	0.27 (0.16)	12.8

each barcoded individual sample (Table S1a) and purification are outlined in supplementary material, section 2. We also generated a draft genome of *Ligula* using Illumina platforms (Omniseq), yielding an assembly comprising 49,218 scaffolds and covering 780 Mb of the genome; gene prediction and annotation for this assembly are ongoing.

2.4. ddRAD data assembly and SNP calling

FastQC (Andrews, 2010) was applied for an initial quality check. The process_radtags program in Stacks v.2.5.3 was used to demultiplex and filter the raw reads with low quality and uncalled bases (Rochette et al., 2019). Also, cut sites, barcodes and adaptors were removed from ddRAD outputs. Next, paired-end read alignment of each sample to the reference genome was performed using the default mode of Bowtie 2 (Langmead and Salzberg, 2012). SNPs calling and assembling loci were carried out from the alignment using the ref_map.pl wrapper program. Data were then filtered for missingness, minor allele frequency (MAF) and linkage disequilibrium (LD) in Stacks (see Supplementary Material, section 3 for details). In sum, we generated three SNP matrices in variant call format (VCF) comprising (1) 44,298 SNPs with a mean coverage per locus of ca. 20×, (2) second linkage disequilibrium filtered data set (0.2 squared coefficient of correlation) including 10,937 SNPs and a mean coverage per locus of ca. 18×, and (3) two outgroups which were added to the third data set comprising 4471 unlinked SNPs and a mean coverage per locus of ca. 25 ×.

2.5. Analysis of sequence diversity

Three gene fragments (Cyt *b*: 405 bp COI: 396 bp, and ND1: 893 bp) were added to an alignment that included 76 sequences of Cyt *b* and COI from Bouzid et al. (2008a, 2008b), Bouzid et al. (2013) and Bouzid et al. (2013) studies. All gene fragments were aligned separately using MUSCLE performed in MEGA v.5 (Tamura et al., 2011). We used the three gene fragments for phylogenetic analyses; however, all genetic distance-based analyses were carried out based on Cyt *b* and COI sequences which had no missingness. Analysis of population genetic diversity (sequence polymorphism, haplotype diversity and nucleotide diversity) was performed in DnaSP v.5 (Librado and Rozas, 2009) on samples grouped according to their phylogenetic relationships.

For nuclear SNP data, we estimated the expected heterozygosity (HE) and observed heterozygosity (HO), private alleles, and inbreeding coefficients among the parasite populations using the R package SambaR (de Jong et al., 2021). The first ddRAD data set (44,298 SNPs) was applied for the analysis as the calculation assumed that the input data were not reduced based on LD filtering. We used Plink v.1.90 (Chang et al., 2015) to convert the VCF file to the bed format. The default data filtering was applied with the "filterdata" function in the R package (see De Jong et al., 2021).

2.6. Phylogenetic analyses

The best-fitting substitution models and partition schemes were estimated for the three mtDNA genes using a greedy algorithm in PartitionFinder v.2.1.1 (Lanfear et al., 2017). Using Bayesian inference (BI) and maximum likelihood (ML), the phylogenetic relationships of the *Ligula* lineages were reconstructed. BI analysis was conducted using MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003) on the data partitioned based on codon positions and using the settings as described in supplementary material, section 4. We inferred the ML tree using IQTREE v.2.1.2 (Minh et al., 2020) under the best-fitting model from PartitionFinder (Table S3) and 1,000 ultrafast bootstrap replicates. We also used the "-bnni" option in IQTREE to minimize the risks of overestimating support values to a minimum. We included *Dibothriocephalus nihonkaiensis* and *D. latus* (Li et al., 2018a), sister groups of *Ligula*, as the outgroups.

We utilised the third data set (4471 unlinked SNPs) to reconstruct the phylogenetic gene tree and the species tree using three approaches. Using PartitionFinder, we first determined the parameters applied as input into RAxML-ng to reconstruct an ML tree (Kozlov et al., 2019). Second, the alignment was used for a Bayesian phylogenetic analysis. BI was completed using ExaBayes v.1.5 (Aberer et al., 2014) under a model partitioning scheme analogous to the ML analyses. Third, species trees were investigated using SVDquartets in PAUP v.4.0a147 (Chifman and Kubatko, 2014). Further details on the assessments of convergence and ML, BI and species trees are provided in supplementary material, section 4.

2.7. Population structure analyses

Using the concatenated COI and Cyt *b* (800 bp), haplotype networks were constructed to visualize relationships among the parasite populations via the Median-Joining algorithm in PopArt v.1.7 (Leigh and Bryant, 2015). Moreover, ddRAD data set was applied to construct a phylogenetic network using a neighbour-net algorithm and 1,000 bootstraps executed in SplitsTree v.4.10 (Huson and Bryant, 2006).

Several clustering analyses were performed to discover the number of distinct genetic groups. First, the model-based evolutionary clustering method was applied in ADMIXTURE (Alexander et al., 2009) via its parallel processing capabilities (AdmixPiPe: Mussmann et al., 2020). Second, a Principal Component Analysis (PCA) was performed using the glPCA function from the adegenet R package (Jombart and Collins, 2015). Next, we applied the TESS3R package in R to provide additional estimates of the role of geography in the genetic structure (Caye et al., 2018). TESS3R explores the implications of genetic diversity in natural populations using geographic and genotypic data simultaneously. Lastly, we applied the programs fineRADstructure and RADpianter v.0.2 (Malinsky et al., 2018) to reveal population genetic structure based on the nearest-neighbour haplotype (For details on ADMIXTURE and fineRADstructure analyses, see Supplementary Material, section 5).

2.8. Species delimitation analyses

To detect species boundaries in the *Ligula* species complex using the mtDNA data, species delimitation was performed using three independent approaches (genetic distance and tree-based approaches): the Generalized Mixed Yule Coalescent model (GMYC by Pons et al., 2006), Bayesian implementation of Poisson Tree Processes model (bPTP by Zhang et al., 2013), and Assemble Species by Automatic Partitioning (ASAP by Puillandre et al. 2021 Supplementary Material, section 6).

To make comparisons among species delimitation models using SNP data, we conducted Bayes Factor Delimitation (BFD, Leaché et al., 2014) implemented in SNAPP in BEAST2 (Bouckaert et al., 2014; Bryant et al., 2012). Comparing models by marginal likelihood scores, this approach assesses support for the alternative model using Bayes factors under the multispecies coalescent (MSC) model. We tested a nested set of hypotheses up to the maximum potential number of species, where every lineage was considered a different species (Supplementary Material, section 6).

2.9. Divergence time and phylogeographic structure

For mtDNA, divergence dates were estimated using a coalescent-based model implemented in BEAST v.1.8.2 (Drummond et al., 2012). Due to a lack of available fossil or geographic events, we used two common rates of substitution for the mt-genes (Cyt b = 0.0195, COI = 0.0225 substitutions/site/Myr), which were previously estimated for species of *Taenia* tapeworms (Cyclophyllidea). In addition, we applied the divergence time between *Spirometra* and *Diphyllobothrium* genera (at the base of Diphyllobothriidea), which is normally distributed with a mean of 11.47 Mya, a standard deviation of 1.14 Mya and a 95 % interval of 9.58–13.36 Mya. The best-fit partitioning schemes were assigned based on the Akaike information criterion (AICc) in PartitionFinder and the analysis was run using the settings reported in the supplementary material, section 7.

To infer the divergence time among major tapeworm lineages using SNP data, we first estimated the substitution rates for protein-coding sequences (CDS) of the *L. intestinalis* genome based on the methods suggested by Wang et al. (2015). Protein sequences of 10 flatworms

(*D. latus*, *Spirometra erinaceieuropaei, Taenia solium, T. asiatica, T. saginata, Echinococcus granulosus, E. multilocularis, Hymenolepis microstoma, Schistosoma japonicum, S. mansoni*) were obtained from WormBase ParaSite (https://wormbase.org/). After concatenating all CDS parts, they were aligned together using MAFFT v.7.4 ('L-INS-I' algorithm) (Katoh and Standley, 2013) and their gaps were discarded by Trimal v.1.4 (Capella-Gutiérrez et al., 2009). The best evolution models for the CDS alignments were calculated using PartitionFinder (GTR + I + G). ML trees were reconstructed with 400 bootstrapping replicates using RAxML under the best fit models. The substitution rate (branch-length) for *Ligula* was estimated to be 0.0201 which was 1.02-fold higher than that of *T. solium* (0.0196 mutations per site) (Fig. S1). Wang et al. (2015) calculated the mutation rate/site/year in the *T. solium* genome (2.82×10^{-9}). Therefore, we estimated 2.89×10^{-9} substitutions/site/year for the *Ligula* genome. We used Beast v1.8 with 141 sequences, including two *D. latus* outgroups to reconstruct the divergence dates of the sequences under a birth–death process as a tree prior and the uncorrelated relaxed molecular clock model for the mutation rate (2.89×10^{-9}). The analysis was run using the settings presented in the supplementary material, section 7.

2.10. Historical biogeography and ancestral host reconstructions

Inferring ancestral ranges and the historical biogeography of *Ligula* was done based on three models of biogeographical range extension, including Bayesian inference (BAYAREA-like), Dispersal-Vicariance (DIVA-like), and Dispersal-Extinction-Cladogenesis (DEC) models, all estimated by the BioGeoBEARS R package (Matzke, 2013). As BioGeoBEARS requires an ultrametric tree of species/population, we used the BEAST chronogram from both mtDNA and ddRAD analyses, and all outgroups and specimens were pruned using Mesquite v.3.7 (Maddison and Maddison, 2021).

These models enable the exploration of different possibilities of vicariance, extinction, and dispersal. Furthermore, by incorporating a founder-event parameter (+J) into the analysis, we attained clastogenic dispersal outside of the parental areas (referred to as jump speciation). The *Ligula* species complex was divided into five biogeographic realms: (1) Palearctic, (2) Afrotropical, (3) Indomalayan, (4) Australasian and (5) Nearctic. Inferred ancestral ranges were allowed to occupy up to five areas, thereby facilitating clear estimations. Likelihood-based measures of three biogeographic models along with their modified J types were ranked according to the AIC method to evaluate model quality.

We applied phylogenetic trees from both mtDNA and ddRAD data to reconstruct ancestral host fish affiliations using a Bayesian framework implemented in sMap v.1.0.7 (Bianchini and Sánchez-Baracaldo, 2021). The software uses stochastic mapping analysis to reconstruct the evolutionary history of discrete characters at the branch and node of the tree. In the present study, *Ligula* plerocercoids were collected from 41 fish species representing 18 genera in four orders (Cypriniformes, Osmeriformes, Gobiiformes and Galaxiiformes). Character states were defined based on the host taxonomy at the order level. Therefore, we defined four different states as follows: 1) lineage China infects Osmeriformes, 2) *L. pavlovskii* was associated with Gobiiformes, 3) lineages of Australia and New Zealand infect Gobiiformes and Galaxiiformes, and 4) lineages of Ethiopia, EAST African Rift (EAR) and lineages A and B were associated with Cypriniformes (Supplementary Material, section 8).

2.11. Genome-wide analysis of hybridization and introgression

Treemix v.1.12 (Pickrell and Pritchard, 2012) was applied to examine gene flow among populations. Utilizing allelic frequency data, the analysis generated an ML tree, followed by inferring the historical migratory events between populations. We assumed 10 migration events and the calculation of each migration event was done separately. We used the third ddRAD data set and grouped it into 11 populations based on phylogenetic analysis. The R package 'optM' (Fitak, 2021) was then used to optimize the number of migration events according to Evanno et al. (2005) (Supplementary Material, section 9). Visualization of the graph results was accomplished by the plot function (plotting_funcs.R) in R. We also employed the ABBA-BABA (D-statistics) and F statistics tests to further investigate patterns of gene flow among populations (Patterson et al., 2012). This analysis provides a basic yet comprehensive framework for deciphering deviation from a strictly bifurcating phylogeny. ABBA-BABA statistics were calculated in DSUITE v.0.4 (Malinsky et al., 2021) for 139 specimens from 11 lineages and two specimens from *D. latus* as outgroup. Patterson's D tests and f4 -ratio statistics are computed by DTRIOS, which provides information about trios of lineages with significant D values based on Z-score, calculated with a book-jackknife technique, taking linkage among sites into consideration. We left the option jknum (the number of jackknife blocks to divide the genome into) at its default value (20 blocks of 4471 SNPs), because the results were not affected by different tested values (20,30,40, 50). The resulting p-value was adjusted by applying the Benjamin-Hochberg method for multiple testing correction via false discovery rate control. To facilitate the evaluation of correlated f4- ratio statistics, the FBRANCH command implemented the calculation of f-branch statistics and its output was plotted with dtools.py. We conducted a sliding-windows analysis for the lineage trios with significantly elevated D by means of DINVESTIGATE command to determine whether the admixture signal was limited to particular contigs (Windows size = 2. Step = 1) (Malinsky et al., 2021).

2.12. Demographic history

The Extended Bayesian Skyline Plot (EBSP) (Heled and Drummond, 2008) was used to separately analyse the demographic changes for each lineage based on the Cyt *b* and COI data for inference of past population dynamics over time. EBSP analyses were implemented in BEAST v.2 under the GTR + G + I (invariant sites) model and a strict molecular clock. We employed the same mutation rate as in the divergence time analysis. We used the plotEBSP R function to draw the skyline plot (Heled, 2015). The *X*-axes indicate time reported in units of thousands of years before present (BP), while the *Y*-axes display mean effective population size (*Ne*) in millions of individuals divided by generation time plotted on a log scale. Moreover, the stairway plot2 (Liu and Fu, 2020) was used for estimating contemporary effective population sizes and changes in effective population size of the *Ligula* species complex over time. SambaR (R package) was used to convert the first ddRAD data file into folded SFS files (see details in Supplementary Material, section 10).

3. Results

3.1. Summary statistics

We compared the genetic diversity of all parasite populations based on the mtDNA (162 specimens) and ddRAD (139 specimens) data. The concatenated mtDNA matrix with 162 *Ligula* individuals and 800 bp aligned positions (404 bp Cyt *b* and 396 bp COI) was characterized by 551 monomorphic sites and 249 polymorphic sites (23 singletons and 226 parsimony informative sites). We also added 62 ND1 sequences (891 bp) to the concatenated data set for all phylogenetic analyses based on ML and BI methods. Estimates of nucleotide diversity for the two mitochondrial genes ranged from 0.0025 to 0.0196 and the haplotype diversity was between 0.61 and 1. The mean observed and expected heterozygosity ranged between 0.0005 and 0.044 and 0.0023–0.104, respectively. Additionally, populations from China, Australia and New Zealand, and Central and South Africa (CSA) demonstrated the lowest levels of haplotype diversity and observed and expected diversities (except *L. alternans*). The *Ligula* population from Ethiopia (12.8) showed the highest number of private alleles; however, populations in Lineage

B, Australia and New Zealand, *L. alternans* and CSA had the fewest private alleles (5.8) (Table 1).

3.2. Phylogenetic relationships and species trees of Ligula spp

BI and ML analyses of the three mitochondrial genes (1691 bp) generated consistent trees with a considerably well-supported phylogenetic structure (Fig. 2A). Similarly, the reconstructed phylogenies based on ddRAD data (4471 SNP) showed identical topologies in the BI and ML methods (Fig. 2B). The phylogenetic structure revealed maximum support for all lineages in both mtDNA and ddRAD data sets. In all reconstructions, the *Ligula* population from Ethiopia was strongly supported by high posterior probability (pp = 1) and bootstrap support (BS = 100) as the basal lineage to all the tapeworms. Parasite populations from the Nearctic separated from other lineages with maximal statistical robustness. All phylogenetic trees from mtDNA and ddRAD data disclosed a sister group relationship between populations from EAR (Tanzania and Kenya) and CSA populations. The phylogenetic relationships of the lineages derived from mtDNA data were in agreement with the phylogeny of the ddRAD data (except China and *L. pavlovskii* lineages). Within mtDNA, *L. pavlovskii* diverged from Lineage A and *L. alternans* with high statistical support. However, *L. pavlovskii* was distinctly separated from Lineage A, China, *L. alternans*, Lineage B, Australia and New Zealand. To sum up, our results added five new distinct lineages (*L. pavlovskii*, Australia, New Zealand, EAR and CSA) to the six previously defined *Ligula* lineages (Bouzid et al., 2008a, 2008b; Štefka et al., 2009). The SVDquartet species trees recovered the same topology as the ML and BI analyses but with variable posterior probability and bootstrap values (Fig. 2).

3.3. Population genetic structure and host usage

A total of 133 haplotypes were found based on the Cyt *b* and COI sequences in 162 individuals. The mitochondrial haplotype network based on TCS analysis with a 95 % parsimony connection limit revealed the same 11 distinct haplogroups as in the phylogenetic analysis (Fig. 3A, connection limit = 12 mutations). Individual haplogroups showed a high level of host specificity. Although many lineages were retrieved from Cypriniformes, the genera (and families) of the fish rarely overlapped between lineages even in areas of sympatry (such as Europe). Haplogroup A is widely distributed in Europe, Iran and Russia and included *Ligula* populations from Cypriniformes (*A. brama, R. rutilus, B. bjoerkna, S. erythrophthalmus, S. cephalus, A. alburnus, P. phoxinus* and *C. carassius*). The *L. alternans* haplogroup was connected to haplogroup A with 39 mutational steps and comprised parasite populations from Iran, China, and far east Russia, which were associated with *Hemiculter* (Cypriniformes: Xenocyprididae) and *Neosalanx* (Osmeriformes: Salangidae). Haplogroup B was linked to Australia and China with 30 and 36 mutational steps, respectively. Furthermore, haplotypes from Australia and New Zealand diverged from each other with 41 mutational steps and included parasite populations from Galaxiiformes (*Galaxias*) and Gobiiformes (*Gobiomorphus*). Lineage B and the haplogroup in China were found in Cypriniformes (*Rhodeus, Gobio* and *Pseudophoxinus*) and Osmeriformes (*Neosalanx*), respectively. The haplogroup in EAR comprised parasite populations of *E. sardella* that diverged from the haplogroup in CSA (*E. anoplus* and *E. paludinosus*) by 22 mutational steps. The *Ligula* populations from Nearctic Cypriniformes (*S. atromaculatus* from Canada and *R. osculus* from Oregon) split from the other lineages by 55 and 120 mutational steps. Furthermore, haplogroups of *L. pavlovskii* from Gobiiformes (*N. fluviatilis, N. melanostomus, A. fluviatilis, P. minus* and *P. microps*) and Ethiopia from Cypriniformes (*L. intermedius, L. tsanenis, L. brevicephalus* and *E. humilis*) separated from each other by 110 mutational steps (Fig. 3A).

Contrary to the mitochondrial haplotype network, the phylogenetic network based on ddRAD data clustered Australia and New Zealand together, revealing 10 well-supported clusters for the *Ligula* species complex (Fig. 3B). Ethiopia was the first lineage branching at the base. EAR and CSA split from each other with 98 % bootstrap support value,

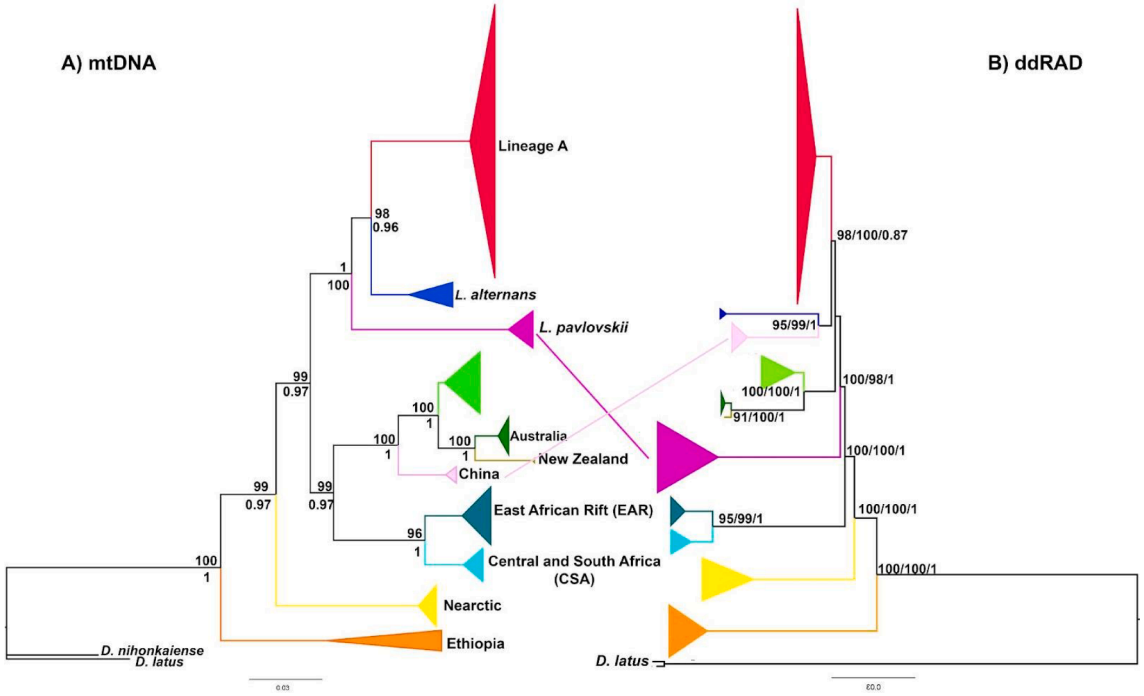


Fig. 2. Phylogenetic relationships of the *Ligula* species complex. A) Bayesian tree reconstructed using the concatenated mitochondrial genes (Cyt *b*, COI and ND1) using two outgroups (*D. latus* and *D. nihonkaiensis*). Lineage positions and branching patterns are concordant with the ML tree. Nodal supports at each node represent support values of BI (above branches) and ML (below branches). B) Reconstructed phylogenetic tree based on ddRAD. Bootstrap support values and Bayesian posterior probabilities are given at the nodes (RAxML/svdquartets/ExaBayes). Diagonal lines represent incongruences between the mtDNA and ddRAD topologies. Topology congruence tests demonstrated significant differences between the mtDNA and ddRAD trees ($P < 0.05$).

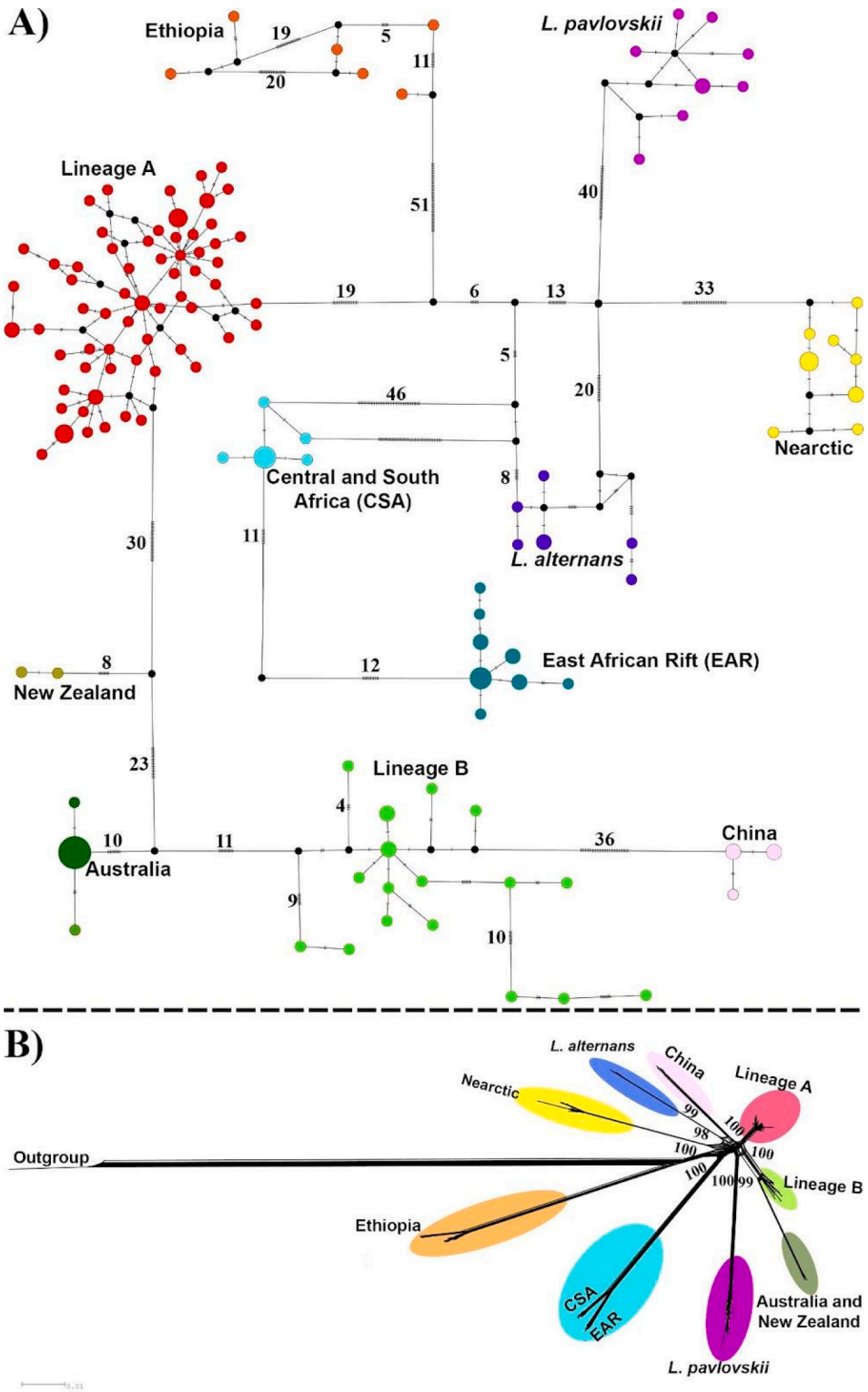


Fig. 3. Population relationships. A) Median-joining network showing relationships among the *Ligula* species complex, reconstructed based on the concatenated Cyt *b* and COI sequences (801 bp). Circle sizes are proportional to haplotype frequencies. Black circles indicate unsampled or extinct haplotypes. Mutational steps are displayed by numbers and dashed symbols along each line. B) Phylogenetic network with bootstrap support values for the ddRAD data showing 10 distinct clusters in the *Ligula* species complex. *D. latus* was used as an outgroup.

diverging from *L. pavlovskii* and Ethiopia with 100 % bootstrap support value. Similarly, Australia and New Zealand differed from Lineage B. Also, Lineage A demonstrated a divergence from all parasite populations with strong bootstrap value (100 %).

Population structure analyses showed the three new clusters (*L. pavlovskii*, EAR, CSA) were clearly separated from the five previously identified lineages (Lineage A, *L. alternans*, Australia, Lineage B and China) (Fig. 4). Cross-validation of the admixture analysis revealed an optimal number of K = 10 clusters (Fig S2). All lineages were separated from each other, although some admixture was observed between EAR and CSA (Fig. 4A). PCA patterns were mainly analogous to the clustering recognised by admixture. The first two PCs represented 22.6 % and 18.1 % of the total genetic variation, separated into 10 groups. EAR was close to CSA and had an overlap on both PC1 and PC2 axes. The third PC (16 %) clearly discriminated Lineage A, Lineage B and Ethiopia from the remaining groups (Fig. 4B). Similarly, the cross-entropy curve of the

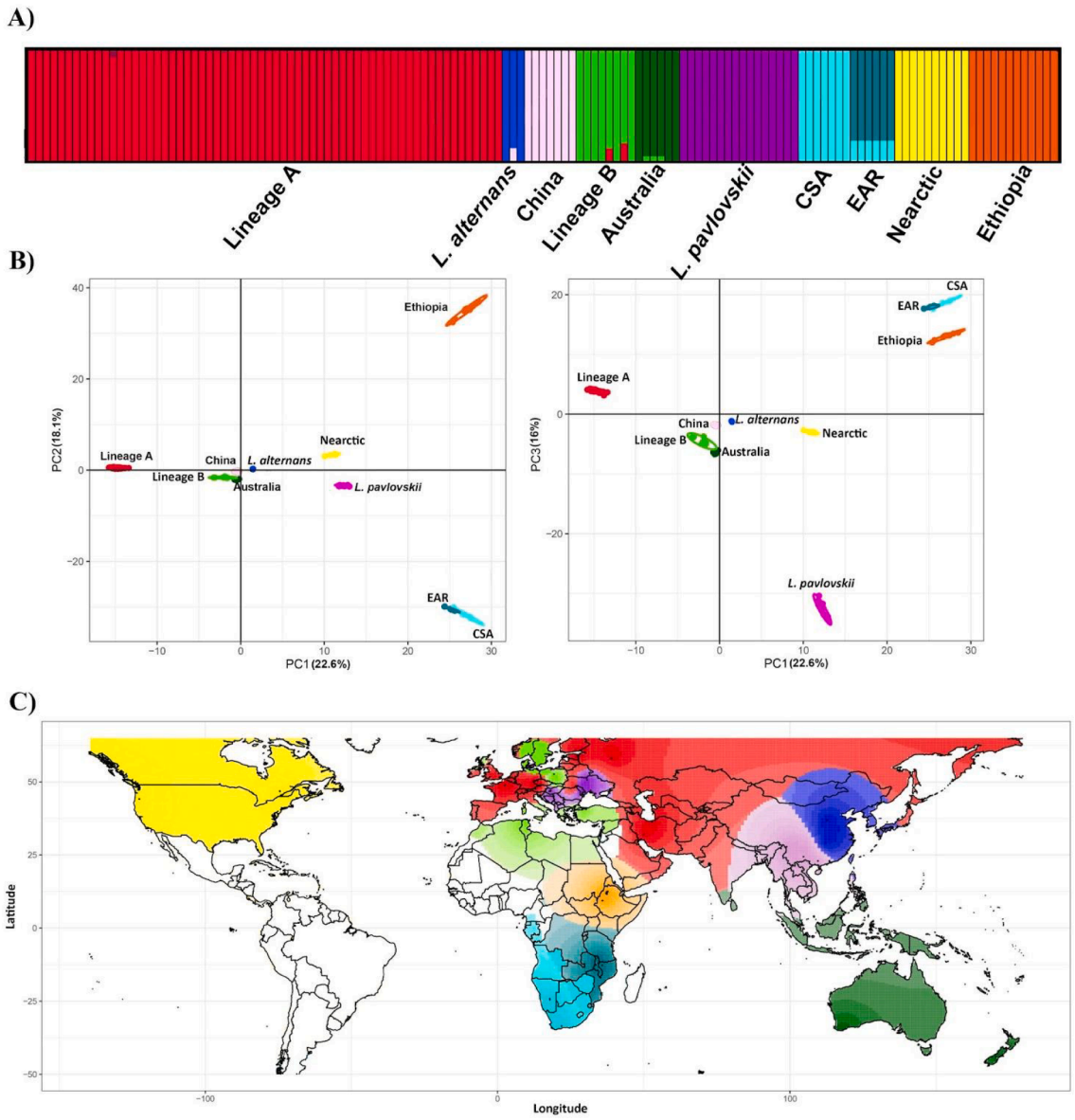


Fig. 4. Genetic structure of the populations. A) Population structure based on the ADMIXTURE analysis. Results are shown at K = 10. Each vertical (100 %) stacked column indicates an individual representing the proportions of ancestry in K-constructed ancestral populations. B) PCA of *Ligula* sp. samples with 10,937 SNPs, where the first three components explain 22.6 %, 18.6 % and 16 % of the variance, respectively. C) Interpolated spatial genetic structure according to the individual ancestry coefficients of TESS3R. The colour gradient shows the degree of difference among individuals.

spatial population genetic structure recovered a likely number of K = 10 genetic clusters (Fig. S3). We spatially interpolated the ancestry coefficients projected onto a map to describe the global genetic structure in our data (Fig. 4C). The resulting spatial pattern revealed 10 components of ancestry in the sampling space.

The analysis of shared ancestry in the fineRADstructure confirmed the results of the cluster analysis with K = 10, with significant support in the dendrogram. Although the recent co-ancestry was greatest between EAR and CSA, the two lineages formed distinct groups in the analysis. Moreover, a single sample from New Zealand showed a higher level of shared genetic background with the cluster from Australia. These two clusters had relatively higher co-ancestry with Lineage B and China. Moreover, lineages A and B exhibited a moderate common ancestry with each other, whereas the lowest level of co-ancestry sharing was discovered among Ethiopia, *L. pavlovskii*, Nearctic, EAR and CSA. This is consistent with the admixture results where a level of admixed population was observed between EAR and CSA as well as Australia and Lineage B (Fig. 5).

3.4. Species delimitation analyses

The species delimitation methods (ASAP, GMYC, bPTP and BFD) produced estimates that ranged from 10 to 15 species (GMYC and bPTP) (Fig. 6). ASAP analysis with the best score delimited 10 and 12 species based on Cyt *b* (lowest score = 2.5), COI (lowest score = 2.5) and ND1 (lowest score = 4). In the ND1 fragment, out of the 10 top-scoring models, three showed distance thresholds of 5.7 %, 1.4 % and 2.7 % with seven, 13 and 12 species. The third was selected as the most likely initial estimate of species diversity. Tree-based analyses identified the same composition and number of species. The results of the BFD hypothesis testing are shown in Fig. S4 and Fig. 6. The ddRAD data set favoured 10 candidate species based on marginal likelihood (3820.21, Fig. S4) and BF (32.5). Overall, all delimitation analyses determined Lineage A, *L. alternans*, *L. pavlovskii*, China and Nearctic as different species. However, GMYC and bPTP over-split Lineage B and Ethiopia into three candidate species. In addition, all the methods (except ASAP based on Cyt *b* and BFD) distinguished a species boundary between

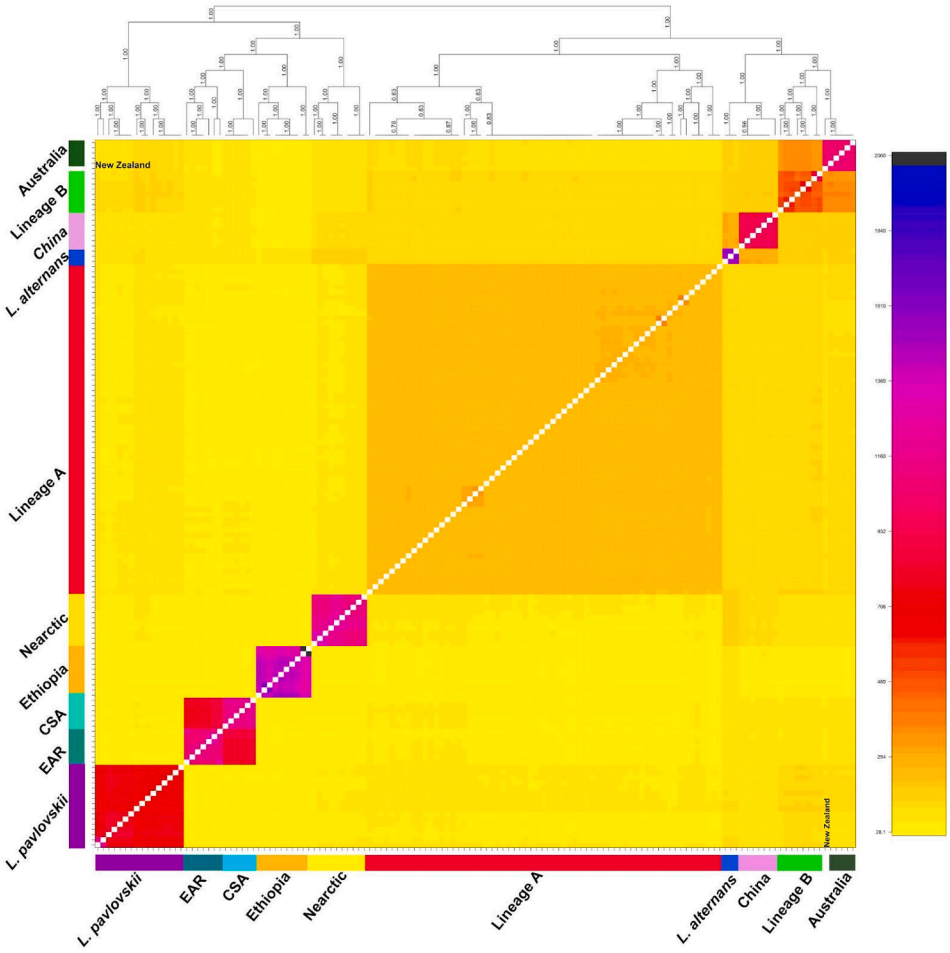


Fig. 5. Pairwise genetic similarity among individuals indicated by FineRADStructure co-ancestry matrix. Inferred populations are clustered together in their accompanying dendrogram.

Australia and New Zealand (Fig. 6).

3.5. Divergence time and phylogeographic structure

Estimate of the divergence times using mitochondrial genes was consistent with the time-calibrated tree of ddRAD, strongly supporting the basal cladogenesis between lineages of the *Ligula* species complex in late Miocene and early Pliocene (mtDNA = 4.99 Mya, 95 % highest posterior density intervals (HPD): 4.76–5.27 Mya and ddRAD = 5.05 Mya, HPD: 4.56–5.49), with the Ethiopian population emerging as the basal lineage to all *Ligula* populations (Fig. 7). Subsequent cladogenetic events dating to mid-Pliocene at approximately 3.35–4.01 Mya (mtDNA = 4.01 Mya, HPD = 4.87–3.10 Mya and ddRAD = 3.35 Mya, HPD = 3.04–3.78 Mya) isolated the ancestor of Nearctic lineage from other populations. Both mtDNA and ddRAD results revealed a sister relationship between EAR and CSA which diverged from each other in early Pleistocene at 1.01–1.14 Mya (mtDNA = 1.14 Mya, HPD: 0.71–1.41 Mya and ddRAD = 1.01 Mya, HPD: 0.66–1.29 Mya). Furthermore, both dated trees strongly supported a sister relationship between Australia and New Zealand which diverged from Lineage B at ~1.12–1.23 Mya (mtDNA = 1.12 Mya, 95 % HPD: 0.78–1.81 Mya and ddRAD = 1.23 Mya, HPD: 1.11–1.49 Mya). In mtDNA, *L. pavlovskii* diverged from *L. alternans* and Lineage A at ~2.41 Mya (Fig. 7A, HPD: 0.78–2.96 Mya); however, ddRAD revealed that *L. pavlovskii* split from the other six lineages at ~2.48 Mya (Fig. 7B, HPD: 2.29–2.88 Mya). Furthermore, in mtDNA, China diverged from Australia, New Zealand and Lineage B at ~1.71 Mya (Fig. 7A, HPD: 1.11–2.17 Mya), while in ddRAD, China formed a sister relationship to *L. alternans*, diverging at ~1.52 Mya (Fig. 7B, HPD: 1.09–1.85 Mya).

3.6. Historical biogeography and gene flow analyses

Ancestral range estimation analyses were applied to chronograms resulting from both the mtDNA and ddRAD data sets (Fig. 8A). A comparison of the six biogeographic models with the likelihood ratio test and AIC indicated that DIVALIKE + J was the top fitting model for both the mtDNA and ddRAD data sets (Tables S4). The preferred hypothesis would be that the common ancestor was widespread throughout the Palearctic, Afrotropical and Nearctic biogeographic realms between the late Miocene and early Pliocene. Moreover, vicariance took place between Nearctic and Palearctic in mid-Pliocene. With respect to anagenetic or gradual speciation processes, our parameters indicate that range contraction made an equal contribution as did range expansion in both data sets (Table S4). The ancestor of EAR and CSA emerged through a rare jump dispersal event from the Palearctic to the Afrotropical in late Pliocene. In addition, the China lineage was formed as a result of founder event speciation between Palearctic and Indomalaya in mid-Pleistocene. Finally, Australia and New Zealand were genetically isolated from their ancestor populations in the Palearctic (Fig. 8A).

Of the three-character models evaluated using sMap, the ER model was the best supported (BIC, Table S5) for both mtDNA and ddRAD data sets. The Bayesian stochastic character mapping of ancestral host fish associations demonstrated an ancient affiliation with cypriniforms at the root of all *Ligula* lineages (Fig. 8B). In both data sets, mean marginal posterior estimates showed that cypriniforms contributed as the most recent common ancestral host for the parasite with 81.4 % and 82.3 % in

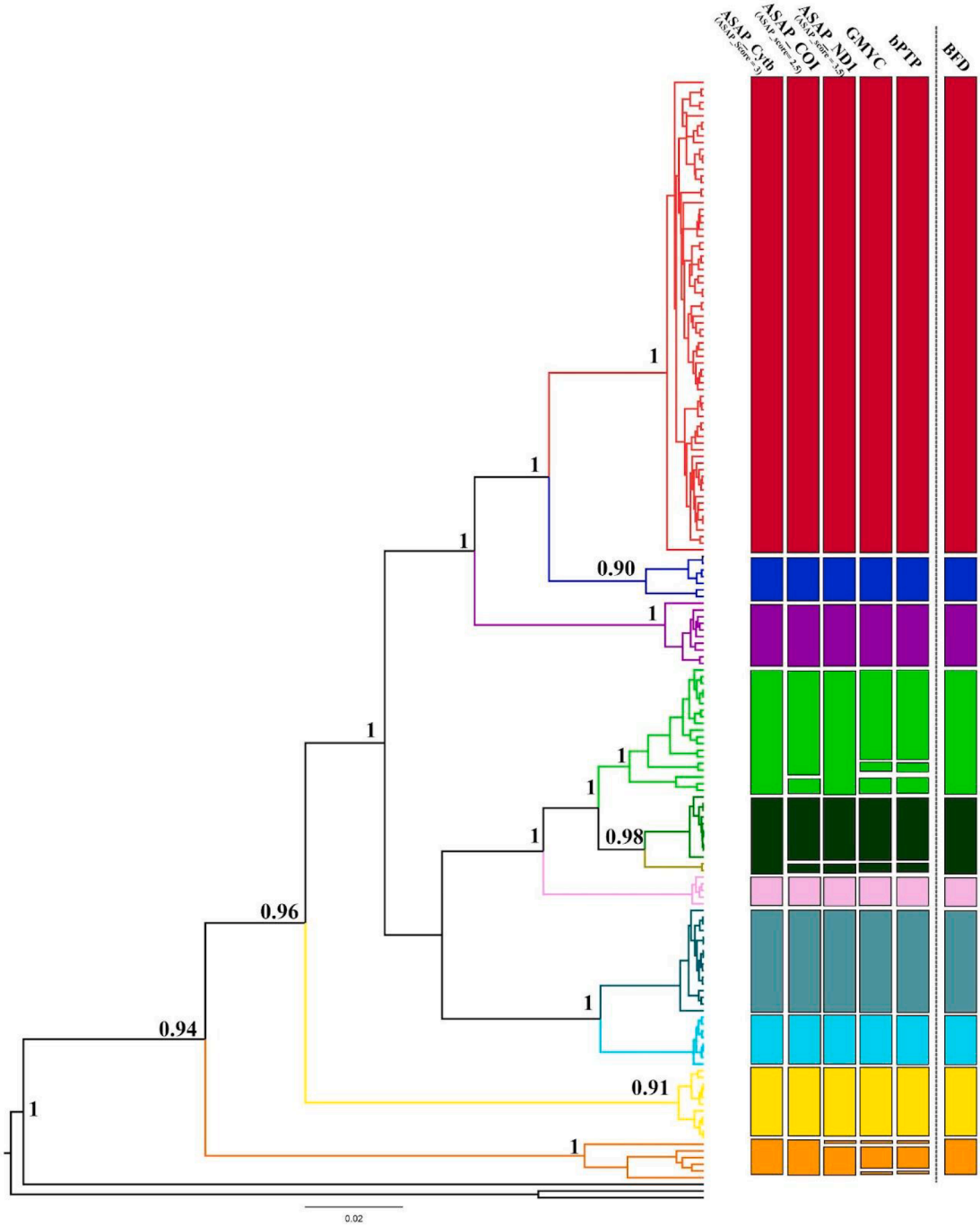


Fig. 6. Multi-approach species delimitation of *Ligula* species complex. Results of the mtDNA method (ASAP, GMYC, bPTP) are compared to the results of genome-wide SNP methods (BFD). Coloured bars depict results from different methods, indicating congruence among them.

mtDNA and ddRAD, respectively. Also, the most recent common ancestor (MRCA) of *Ligula* had the highest host switching from cypriniforms to the other fish orders. Furthermore, *Ligula*'s MRCA had the second highest average time spent (mtDNA = 9.1 %, ddRAD = 12.1 %) in Gobiiformes and Galaxiiformes, and the lowest average time spent in Osmeriformes (Table S4, Supporting Material, mtDNA = 5.5 % and ddRAD = 6.1 %). A host shift from cypriniforms towards Gobiiformes and Galaxiiformes was discovered in both data sets for the MRCA of Lineage B, Australia and New Zealand. In addition, we identified the occurrence of several host shifts for the MRCA of China and *L. alternans*. On the other hand, the mtDNA chronogram revealed the MRCA of *L. pavlovskii* experienced three host shifts from Cypriniformes to Gobiiformes and Osmeriformes, while ddRAD stochastic map detected a host shift between Cypriniformes and Gobiiformes (Fig. 8B).

3.7. Genome-wide hybridization and introgression

The estimated species tree from Treemix recovered similar topology (Fig. 9A) and relative branch lengths to the phylogenetic trees. The simplest model that described the data was the model that assumed three migrations. No significant migration events were retrieved by more complex models (Fig. S5). The results revealed gene flow from Nearctic into *L. alternans* and China populations. Moreover, two gene flow events were detected from Lineage B into China and *L. pavlovskii*. This finding shows how introgression affects phylogenetic relationships in closely related species. To determine possible introgression events, we

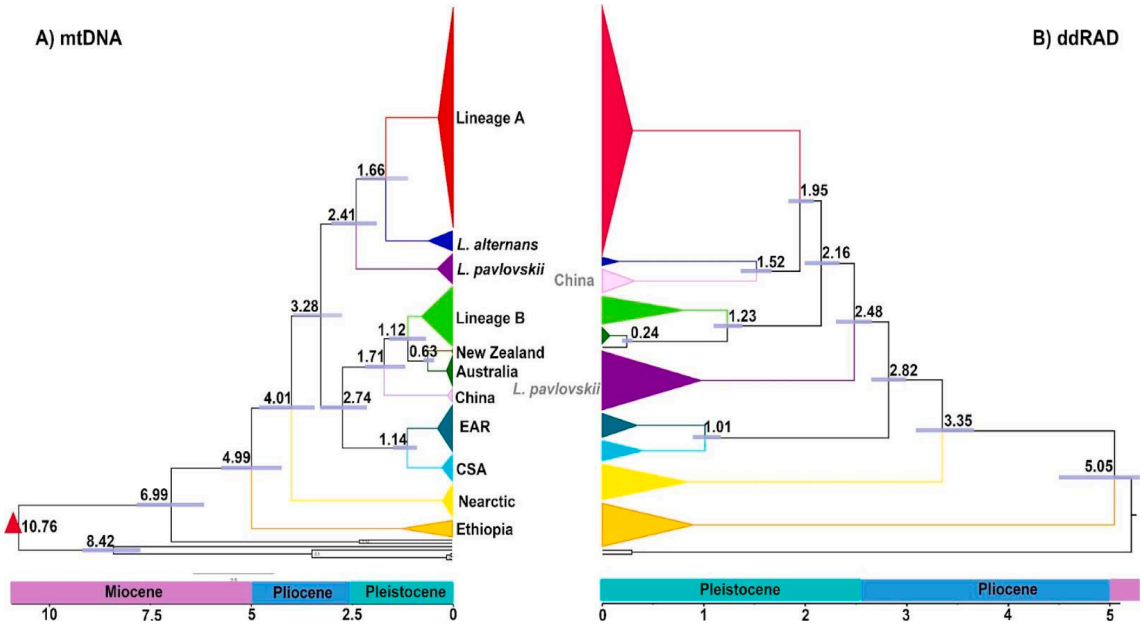


Fig. 7. Chronograms obtained from the dating analyses. A) Concatenated data set of three mitochondrial genes (Cyt *b*, COI, and ND1) containing 165 *Ligula* sequences. Numbers on each branch indicate divergence times. Calibration point is displayed by a red triangle. B) Divergence time estimated from the ddRAD data set (4471 unlinked SNPs). The bars on the nodes represent 95% HPD for divergence times. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

performed ABBA-BABA tests for 10 trios of closely related phylogroups. D-statistics for 47 of the trios were significantly different from zero ($Z > 3$, p-adjusted < 0.000919) and the values ranged between 0.244 and 0.492 (Table S6). Based on the sliding window analysis, SNPs demonstrating the ABBA-BABA patterns for each trio were distributed across 25–29 different contigs (Table S3). Nevertheless, it is possible that a single gene flow event leads to multiple elevated f4-ratio and D-statistics results. When we calculated a branch-specific statistic fb(C) that partially disentangled the correlated f4-ratio results, we found that only two out of the eight branches exhibited significant excess sharing of derived alleles at least with one other lineage (Fig. 10B). An internal branch that represented Lineage A and China exhibited an excess sharing of derived alleles with Australia, New Zealand, *L. pavlovskii* and Lineage B. Another fb(C) signal involves an internal branch as the common ancestor of Lineage B with *L. pavlovskii*, EAR and CSA (Fig. 9B). These results could also be attributed to incomplete lineage sorting. It must be noted that even a single introgression event may result in significant fb(C) values in multiple related phylogroups (Malinsky et al., 2018).

3.8. Historical and contemporary demographics

The demographic history based on two mtDNA genes (Cyt *b* and COI) and ddRAD data sets is presented for nine parasite lineages in Fig. 10. Two lineages (*L. alternans* and New Zealand) were not plotted due to their low effective sample sizes (ESS ⟨200) and low sample sizes. EBSP based on the mtDNA suggested that demographic expansion took place at approximately 100 kya (thousand years ago) for lineages A and B (Fig. 10 A and B). Moreover, the stairway plot based on the ddRAD data showed a clear sign of ancestral bottleneck between 1.5 kya and 2 kya followed by a range expansion for these lineages. China showed a population expansion after experiencing a bottleneck between 4 kya and 8 kya. Australia demonstrated a trend indicative of recent population shrinkage. Our mtDNA results uncovered a rapid population expansion of approximately 100 kya for *L. pavlovskii*; however, ddRAD revealed a bottleneck at ~ 50 kya with a decreased pattern from 6 kya onwards. A stable population size is indicated for EAR and CSA during the last 3 kya. In addition, a recent population reduction and a strong signal of demographic expansion were detected for the Nearctic and Ethiopian lineages.

4. Discussion

Host specificity, an inherent feature of parasitic and symbiotic organisms, can be a product of co-diversification (Engelstädter and Fortuna, 2019), but it is often the result of a more convoluted evolutionary history including host switches and dispersal events (Hoberg and Brooks, 2008; Matthews et al., 2022). The latter should be peculiar to parasites where the host lineages are much older than the parasite lineages, such as the *L. intestinalis* species complex. Using our example of a speciating parasite lineage with a global distribution, we demonstrated how host specificity of individual *Ligula* lineages was formed by different historical events involving long-distance dispersals, host switches and occasional introgression. Our analyses based on a dense SNP matrix and mtDNA data provided a clear picture of the cryptic diversity within the *L. intestinalis* species complex and a robust time-calibrated phylogeography. By analysing biogeographic and demographic patterns of individual lineages, we revealed patterns of the parasite's speciation and dispersal throughout its distribution range. Below, we discuss the results focusing on individual factors contributing to the parasite's diversification.

4.1. Phylogenetic relationships and occasional mito-nuclear discordance

Ligula phylogenies based on ddRAD and mtDNA data revealed 11 distinct lineages and showed a generally well-supported phylogenetic structure for the *Ligula* species complex. We characterized several new lineages from previously unexplored areas (CSA, EAR) or lineages that were not analysed in the global context before (*L. pavlovskii,* New Zealand; see Lagrue et al., 2018; Vitál et al., 2021). In addition, several differences in the phylogeny of the group were recovered compared to previous studies based on fewer loci and less material (Bouzid et al., 2008a, 2008b; Štefka et al., 2009). Ethiopia formed a basal lineage to other populations in our analysis, contrary to the findings of the previous study (Bouzid et al., 2008a, 2008b) where Nearctic samples (from Canada and Mexico) were at the basal position. Our data set also allowed

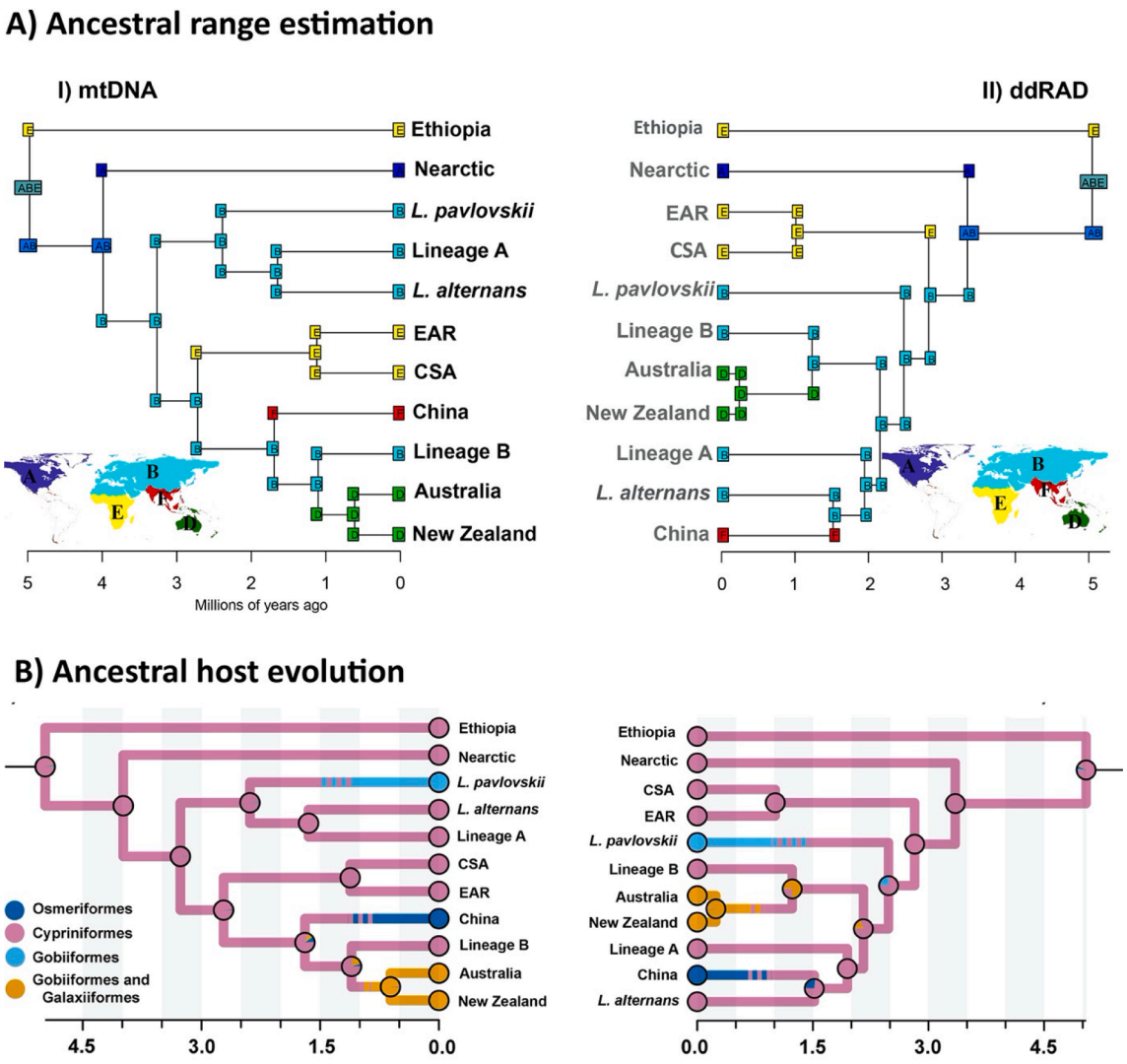


Fig. 8. Ancestral range and host estimations. A) Ancestral range estimation for the tapeworm chronograms in BioGeoBEARS under the DIVALIKE + J model. Squares at nodes indicate the highest ML probability in the area prior to the sudden speciation event, while squares on branches refer to the states of descendant lineages immediately after the speciation event. Squares showing more than one letter represent ancestral areas with more than one biogeographical area. B) Bayesian stochastic character map displaying the ancestral host evolution based on mtDNA (I) and ddRAD (II). Posterior probabilities for each node in different states are depicted by pie charts.

for a more fine scale distinction of the lineages previously clustered with Lineage B (Bouzid et al., 2008a, 2008b).

Only two cases of discordance between mitochondrial and nuclear-ddRAD phylogenies were found, with *L. pavlovskii* and China showing different topologies in each data set. To exclude a possibility that the incongruence is only due to the lack of signal in the smaller mtDNA data set, we performed a congruence test between mtDNA and ddRAD topologies and found the topology differences were significant (AU and SH tests). Similar mito-nuclear phylogenetic discordances have also been reported in several other metazoan taxa (e.g. Bouzid et al., 2008a, 2008b; Shaw, 2002; Weigand et al., 2017), often credited to the molecular signature of selection (Morales et al., 2015), incomplete lineage sorting (Degnan and Rosenberg, 2009) or introgressive hybridization (Bisconti et al., 2018). Our analyses of historical gene flow indicated past introgression events between multiple lineages (see the discussion below). Thus, we suggest introgressions as the mechanism responsible for the emergence of mito-nuclear discordances for China and *L. pavlovskii* lineages.

4.2. Species delimitation reveals extensive cryptic diversity in *L. Intestinalis* s.L. Populations

The taxonomy of *Ligula* tapeworms has always been complicated due to the general uniformity of their plerocercoids and adults (Dubinina, 1980), resulting in a lack of suitable morphological features to determine species boundaries using traditional methods, except for an already known trait, the duplicated reproductive complexes in *L. alternans* (Luo et al., 2003). Indeed, numerous species of *Ligula* have been proposed to accommodate plerocercoids from different fish hosts (see Dubinina, 1980 for their extensive list), but almost all these taxa have been synonymised with *L. intestinalis*. This species was then considered as a euryxenous parasite of cyprinoid fishes, maturing in several fish-eating birds (Kuchta and Scholz, 2017; Li et al., 2018a, 2018b; Petkevičiūtė, 1992).

Here, we used a range of species delimitation analyses to identify the putative species in the *Ligula* complex. The results based on ddRAD data revealed a slightly lower number of candidate species than the non-recombining mtDNA. Our findings also indicated that although different methods of species delimitation produced highly consistent species boundaries, they were not entirely similar. Carstens et al. (2013)

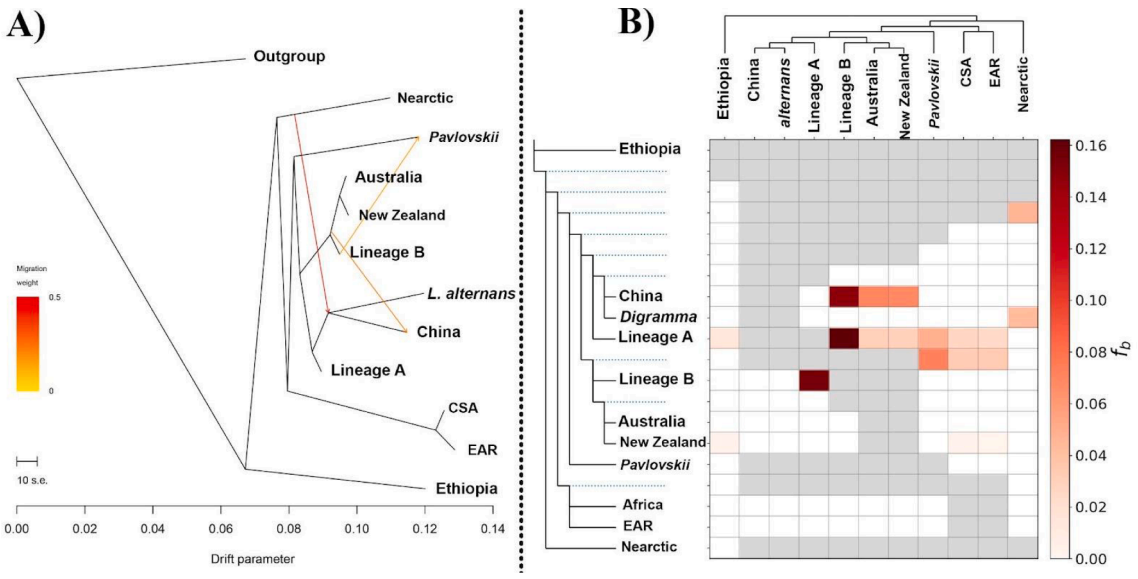


Fig. 9. Patterns of historical gene flow A) ML tree from Treemix, representing the relationships between the 11 *Ligula* lineages. The scale bar displays 10 times the average standard error for entries estimated in the covariance matrix. Migration arrows (events) are coloured based on their weight. The fraction of ancestry from the migration edge is referred to as migration weight. B) Results of genome-wide *f*-branch among the 11 *Ligula* lineages. Columns represent tips in the tree topology and rows indicate tree nodes. Cells display fb statistics between nodes (rows) and tips (columns). When comparisons cannot be made, the cells are left empty and grey. F-branch (fb(C)) statistics in our data suggest excess allele sharing among tips of the phylogenetic tree (representing lineages which are arranged horizontally at the top) and between each tip and node in the tree (arranged vertically on the left).

argued that such incongruence either signifies a difference in the power of one or more of the methods to detect cryptic lineages or that assumptions of the species delimitation approach have been violated. They suggested that species delimitation inferences should be drawn cautiously in order to avoid delimiting entities that fail to accurately represent evolutionary lineages. Therefore, taking the conservative approach, we identified at least 10 putative species for the *Ligula* species complex as follows: (1) *L. intestinalis* Lineage A; (2) *L. alternans* (syn. *Digramma alternans*); (3) *L. intestinalis* Lineage B; (4) *Ligula* sp. 1 from Australia and New Zealand (possibly-two species lineages); (5) *L. pavlovskii*; (6) *Ligula* sp. 2 from China; (7) *Ligula* sp. 3 from EAR; (8) *Ligula* sp. 4 from CSA (9); *Ligula* sp. 5 from the Nearctic; and (10) *Ligula* sp. 6 from Ethiopia. In subsequent analyses, we contrasted these evolutionarily significant lineages with their host and geographic spectra and demographic histories.

4.3. Dated phylogeography reveals historical dispersal and vicariance due to climatic oscillations and host switching

Climatic fluctuations have shaped the patterns of divergence in a multitude of species (Hewitt, 2001). In particular, quaternary glaciation cycles and climatic oscillations had a major impact on both free-living and parasitic species, leading to the extinction of several parasites (Hoberg et al., 2012; Li et al., 2018a) and influencing their evolution and distribution (e.g. Galbreath et al., 2020; Nazarizadeh et al., 2022b). Based on our mtDNA and ddRAD molecular dating, the ancestor of *Ligula* is estimated to have appeared between the late Miocene and early Pliocene (Fig. 7), and was found to parasitise one of the early diverging cyprinid species (*Barbus*, of the subfamily Barbinae). The divergence time suggests that the parasite complex did not follow the evolution of its intermediate host as the divergence of Cypriniformes had already occurred during the late Jurassic (154 Mya) (Tao et al., 2019). Moreover, the historical biogeography model revealed that the divergence of parasite populations from Afrotropical, Nearctic and Palearctic realms was recovered as two vicariant events (Fig. 8). This allopatric speciation could be associated with the isolation of migratory waterbirds due to the biogeographical realms becoming isolated by major barriers. According to our results, the first vicariance occurred between Afrotropical, Palearctic and Nearctic regions at approximately 5 Mya (Fig. 8) when desertification increased and the decline of global temperatures accelerated, resulting in the development of the Northern Hemisphere ice sheets and the onset of the ice ages. Moreover, the considerable expansion of the Sahara Desert had by then impeded *trans*-Saharan migration and long-distance dispersal for several migratory waterbirds (Bruderer and Salewski, 2008).

Ligula plerocercoids mature very quickly into adults in bird guts following ingestion of an infected fish (Loot et al., 2001). The adult phase lasts up to five days (Dubinina, 1980), during which the eggs can be dispersed. Consequently, a lack of suitable freshwater lake habitats within a few days flight of the final host represents a strong barrier against gene flow of the parasite, as documented by the lack of gene flow in Lineage A between Europe and a recently introduced population in North Africa (Bouzid et al., 2013; Štefka et al., 2009). Thus, aridification of the environment accompanied by reduced or ceased migration of the definitive host likely caused the major vicariance mechanisms in the evolution of *Ligula*.

The second vicariant event took place between the Nearctic and Palearctic during the late-Pliocene when the Nearctic lineage diverged from the other populations (Fig. 8). This lineage was recovered from the Nearctic leuciscids of two subfamilies (*Semotilus* - Plagopterinae, and *Rhinichthys* - Pogonichthyinae) which are thought to have originated in ancient Europe and dispersed to North America during the Late Cretaceous to Paleocene (see Imoto et al., 2013). In contrast to the intermediate hosts which had evolved a long time before *Ligula*'s speciation, changes in the behaviour of the definitive hosts could have accelerated parasite's diversification from Pliocene to Pleistocene (Hirase et al., 2016). It has been proposed that glaciations possibly shifted the migratory routes and migratory behaviour of several North American bird species (Zink and Gardner, 2017). In addition, ecological competition and physical isolating barriers between the Palearctic and Nearctic during the late Pliocene or early Pleistocene were determining factors (Naughton, 2003). On the other hand, the results of TreeMix showed historical gene flow between the population from the Nearctic and the MRCA of lineages of China and *L. alternans* (Fig. 9A). This introgression likely occurred after the second vicariant event when both the Nearctic and Palearctic regions were connected through the Bering Land Bridge,

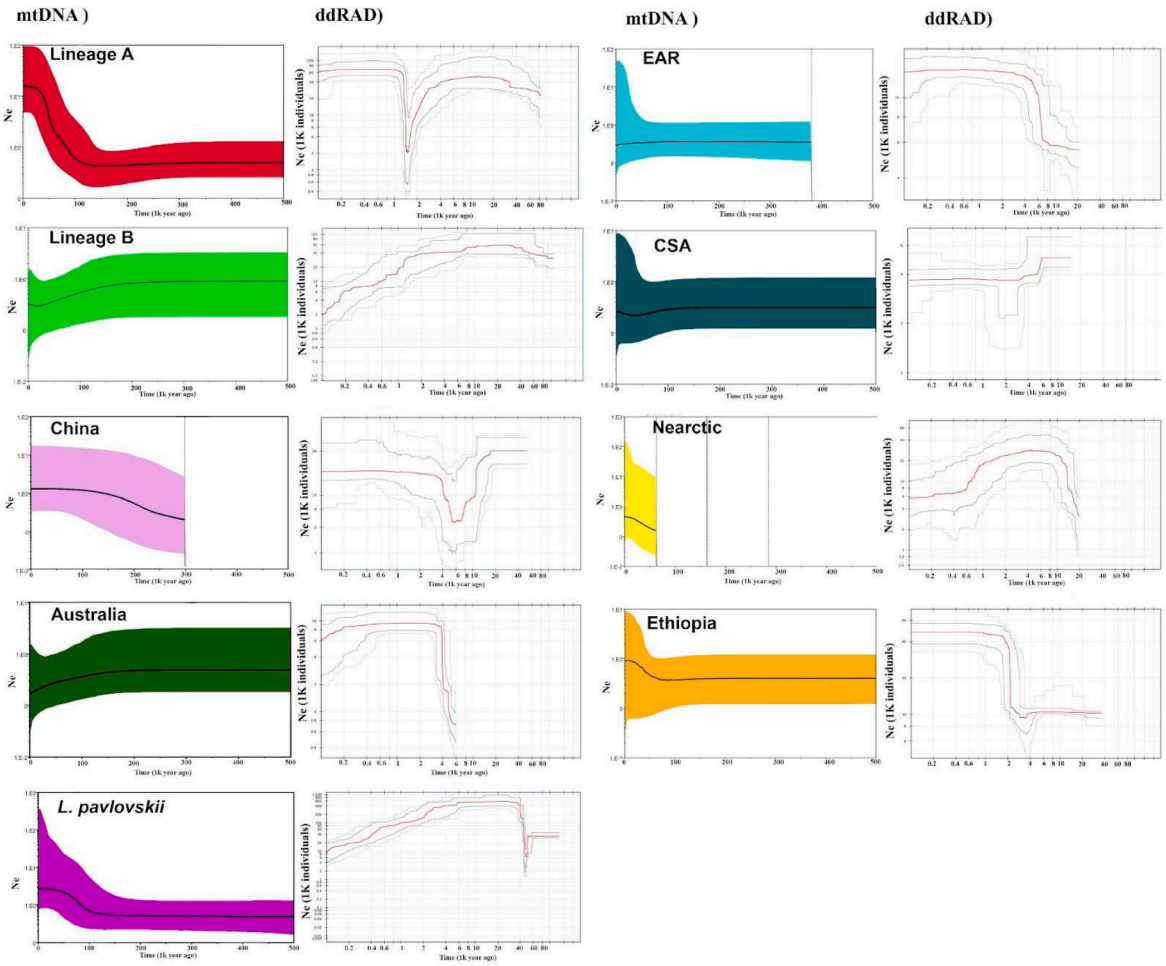


Fig. 10. Demographic history analyses. The plot on the left is EBSP based on two mtDNA genes (Cyt *b* and COI) for each main lineage. Bold line represents the median and the coloured area indicates 95% confidence interval. On the right, the stairway plot shows calculated changes in effective population size (Ne) for each lineage over time estimated from site frequency spectrum data. The grey lines above and below the red line (mean estimation) indicate 75% (dark) and 95% (light) confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which allowed for a passage of many species between the two realms (Beaudoin and Reintjes, 1994).

Founder event speciation is a form of allopatric speciation in which a small number of individuals establishes a new population beyond the existing range of the main population (Matzke, 2014). The results of the historical biogeographic model revealed that a jump dispersal event took place from the Palearctic to the Afrotropical in the early Pleistocene, forming the MRCA of EAR and CSA (Fig. 8). In line with the present study, several fossil records of waterbirds have been discovered from the early Pleistocene in EAR, the Upper Pliocene in Ethiopia as well as Miocene deposits in Kenya (Brodkorb and Mourer-Chauviré, 1982; Dyke and Walker, 2008; Prassack et al., 2018). Our results do not support the theory of *Ligula* colonization of Africa in recent decades (Gutiérrez and Hoole, 2021; Kihedu et al., 2001). Instead, we suggest that the parasite had already colonized sub-Saharan Africa in the early Pleistocene. Moreover, the dated tree based on both mtDNA and ddRAD data indicated that parasite populations from EAR and CSA diverged from each other in the late Pleistocene (Fig. 7). EAR lineage showed host specificity to *Engraulicypris sardella* and *Rastrineobola argentea,* whereas CSA lineage to barbels (*Enteromius anoplus, E. paludinosus, E. trimaculates* and *Barbus neefi*). Whilst barbels are a common host for *Ligula* in many regions (Barson and Marshall, 2003; Emaminew et al., 2014) and represent one of its ancestral hosts, a switch to the hosts endemic to the EAR region could have played a role in the differentiation of the EAR lineage.

Our biogeographic analyses also suggested the occurrence of sympatric speciation in *Ligula*. The historical biogeographic model revealed that three lineages (Lineage A, Lineage B and *L. pavlovskii*) were formed as a result of two sympatric speciation events in the western Palearctic. It also revealed that the MRCA of these lineages were historically distributed in the Palearctic (Fig. 8A and B). Moreover, some samples from these lineages were found in the same waterbodies. Lineages A and B were found in two different groups of cyprinid fish and *L. pavlovskii* was discovered in gobies, i.e., members of the distantly related fish order Gobiiformes. Similarly, Bouzid et al. (2008a, 2008b) revealed that the Euro-Mediterranean lineages (lineages A and B) live in sympatry and infect the same definitive host. They also suggested that reproductive isolation formed a significant genetic barrier between the two lineages. Moreover, a local genetic population study on *Ligula* provides preliminary evidence of a host-related ecological differentiation in Lineage A at an intra-lineage level (Nazarizadeh et al., 2022a). Reproductive isolation in sympatry is mainly a consequence of the interaction between various pre- and post-zygotic barriers (Nosil, 2012). The proximate mechanisms of differentiation in sympatric lineages of *Ligula* are yet to be explored; however, for instance, Henrich and Kalbe (2016) indicated that postzygotic ecological selection is more likely the cause of restricted gene flow between *Schistocephalus solidus* and *S. pungitii*, sister species of tapeworms related to *Ligula*.

The phenomena of geographical dispersal and host-switching are common in parasites at different spatial and temporal scales (Hoberg

and Brooks, 2008). Cyclical and episodic climatic changes are predicted to influence ecological release and the response of populations to the relaxation of isolation mechanisms that increase host-switching potential (e.g. Brooks et al., 2006; Hoberg et al., 2002). Through host-switching, a small proportion of a species moves into a new geographical area, which may be followed by speciation through peripheral isolation and addition of a new species to the parasite's host range (Huyse et al., 2005). The results of divergence times and historical biogeography confirmed that the latest "jumping speciation" in *Ligula* occurred from the Palearctic to Indomalayan and Australian realms in the mid-Pleistocene (approximately 2.0 to 1.2 mya). The host switch from Cypriniformes to native freshwater fish (Osmeriformes in China, Galaxiiformes and Gobiiformes in Australia and New Zealand) led to "peripheral isolate speciation" (Frey, 1993), a pattern previously reported in several cestode species (Brooks, 1989; Zarlenga et al., 2006).

4.4. Genome-wide hybridization and introgression

Results of the TreeMix introgression and the f-branch tests were highly consistent, with the f-branch statistics revealing a higher number of gene flow events. Introgression results suggested the presence of several ancestral ghost lineages between Ethiopia and other lineages (Fig. 10B). Ghost lineages, indicative of the existence of unsampled or extinct lineages in a species, significantly influence gene flow detection (Tricou et al., 2022). TreeMix detects introgression events based on the known lineages and branches in a tree and does not consider ghost lineages (Tricou et al., 2022).

Historical and contemporary patterns of hybridization have been documented for many tapeworms (Bello et al., 2021; Easton et al., 2020; Landeryou et al., 2022). Henrich and Kalbe (2016) confirmed natural hybridization between *S. solidus* and *S. pungitii* from both allopatric and sympatric populations. Our results detected historical hybridization in both sympatric (Lineage A, Lineage B and *L. pavlovskii*) and allopatric lineages (Nearctic, China, Lineage A, Australia and New Zealand). In addition, we discovered significant positive D-statistic values, indicating an excess of derived allele sharing between China and Lineage B, in line with the TreeMix migration edge (Fig. 10B). We suggest that historical introgressive hybridization might have enabled *Ligula* to infect a broader range of host species (Osmeriformes, Gobiiformes and Galaxiiformes), similar to adaptation of hybrids of nonparasitic taxa to novel ecological conditions where their parents could not persist (Thaenkham et al., 2022). However, future testing of specific questions related to adaptive introgression will require extended sampling and a well-resolved and annotated genome assembly.

4.5. Demographic history

Demographic analyses based on mtDNA and ddRAD data showed mutually consistent results for most lineages, with SNP based results providing a more recent picture (Fig. 10). EBSP showed Lineage A and Lineage B experienced rapid range expansions (Fig. 10A) during the Last Interglacial Period (LIG; Otto-Bliesner et al., 2006). Moreover, during the Medieval Warm Period (MWP; Crowley and Lowery, 2000), these lineages exhibited rapid expansions following a bottleneck event (Fig. 10B). Although these very recent events are not straightforward to explain without more comprehensive data and analyses, they fit into the general picture of a volatile population history expected for parasitic taxa, involving bottlenecks and expansions following host-switches as well as expansions due to post-glacial host range expansions. Periods of geographical and host range expansions allow parasites to explore their capacity for generalism, while periods of isolation force parasites to specialise on specific hosts. This specialisation will modify the "sloppy fitness space" (Agosta and Klemens, 2008; Araujo et al., 2015) of the parasite due to the parasite adapting to local environments.

Bottlenecks found in *Ligula* lineages, particularly those associated with lineage formation, could be initiated by host switches followed by population expansion in case of a successful specialisation. During a host switch, it is likely that the host immune system can decrease overall parasite load, while at the same time it selects for more immune-resistant parasite's genotypes (Levin et al., 1999). *Ligula* interacts with the immune and metabolic systems of the fish host intensively, leading to the castration of the host and redirecting energy resources towards the parasite's growth (Yoneva et al., 2015). Future analysis of coding genetic differences and dating their origin, ideally using whole genome data, could allow testing the hypotheses of host-switch associated bottlenecks.

Declaration of Competing 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 paper.

Acknowledgements

We are grateful to Anna Mácová and Roman Hrdlička for their help with the fieldwork. The authors also wish to thank Eva Čisovská, Košice, Slovakia for providing *D. latus* samples to generate outgroup sequences. Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic. Some of the samples were collected by other colleagues and provided to AH for this project and we would like to acknowledge them here: The following people who contributed to the project funding and/or the sampling in DRC, Auguste Chocha Manda, Tine Huyse, Gyrhaiss Kapepula Kasembele, Clément Kalombo Kabalika, Fidel Muterezi Bukinga, Eliane Ngoma Mavungu, Jos Snoeks, Maarten Vanhove, Filip Volckaert, Emmanuel Vreven and the two projects that funded the sampling: the South Initiative "Renforcement des capacités locales pour une meilleure évaluation biologique des impacts miniers au Katanga (RD Congo) sur les poissons et leurs milieux aquatiques" (ZRDC2014MP084) of the University Development Cooperation of the Flemish Interuniversity Council (VLIR-UOS) and the Mbisa Congo project, a framework agreement project of the Royal Museum for Central Africa and the Belgian Development Cooperation. Colleagues in Namibia, Annette Amakali, Martin Tjipute and Matthew Hanghome (University of Namibia) and South Africa, Prof. Wilmien Luus-Powell and her team (University of Limpopo), were involved in collecting part of the African samples. Sampling in South Africa, DRC and Namibia was supported by the National Research Foundation of South Africa (Grant Number 101054).

Funding

The research was supported by a grant from the Czech Science Agency (no. GA19-04676S).

Benefit-Sharing Statement

All authors took part in field-collecting samples for the research via a collaborative effort.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107677.

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31 (10), 2553–2556. https://doi.org/10.1093/molbev/msu236.

Agosta, S.J., Klemens, J.A., 2008. Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. Ecol. Lett. 11 (11), 1123–1134. https://doi.org/10.1111/j.1461-0248.2008.01237.x.
Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19 (9), 1655–1664. https://doi.org/10.1101/gr.094052.109.
Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/ fastqc.
Angst, P., Ebert, D., Fields, P.D., 2022. Demographic history shapes genomic variation in an intracellular parasite with a wide geographical distribution. Mol. Ecol. 31 (9), 2528–2544. https://doi.org/10.1111/mec.16419.
Araujo, S.B.L., Braga, M.P., Brooks, D.R., Agosta, S.J., Hoberg, E.P., von Hartenthal, F.W., Boeger, W.A., 2015. Understanding host-switching by ecological fitting. PLoS One 10 (10), e0139225. https://doi.org/10.1371/journal.pone.0139225.
Barson, M., Marshall, B.E., 2003. The occurrence of the tapeworm *Ligula intestinalis* (L.), in *Barbus paludinosus* from a small dam in Zimbabwe. Afr. J. Aquat. Sci. 28 (2), 175–178. https://doi.org/10.2989/16085910309503782.
Beaudoin, A.B., Reintjes, F.D., 1994. Late Quaternary Studies in Beringia and Beyond, 1950-1993: An Annotated Bibliography.Provincial Museum of Alberta Archeological Survey Occasional Paper NO. 35.
Bello, E., Palomba, M., Webb, S.C., Paoletti, M., Cipriani, P., Nascetti, G., Mattiucci, S., 2021. Investigating the genetic structure of the parasites *Anisakis pegreffii* and *A. berlandi* (Nematoda: Anisakidae) in a sympatric area of the southern Pacific Ocean waters using a multilocus genotyping approach: first evidence of their interspecific hybridization. Infect. Genet. Evol. 92, 104887 https://doi.org/10.1016/j.meegid.2021.104887.
Bianchini, G., Sánchez-Baracaldo, P., 2021. sMap: Evolution of independent, dependent and conditioned discrete characters in a Bayesian framework. Methods Ecol. Evol. 12 (3), 479–486. https://doi.org/10.1111/2041-210X.13540.
Bisconti, R., Porretta, D., Arduino, P., Nascetti, G., Canestrelli, D., 2018. Hybridization and extensive mitochondrial introgression among fire salamanders in peninsular Italy. Sci. Rep. 8, 13187. https://doi.org/10.1038/s41598-018-31535-x.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 10 (4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537.
Bouzid, Wafa, Štefka, J., Hypša, V., Lek, S., Scholz, T., Legal, L., Hassine, O.K. Ben, Loot, G., 2008. Geography and host specificity: Two forces behind the genetic structure of the freshwater fish parasite *Ligula intestinalis* (Cestoda: Diphyllobothriidae). Parasitol. Int. 38, 12, 1465–1479. https://doi.org/10.1016/j.ijpara.2008.03.008.
Bouzid, W., Lek, S., Mace, M., Ben Hassine, O., Etienne, R., Legal, L., Loot, G., 2008a. Genetic diversity of *Ligula intestinalis* (Cestoda: Diphyllobothriidea) based on analysis of inter-simple sequence repeat markers. J. Zool. Syst. Evol. Res. 46 (4), 289–296. https://doi.org/10.1111/j.1439-0469.2008.00471.x.
Bouzid, W., Štefka, J., Bahri-Sfar, L., Beerli, P., Loot, G., Lek, S., Haddaoui, N., Hypša, V., Scholz, T., Dkhil-Abbes, T., 2013. Pathways of cryptic invasion in a fish parasite traced using coalescent analysis and epidemiological survey. Biol. Invasions. 15 (9), 1907–1923. https://doi.org/10.1007/s10530-013-0418-y.
Brodkorb, P., Mourer-Chauviré, C., 1982. Fossil Anhingas (Aves: Anhingidae) from Early Man Sites of Hadar and Omo (Ethiopia) and Olduvai Gorge (Tanzania). Geobios 15 (4), 505–515. https://doi.org/10.1016/S0016-6995(82)80071-5.
Brooks, D.R., 1989. A summary of the database pertaining to the phylogeny of the major groups of parasitic platyhelminths, with a revised classification. Can. J. Zool. 67 (3), 714–720. https://doi.org/10.1139/z89-103.
Brooks, D.R., León-Règagnon, V., McLennan, D.A., Zelmer, D., 2006. Ecological fitting as a determinant of the community structure of platyhelminth parasites of anurans. Ecol. 87, sp7, S76–S85. https://doi.org/10.1890/0012-9658(2006)87[76:EFAADO]2.0.CO;2.
Bruderer, B., Salewski, V., 2008. Evolution of bird migration in a biogeographical context. J. Biogeogr. 35 (11), 1951–1959. https://doi.org/10.1111/j.1365-2699.2008.01992.x.
Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Mol. Biol. Evol. 29 (8), 1917–1932. https://doi.org/10.1093/molbev/mss086.
Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22 (17), 4369–4383. https://doi.org/10.1111/mec.12413.
Caye, K., Jay, F., Michel, O., François, O., 2018. Fast inference of individual admixture coefficients using geographic data. Ann. Appl. Stat. 12 (1), 586–608. https://doi.org/10.1214/17-AOAS1106.
Chang, C. C., Chow, C.C., Tellier, L.C. A.M., Vattikuti, S., Purcell, S. M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4, 1, s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8.
Chapman, A., Hobbs, R.P., Morgan, D.L., Gill, H.S., 2006. Helminth parasitism of *Galaxias maculatus* (Jenyns 1842) in southwestern Australia. Ecol. Freshw. Fish 15 (4), 559–564. https://doi.org/10.1111/j.1600-0633.2006.00198.x.
Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics 30 (23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.
Crowley, T.J., Lowery, T.S., 2000. How warm was the medieval warm period? Ambio 29 (1), 51–54. https://doi.org/10.1579/0044-7447-29.1.51.
de Jong, M.J., de Jong, J.F., Hoelzel, A.R., Janke, A., 2021. SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. Mol. Ecol. Resour. 21 (4), 1369–1379. https://doi.org/10.1111/1755-0998.13339.
Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332–340. https://doi.org/10.1016/J.TREE.2009.01.009.
Ding, F., Gu, S., Yi, M.-R., Yan, Y.-R., Wang, W.-K., Tung, K.-C., 2022. Demographic history and population genetic structure of *Anisakis pegreffii* in the cutlassfish Trichiurus japonicus along the coast of mainland China and Taiwan. Parasitol. Res. 121 (10), 2803–2816. https://doi.org/10.1007/s00436-022-07611-7.
Dobson, A., Lafferty, K.D., Kuris, A.M., Hechinger, R.F., Jetz, W., 2008. Homage to Linnaeus: How many parasites? How many hosts? Proc. Natl. Acad. Sci. U.S.A. 105 (SUPPL. 1), 11482–11489. https://doi.org/10.1073/pnas.0803232105.
Doyle, S.R., Søe, M.J., Nejsum, P., Betson, M., Cooper, P.J., Peng, L., Zhu, X.-Q., Sanchez, A., Matamoros, G., Sandoval, G.A.F., 2022. Population genomics of ancient and modern *Trichuris trichiura*. Nat. Commun. 13 (1), 1–12 https://doi.org/s41467-022-31487-x.
Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29 (8), 1969–1973. https://doi.org/10.1093/molbev/mss075.
Dubinina, M.N., 1980. Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR. Amerind Publishing Company, Delhi, p. 320.
Dyke, G.J., Walker, C.A., 2008. New records of fossil 'waterbirds' from the Miocene of Kenya. Am. Mus. Novit. 2008 (3610), 1–10 https://doi.org/10.1206/0003-0082(2008)3610[1:NROFWF]2.0.CO;2.
Easton, A., Gao, S., Lawton, S.P., Bennuru, S., Khan, A., Dahlstrom, E., Oliveira, R.G., Kepha, S., Porcella, S.F., Webster, J., 2020. Molecular evidence of hybridization between pig and human *Ascaris* indicates an interbred species complex infecting humans. Elife 9, e61562. https://doi.org/10.7554/eLife.61562.
Emaminew, T., Dereje, B., Abdu, M., 2014. Prevalence of *Ligula intestinalis* larvae in Barbus fish genera at Lake Tana. Ethiopia. World J. Fish Mar. Sci. 6 (5), 408–416.
Engelstädter, J., Fortuna, N.Z., 2019. The dynamics of preferential host switching: Host phylogeny as a key predictor of parasite distribution. Evol. 73 (7), 1330–1340. https://doi.org/10.1111/evo.13765.
Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14 (8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.
Fitak, R.R., 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol. Methods Protoc. 6 (1), bpab017. https://doi.org/10.1093/biomethods/bpab017.
Frey, J.K., 1993. Modes of peripheral isolate formation and speciation. Syst. Biol. 42 (3), 373–381. https://doi.org/10.1093/sysbio/42.3.373.
Gabagambi, N.P., Salvanes, A.-G.-V., Midtøy, F., Skorping, A., 2019. The tapeworm *Ligula intestinalis* alters the behavior of the fish intermediate host *Engraulicypris sardella*, but only after it has become infective to the final host. Behav. Processes. 158, 47–52. https://doi.org/10.1016/j.beproc.2018.11.002.
Galbreath, K.E., Toman, H.M., Li, C., Hoberg, E.P., 2020. When parasites persist: tapeworms survive host extinction and reveal waves of dispersal across Beringia. Proc. Royal Soc. B. 287 (1941), 20201825. https://doi.org/10.1098/rspb.2020.1825.
Gutiérrez, J.S., Hoole, D., 2021. Ligula intestinalis. Trends Parasitol. S1471–4922 https://doi.org/10.1016/j.pt.2021.09.005.
Hafner, M., Nadler, S.A., 1990. Cospeciation in Host-Parasite Assemblages: Comparative Analysis of Rates of Evolution and Timing of Cospeciation Events. Syst. biol. 39 (3), 192–204. https://doi.org/10.2307/2992181.
Haukisalmi, V., Hardman, L.M., Fedorov, V.B., Hoberg, E.P., Henttonen, H., 2016. Molecular systematics and H olarctic phylogeography of cestodes of the genus *Anoplocephaloides* Baer, 1923 s. s. (Cyclophyllidea, Anoplocephalidae) in lemmings (Lemmus, Synaptomys). Zool. Scr. 45 (1), 88–102. https://doi.org/10.1111/zsc.12136.
Heled, J., Drummond, A.J., 2008. Bayesian inference of population size history from multiple loci. BMC Evol. Biol. 8 (1), 289. https://doi.org/10.1186/1471-2148-8-289.
Heled, Joseph. 2015. Extended Bayesian skyline plot tutorial for BEAST 2. WWW document] URL http://evomicsorg. wpengine. netdna-cdn. com/wp-content.
Henrich, T., Kalbe, M., 2016. The role of prezygotic isolation mechanisms in the divergence of two parasite species. BMC Evol. Biol. 16 (1), 1–10. https://doi.org/10.1186/s12862-016-0799-5.
Hewitt, G.M., 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Mol. Ecol. 10 (3), 537–549. https://doi.org/10.1046/j.1365-294x.2001.01202.x.
Hirase, S., Yokoyama, Y., Lee, C.-T., Iwasaki, W., 2016. The Pliocene-Pleistocene transition had dual effects on North American migratory bird speciation. Palaeogeogr. Palaeoclimatol. Palaeoecol. 462, 85–91. https://doi.org/10.1016/j.palaeo.2016.09.006.
Hoberg, E.P., Brooks, D.R., 2013. Episodic processes, invasion, and faunal mosaics inevolutionary and ecological time. In: The Balance of Nature and Human Impact (ed.Rohde K), pp. 199–213. Cambridge University Press. https://doi.org/10.1017/CBO9781139095075.021.
Hoberg, E.P., Brooks, D.R., 2008. A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host–parasite systems. J. Biogeogr. 35 (9), 1533–1550. https://doi.org/10.1111/j.1365-2699.2008.01951.x.
Hoberg, E.P., Kutz, S.J., Nagy, J., Jenkins, E., Elkin, B., Branigan, M., Cooley, D., 2002. Protostrongylus stilesi (Nematoda: Protostrongylidae): Ecological isolation and putative host-switching between Dall's sheep and muskoxen in a contact zone.

Comp. Parasitol. 69, 1, 1–9. https://doi.org/10.1654/1525-2647(2002)069[0001:PSNPEI]2.0.CO;2.
Hoberg, E.P., Galbreath, K.E., Cook, J.A., Kutz, S.J., Polley, L., 2012. Northern Host–Parasite Assemblages: History and Biogeography on the Borderlands of Episodic Climate and Environmental Transition. Adv. Parasitol. 79, 1–97. https://doi.org/10.1016/B978-0-12-398457-9.00001-9.
Hoole, D., Carter, V., Dufour, S., 2010. *Ligula intestinalis* (Cestoda: Pseudophyllidea): an ideal fish-metazoan parasite model? Parasitology 137 (3), 425–438. https://doi.org/10.1017/S0031182010000107.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23 (2), 254–267. https://doi.org/10.1093/molbev/msj030.
Huyse, T., Poulin, R., Théron, A., Theron, A., 2005. Speciation in parasites: a population genetics approach. Trends Parasitol. 21 (10), 469–475. https://doi.org/10.1016/j.pt.2005.08.009.
Imoto, J.M., Saitoh, K., Sasaki, T., Yonezawa, T., Adachi, J., Kartavtsev, Y.P., Miya, M., Nishida, M., Hanzawa, N., 2013. Phylogeny and biogeography of highly diverged freshwater fish species (Leuciscinae, Cyprinidae, Teleostei) inferred from mitochondrial genome analysis. Gene 514 (2), 112–124. https://doi.org/10.1016/j.gene.2012.10.019.
Johnson, K.P., Adams, R.J., Clayton, D.H., 2002. The phylogeny of the louse genus Brueelia does not reflect host phylogeny. Biol. J. Linn. Soc. 77 (2), 233–247. https://doi.org/10.1046/j.1095-8312.2002.00107.x.
Jombart, T., Collins, C., 2015. Analysing genome-wide SNP data using adegenet 2.0. 0.
Kalbe, M., Eizaguirre, C., Scharsack, J.P., Jakobsen, P.J., 2016. Reciprocal cross infection of sticklebacks with the diphyllobothriidean cestode *Schistocephalus solidus* reveals consistent population differences in parasite growth and host resistance. Parasites Vectors. 9 (1), 1–12. https://doi.org/10.1186/s13071-016-1419-3.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.
Kennedy, C.R., Shears, P.C., Shears, J.A., 2001. Long-term dynamics of *Ligula intestinalis* and roach *Rutilus rutilus*: a study of three epizootic cycles over thirty-one years. Parasitology 123 (3), 257–269. https://doi.org/10.1017/S0031182001008538.
Kihedu, K.J., Mlay, M.K.L., Mwambungu, J.A., Ngatunga, B.P., 2001. Drifting long line, a potential fishing method for the northern part of Lake Nyasa/Malawi/Niassa. Lake Malawi Fisheries Management Symposium. 121.
Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinform. 35 (21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.
Kuchta, R., Scholz, T., 2017. Diphyllobothriidea Kuchta, Scholz, Brabec Bray, 2008. In: Planetary Biodiversity Inventory (2008-2017): Tapeworms from Vertebrate Bowels of the Earth. J.N. Caira & K. Jensen (eds). University of Kansas, Natural History Museum, Special Publucation No. 25, Lawrence, KS USA, pp. 167-189. https://www.marinespecies.org/aphia.php?p=sourcedetails&id=283093.
Lagrue, C., Presswell, B., Dunckley, N., Poulin, R., 2018. The invasive cestode parasite *Ligula* from salmonids and bullies on the South Island, New Zealand. Parasitol. Res. 117 (1), 151–156. https://doi.org/10.1007/s00436-017-5684-7.
Landeryou, T., Rabone, M., Allan, F., Maddren, R., Rollinson, D., Webster, B.L., Tchuem-Tchuenté, L.-A., Anderson, R.M., Emery, A.M., 2022. Genome-wide insights into adaptive hybridisation across the *Schistosoma haematobium* group in West and Central Africa. PLOS Negl. Trop. Dis. 16 (1), e0010088. https://doi.org/10.1371/journal.pntd.0010088.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34 (3), 772–773. https://doi.org/10.1093/molbev/msw260.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357–359 https://doi.org/10.1038/nmeth.1923.
Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63 (4), 534–542. https://doi.org/10.1093/sysbio/syu018.
Leigh, J.W., Bryant, D., 2015. POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6 (9), 1110–1116. https://doi.org/10.1111/2041-210X.12410.
Levin, B.R., Lipsitch, M., Bonhoeffer, S., 1999. Population biology, evolution, and infectious disease: convergence and synthesis. Science 283 (5403), 806–809. https://doi.org/10.1126/science.283.5403.806.
Li, W.X., Fu, P.P., Zhang, D., Boyce, K., Xi, B.W., Zou, H., Li, M., Wu, S.G., Wang, G.T., 2018b. Comparative mitogenomics supports synonymy of the genera *Ligula* and *Digramma* (Cestoda: Diphyllobothriidae). Parasites Vectors. 11 (1), 1–11. https://doi.org/10.1186/s13071-018-2910-9.
Li, L., Lü, L., Nadler, S.A., Gibson, D.I., Zhang, L.-P., Chen, H.-X., Zhao, W.-T., Guo, Y.-N., 2018a. Molecular phylogeny and dating reveal a terrestrial origin in the early carboniferous for ascaridoid nematodes. Syst. Biol. 67 (5), 888–900. https://doi.org/10.1093/sysbio/syy018.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25 (11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187.
Liu, X., Fu, Y.-X., 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21 (1), 1–9. https://doi.org/10.1186/s13059-020-02196-9.
Loot, G., Aulagnier, S., Lek, S., Thomas, F., Guégan, J.-F., 2002. Experimental demonstration of a behavioural modification in a cyprinid fish, *Rutilus rutilus* (L.), induced by a parasite, *Ligula intestinalis* (L.). Can. J. Zool. 80 (4), 738–744. https://doi.org/10.1139/z02-043.
Loot, G., Lek, S., Brown, S.P., Guégan, J.-F., 2001. Phenotypic modification of roach (Rutilus rutilus L.) infected with Ligula intestinalis L. (Cestoda: Pseudophyllidea). J. Parasitol. 87 (5), 1002–1010. https://doi.org/10.1645/0022-3395(2001)087[1002:PMORRR]2.0.CO;2.
Luo, H.Y., Nie, P., Yao, W.J., Wang, G.T., Gao, Q., 2003. Is the genus *Digramma* synonymous to the genus *Ligula* (Cestoda: Pseudophyllidea)? Parasitol. Res. 89 (5), 419–421. https://doi.org/10.1007/s00436-002-0802-5.
Mácová, A., Hoblíková, A., Hypša, V., Stanko, M., Martinů, J., Kvičerová, J., 2018. Mysteries of host switching: Diversification and host specificity in rodent-coccidia associations. Mol. Phylogenet. Evol. 127, 179–189. https://doi.org/10.1016/j.ympev.2018.05.009.
Maddison, W.P., Maddison, D.R., 2021. Mesquite: a modular system for evolutionary analysis. Version 3.70. http://www.mesquiteproject.org. *62*, 1103–1118.
Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: population inference from RADseq data. Mol. Biol. Evol. 35 (5), 1284–1290. https://doi.org/10.1093/molbev/msy023.
Malinsky, M., Matschiner, M., Svardal, H., 2021. Dsuite-Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21 (2), 584–595. https://doi.org/10.1111/1755-0998.13265.
Martinů, J., Hypša, V., Štefka, J., 2018. Host specificity driving genetic structure and diversity in ectoparasite populations: Coevolutionary patterns in Apodemus mice and their lice. Ecol. Evol. 8 (20), 10008–10022. https://doi.org/10.1002/ece3.4424.
Matthews, A.E., Barnett, C.J., Boves, T.J., 2022. Differential survival and dispersal of avian feather mites with contrasting host specificities. Ecol. Entomol. 47 (5), 864–871. https://doi.org/10.1111/een.13176.
Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63 (6), 951–970. https://doi.org/10.1093/sysbio/syu056.
Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R Package, Version *0.2*, *1*, 2013.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37 (5), 1530–1534. https://doi.org/10.1093/molbev/msaa015.
Morales, H.E., Pavlova, A., Josephm, L., Sunnucks, P., 2015. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. Mol. Ecol. 24 (11), 2820–2837. https://doi.org/10.1111/mec.13203.
Mussmann, S.M., Douglas, M.R., Chafin, T.K., Douglas, M.E., 2020. A dmixPipe: population analyses in A dmixture for non-model organisms. BMC Bioinform. 21 (1), 1–9. https://doi.org/10.1186/s12859-020-03701-4.
Naughton, D., 2003. Annotated bibliography of Quaternary vertebrates of northern North America: with radiocarbon dates. University of Toronto Press.
Nazarizadeh, M., Martinů, J., Nováková, M., Stanko, M., Štefka, J., 2022a. Phylogeography of the parasitic mite *Laelaps agilis* in Western Palearctic shows lineages lacking host specificity but possessing different demographic histories. BMC Zool. 7 (1), 1–16. https://doi.org/10.1186/s40850-022-00115-y.
Nazarizadeh, M., Peterka, J., Kubečka, J., Vašek, M., Jůza, T., de Moraes, K.R., Čech, M., Holubová, M., Souza, A.T., Blabolil, P., 2022b. Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies. Folia Parasitol. 69, 18. https://doi.org/10.14411/fp.2022.018.
Nosil, P., 2012. Ecological Speciation. Oxford University Press, 1–304. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001.
Otto-Bliesner, B.L., Marshall, S.J., Overpeck, J.T., Miller, G.H., Hu, A., members, C.L.I.P., 2006. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. Science, 311, 5768, 1751–1753. https://doi.org/10.1126/science.1120808.
Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192 (3), 1065–1093. https://doi.org/10.1534/genetics.112.145037.
Perrot-Minnot, M., Špakulová, M., Wattier, R., Kotlik, P., Düşen, S., Aydoğdu, A., Tougard, C., 2018. Contrasting phylogeography of two Western Palaearctic fish parasites despite similar life cycles. J. Biogeogr. 45 (1), 101–115. https://doi.org/10.1111/jbi.13118.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One. 7, 5, e37135. https://doi.org/10.1371/journal.pone.0037135.
Petkevičiūtė, R., 1992. Comparative cytogenetics of Diphyllobothrium ditremum (Creplin, 1925) and *Ligula intestinalis* (Linnaeus, 1758)(Cestoda: Pseudophyllidea). Syst. Parasitol. 23, 3, 167–173. https://doi.org/10.1007/BF00010869.
Pickrell, J., Pritchard, J., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLOS Genet 8 (11), e1002967. https://doi.org/10.1371/journal.pgen.1002967.
Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Syst. Biol. 55, 4, 595–609. https://doi.org/10.1080/10635150600852011.
Poulin, R., 2011. Evolutionary ecology of parasites. Evolutionary ecology of parasites. Princeton University Press.
Prassack, K.A., Pante, M.C., Njau, J.K., de la Torre, I., 2018. The paleoecology of Pleistocene birds from Middle Bed II, at Olduvai Gorge, Tanzania, and the environmental context of the Oldowan-Acheulean transition. J. Hum. Evol. 120, 32–47. https://doi.org/j.jhevol.2017.11.003.

Puillandre, N., Brouillet, S., Achaz, G., 2021. ASAP: assemble species by automatic partitioning. Mol. Ecol. Resour. 21, 2, 609–620. https://doi.org/10.1111/1755-0998.13281.

Rahmouni, C., Vanhove, M.P.M., Koblmüller, S., Šimková, A., 2022. Molecular phylogeny and speciation patterns in host-specific monogeneans (*Cichlidogyrus*, Dactylogyridae) parasitizing cichlid fishes (Cichliformes, Cichlidae) in Lake Tanganyika. Parasitol. Int. 52, 6, 359–375. https://doi.org/10.1016/j.ijpara.2021.12.004.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28, 21, 4737–4754. https://doi.org/10.1111/mec.15253.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574. https://doi.org/10.1093/bioinformatics/btg180.

Shaw, K.L., 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. U.S.A. 99 (25), 16122–16127. https://doi.org/10.1073/pnas.24258589.

Šimková, A., Morand, S., Jobet, E., Gelnar, M., Verneau, O., 2004. Molecular phylogeny of congeneric monogenean parasites (Dactylogyrus): a case of intrahost speciation. Evol. 58, 5, 1001–1018. https://doi.org/10.1111/j.0014-3820.2004.tb00434.x.

Šimková, A., Řehulková, E., Choudhury, A., Seifertová, M., 2022. Host-Specific Parasites Reveal the History and Biogeographical Contacts of Their Hosts: The Monogenea of Nearctic Cyprinoid Fishes. Biology, 11, 2, 229. https://doi.org/10.3390/biology11020229.

Štefka, J., Hypša, V., Scholz, T., 2009. Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: Microsatellite study of Ligula intestinalis (Cestoda). Mol. Ecol. 18, 6, 1187–1206. https://doi.org/10.1111/j.1365-294X.2008.04074.x.

Štefka, J., Hoeck, P.E.A., Keller, L.F., Smith, V.S., 2011. A hitchhikers guide to the Galápagos: co-phylogeography of Galápagos mockingbirds and their parasites. BMC Evol. Biol. 11, 1, 284. https://doi.org/10.1186/1471-2148-11-284.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 10, 2731–2739. https://doi.org/10.1093/molbev/msr121.

Tao, W., Yang, L., Mayden, R. L., He, S., 2019. Phylogenetic relationships of Cypriniformes and plasticity of pharyngeal teeth in the adaptive radiation of cyprinids. Sci. China Life Sci. 62, 4, 553–565. https://doi.org/10.1007/s11427-019-9480-3.

Techer, M.A., Roberts, J.M.K., Cartwright, R.A., Mikheyev, A.S., 2022. The first steps toward a global pandemic: Reconstructing the demographic history of parasite host switches in its native range. Mol. Ecol. 31, 5, 1358–1374. https://doi.org/10.1111/mec.16322.

Thaenkham, U., Chaisiri, K., Hui En Chan, A., 2022. Overview of Parasitic Helminth Diversity: How Molecular Systematics Is Involved. In: Molecular Systematics of Parasitic Helminths (U. Thaenkham, K. Chaisiri, A. Hui En Chan (eds.); pp. 61–86). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-1786-8_3.

Tricou, T., Tannier, E., de Vienne, D.M., 2022. Ghost lineages can invalidate or even reverse findings regarding gene flow. PLoS Biol. 20 (9), e3001776.

Vitál, Z., Boross, N., Czeglédi, I., Preiszner, B., Erős, T., Molnár, K., Cech, G., Székely, C., Sándor, D., Takács, P., 2021. First genetically verified occurrence of *Ligula pavlovskii* outside its native range and characteristics of its infection in *Neogobius fluviatilis*. J. Great Lakes Res. 47, 1, 236–241. https://doi.org/10.1016/j.jglr.2020.10.008.

Wacker, S., Larsen, B. M., Karlsson, S., Hindar, K., 2019. Host specificity drives genetic structure in a freshwater mussel. Sci. Rep. 9, 1, 1–7. https://doi.org/10.1038/s41598-019-46802-8.

Wang, S., Wang, S., Luo, Y., Xiao, L., Luo, X., Gao, S., Dou, Y., Zhang, H., Guo, A., Meng, Q., 2016. Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. Nat. Commun. 7, 1, 1–12. https://doi.org/10.1038/ncomms12845.

Weigand, H., Weiss, M., Cai, H., Li, Y., Yu, L., Zhang, C., Leese, F., 2017. Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. Mol. Ecol. 26, 20, 5705–5715. https://doi.org/10.1111/mec.14292.

Yoneva, A., Scholz, T., Młocicki, D., Kuchta, R. 2015. Ultrastructural study of vitellogenesis of *Ligula intestinalis* (Diphyllobothriidea) reveals the presence of cytoplasmic-like cell death in cestodes. Front. Zool. 12, 1, 1–9. https://doi.org/10.1186/s12983-015-0128-7.

Zarlenga, D.S., Rosenthal, B.M., La Rosa, G., Pozio, E., Hoberg, E.P., 2006. Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. Proc. Natl. Acad. Sci. U.S.A. 103, 19, 7354–7359. https://doi.org/10.1073/pnas.0602466103.

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29 (22), 2869–2876. https://doi.org/10.1093/bioinformatics/btt499.

Zink, R.M., Gardner, A.S., 2017. Glaciation as a migratory switch. Sci. Adv. 3, 9, e1603133. https://doi.org/10.1126/sciadv.160313.

Supplemental Information for:

Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite - the *Ligula intestinalis* species complex

Masoud Nazarizadeh[1,2], Milena Nováková2, Géraldine Loot[3], Nestory P. Gabagambi[4], Faezeh Fatemizadeh[5], Odipo Osano[6], Bronwen Presswell[7], Robert Poulin[7], Zoltán Vitál[8], Tomáš Scholz[1,2], Ali Halajian[9], Emiliano Trucchi[10], Pavlína Kočová[11], Jan Štefka[1,2*]

Table of Contents:

Section 1

Polymerase chain reaction (PCR) and mitochondrial sequencing

Each gene fragment was amplified in a 12 ul PCR containing 6.25ul 2x concentrated PPP Master Mix (Top-Bio, CZ), 1 ul of extracted DNA and 10 pM of forward and reverse primers. The amplification and sequencing primers used in this study are listed in Table S3. The protocol for amplification of Cyt *b* (404-900 bp) and COI genes (396 bp) consisted of an initial denaturation step for 15 min at 94 °C, followed by 30 cycles of 30 s denaturation at 94 °C, annealing for 45 s at 50 °C, and extension for 45 s at 72 °C. The last elongation step was carried out for 10 min at 72 °C. The same procedure was applied for the ND1 gene (891 bp), except for the annealing temperature which was set at 55 °C. VWR ExoCleanUp FAST PCR reagent (VWR, USA) was used to enzymatically clean up PCR products following the manufacturer's instructions. Purified PCRs were Sanger sequenced in a commercial laboratory (Seqme, CZ) using PCR primers.

Section 2

ddRAD library preparation

ddRAD libraries were prepared following the Peterson et al. (2012) protocol with modifications specified below. Briefly, genomic DNA was digested at 37 °C for 2 h using the restriction enzymes NspI and MluCI (NEB). Complete digestion was verified on 1,5% agarose gel. The Ligation step for 30ul reactions contained 0.2uM P1-flex adapter, 1uM P2-flex adapter, 1X T4 ligase buffer, 400U T4 DNA ligase, ddH2O and standardizing amount of digested DNA per sample in a range of 76-100ng for each individual library. Reaction conditions were: 22°C for 60 minutes following by enzyme heat inactivation at 65°C for 10 minutes at ramp rate -1°C/min. A total of 48 uniquely barcoded samples in each library were pooled into one tube, vortexed and distributed equally into 18 strip tubes. Two steps of magnetic beads purification were carried out. First, all samples were diluted by 10 ul ddH2O and combined into two tubes before magnetic plate incubation. Then, the resulting DNA was second cleaned and diluted by 16 ul ddH2O. The Contents of both tubes were combined together before final magnetic plate incubation. Finally, pooled samples were verified with Qubit 2.0 Fluorometer and size selected in a range of 275-325 bp (Pippin Prep, Sage Science). Finally, PCR reactions were performed combining 6.25ng of size-selected sample, PCR primers 1 and 2 at final concentration 0.2uM each, the recommended amount of 5X-HF buffer, dNTPs, ddH2O and Phusion polymerase, and run on 12 cycles. Quality and concentrations of final libraries were assessed using agarose gel electrophoresis and a Qubit 2.0 fluorimeter, respectively. Libraries were sequenced using Illumina Novaseq 150 PE reads in a commercial laboratory (Novogene, UK).

Section 3

SNP filtering of assembled ddRAD data

ddRAD assembly was aided by mapping the reads against a draft genome of *L. intestinalis* obtained in the frame of a concurrent study. After assembly in the Stacks (Rochette et al., 2019), we applied several different filters to the population component of stacks to keep loci that were present across 90% of all populations (p), in a minimum of 80% of individuals from each population (r=0.80) where the minor allele frequency (MAF) was 0.05 and demonstrated a maximum observed heterozygosity of 80% (max_obs_het= 0.80). As regards PCA, phylogeny analyses and FST-based analyses, we discarded SNPs in linkage disequilibrium with the squared coefficient of correlation 0.5 in 50-SNP sliding windows. In addition to the post-processing step, further filtering of variants was carried out in vcftools v0.1.16 (Danecek et al., 2011), removing variants with low (<5x) and high depth of coverage (800x). Moreover, all loci with higher than 10% missingness were discarded.

Section 4

Phylogenetic analyses

The best-fitting substitution models were estimated using the Bayesian Information Criterion (BIC, Table S3). Bayesian and maximum likelihood phylogenetic trees were reconstructed using the selected models in Mrbayes and IQtree, respectively. The Bayesian analysis was carried out with one cold and three heated chains (MC3) for 60 million generations and sampled every 2000th generation with a burn-in of 25%. We examined convergence in the Tracer v1.5 package (Rambaut et al., 2018), which was then visualised with the convergence diagnostic parameters from MrBayes.

Raxml-ng (Kozlov et al., 2019) was run using the GTR+G substitution model to perform 50 ML tree searches with 25 random and 25 parsimony-based starting trees. 1000 nonparametric bootstrap replicates were generated based on the best tree search. Convergence was checked post-hoc using the --bsconverge command in raxml-ng after changing the cutoff value to 0.05; Branch supports were computed and mapped onto the best-scoring ML tree by means of Felsenstein bootstraps (FBP) and Transfer Bootstrap Expectation (TBE) (Lemoine et al., 2018). We used four coupled chains in ExaBayes (Aberer et al., 2014, all with three heated chains) to run two Metropolis-coupling replicates for 1000000 MCMC generations, sampling every 500th generation. The postProcParam tool was used to assess the estimated sample size (ESS) of all parameters and branch lengths, which was confirmed to be adequately sampled using Tracer v.1.6 (Aberer et al. 2014). Eventually, a majority-rule consensus tree was computed using the 'consense' tool from the Exabayes package (burn-in: 25%). Species trees were constructed using SVDquartets (Chifman and Kubatko, 2014) in PAUP v.4.0a147 (Swofford and Sullivan, 2003). Subsampling was performed at 10,000 quartets. The sample quartets were assembled in a species tree with a variant of the FM Quartet as recommended by the developers of SVDquartets (Chifman and Kubatko, 2014). To inspect the incongruence between the topologies obtained from mtDNA and ddRAD data sets, we tested the fit of a constrained SNP topology on the mtDNA dataset using Shimodaira-Hasegawa (SH), Approximately unbiased (AU), and Kishino and Hasegawa (KH) implemented in PAUP.

Section 5

Population genetic structure analyses

the AdmixPiPe (Mussmann et al., 2020) was used with a default thin parameter (t=100) and ran 20 ADMIXTURE replicates for each value of K from 1 – 20 using the second VCF file described in material and methods. The optimal values of K were calculated in ADMIXTURE based on the lowest of cross-validation error (CV) across replicate runs (Alexander & Lange, 2011, Fig. S2). CLUMPK server (http://clumpak.tau.ac.il/) was used to plot clustering individuals into populations.

The python script Stacks2fneRAD.py in the fineRADstructure package was used to calculate the distribution of alleles and SNPs per locus and of the missing data per individual. A maximum number of 10 SNP per locus (13291 loci) and a maximum 25% individual missingness (139 individual) were considered to convert the haplotype file into RADpainter format. Considering that ddRAD is susceptible to batch effect as a result of small deviations among libraries at the size selection step, we assessed whether missing data were driving any library-based structure. The fineRADstructure pipeline was run using default settings, but we raised the number of burn-in iterations to 200000, along with 1000000 iterations, and sampling every 1000. Subsequently, we calculated convergence by 1) assigning individuals to populations for multiple independent runs, 2) tracing plots for the MCMC output of the parameter values to guarantee convergence on the same Bayesian posterior distributions, and 3) generating effective parameter sample sizes (> 100) by running each chain long enough. "FinestructureLibrary.R" function in fineRADstructure was used to plot the co-ancestry heatmap (Malinsky et al., 2018).

Section 6

Species delimitation analyses

To generate an ultrametric gene tree for the analysis, we reconstructed a phylogenetic tree based on the concatenated alignment (Cyt b, COI, and ND1) using BEAST v.1.8.2 (Drummond et al., 2012), under the assumption of uncorrelated lognormal relaxed molecular clock. The analyses were performed under the constant population size coalescent as the tree prior and Ucld mean prior was set to exponential distribution with mean 10 and initial value 1. We executed analyses in two independent runs with 100 million generations, sampling every 5000. We used Tracer v.1.6 for assessing convergence and adequacy of posterior samples (ESS > 200). The first 20-25 % of trees were discarded, followed by combining the rest of trees using LogCombiner v 1.8.2. We subsequently used mean heights for annotation to compute a maximal clade credibility (MCC) tree by TreeAnnotator utility v.1.8.2 (Bouckaert et al., 2014). The GMYC approach (Pons et al., 2006) is based on discerning stochastic birth-death processes (essentially a pure-birth Yule process) between species from neutral coalescent intraspecific processes by analyzing the timing of branching events (successively summarized as combination of independent Poisson models) in single gene trees. Input prerequisites need a well-sampled, well-estimated, ultrametric single neutral locus tree which optimally characterizes the true species genealogy in absence of population structure and population size fluctuation. We used the P2C2M.GMYC r package (Fonseca et al., 2021) to test whether the data has a good fit to the GMYC model. The GMYC approach with a single threshold (GMYC) was carried out in RStudio v. 4.2 (RStudio Team, 2020.) using the Splits R package. The ultrametric tree from the Beast analyses was used for both analysis. Using the function 'multi2di', we avoided having branches of length 0. Contrary to GMYC, bPTP (Zhang et al., 2013) only needs a simple phylogenetic tree rather than an ultrametric tree. Therefore, we used the ML phylogenetic tree reconstructed in the IQTREE. The analyses were run in the bPTP servers available at http://species.h-its.org/ and http://mptp.h-its.org/#/tree. For the genetic distance-based approaches (ASAP, Puillandre et al., 2021), each gene data set was treated independently as a single partition to determine the possible putative species. The ASAP analysis only employs pairwise genetic distances to sort specimens into putative species without an a priori species hypothesis. It performs automatic detection of substantial variations between and within species. We uploaded the aligned sequences of the three genes (Cyt *b*, COI, and ND1) separately to the web server https://bioinfo.mnhn.fr/abi/public/asap. The analysis was conducted based on uncorrected pairwise genetic distances and split groups below 0.01 probability.

Bayes Factor Delimitation (BFD, Leaché et al., 2014) was carried using SNAPP package (Bryant et al., 2012) in BEAST2. The unsampled mutation rates (u, v) were both fixed at 1, alpha at 1, beta at 250 and lambda at 20. A coalescence rate initially set to 10 was sampled and default values were used for all other parameters. To achieve marginal likelihood estimation for each competing model, path sampling was performed for 48 steps in total, followed by MCMC runs of 200,000 generations (sampled every 1,000 steps). We used Bayes factors to rank and compare the resulting marginal likelihood values (Kass and Raftery, 1995). To monitor the possible influence of different priors, the analyses were repeated with default options for the mutation rates (u, v), alpha, beta, lambda, and coalescence rate. Using the highest ranked model from BFD, we performed a final species tree estimation in SNAPP. MCMC was run for 1,000,000 generations, sampling every 1,000 steps. We used Tracer to confirm high ESS values and convergence, and TreeAnnotator v.1.7.5 to construct an MCC tree with a 25% burn-in.

Section 7

Divergence time and phylogeographic structure

Substitutions and clock models were unlinked with the following settings: an uncorrelated lognormal clock for Cytb and COI and a strict clock for ND1. The dynamics of speciation and extinction were modelled by a birth-death process to infer lineage-specific rate heterogeneity. We performed four independent MCMC replicates at 400 million generations, sampled every 5000 stepovers. Tracer 1.5 was applied to evaluate the appropriate amount of burn-in and effective sample sizes of all parameters (ESS>500). Additionally, LogCombiner v.1.8.2 (Rambaut and Drummond, 2015) was used to merge all four tree files after removing a range of 18-26% as burn-in. Finally, a consensus tree with divergence times was generated in TreeAnnotator v1.8.2 with the maximum clade credibility tree and posterior probabilities of 0.95. For ddRAD data set, four replicate MCMC runs were carried out, with the tree and parameter values sampled every 10000 generations over a total length of 5 × 108 steps. We used the diagnostic tool Tracer v1.5 in assessing convergence between runs, estimating an appropriate burn-in period to be implemented, and obtaining reasonable estimates of model parameter variance by ensuring the adequacy of effective sample sizes (larger than 500). All runs were combined using LogCombiner We used FigTree v1.3.1 (Rambaut and Drummond, 2010) to view the sampled maximum clade credibility tree.

Section 8

Ancestral host reconstructions

We used sMap (Bianchini and Sánchez-Baracaldo, 2021) to evaluate three discrete models by means of the Bayesian information criterion (BIC): (i) all rates of transition were the same (ER); (ii) all rates of transition were different (ARD); and (iii) forward and reverse rates of transitions between states were identical (i.e., symmetrical, SYM). ER was a better fit model to both mtDNA and ddRAD data than either the ARD or the SYM model (Table S4). The phylogenetic trees were normalised using the -N/--norm option. We carried out the analysis with two independent parallel MCMC runs, sampled every 10 steps over 1000 generations, a minimum number of 2000 samples were collected from four chain analyses, and 200 samples were burin-in. We checked the convergence of Bayesian analysis using the "ChainMonitor" function (ESS > 800). Stochastic mapping was visualised in TreeViewer V 2.0.1 (https://treeviewer.org/).

Section 9

Demographic history

The Extended Bayesian Skyline Plot was carried in BEAST 2. We ran Markov chains for 600 × 10^7 generations and sampled every 10,000th generation. Log files were analysed in Tracer v.1.7 (Rambaut et al., 2018) to evaluate the convergence of the MCMC analysis with effective sample sizes (ESS larger than 200), combined with 10% burn-in removal during 'skyline' plotting.

We ran Stairway Plot2 (Liu and Fu, 2020) using the two-epoch model, with the recommended 67% of sites used for training and 200 bootstraps on the folded SFS. Singletons were excluded

from the estimation to reduce genotype calling errors. We estimated the mutation rate per generation to be 2.89×10^{-9} according to our evaluation of the divergence time. The analysis is a flexible multi-epoch model derived from the composite likelihood of an observed site frequency spectrum (SFS) and is not based on parameters or predefined models (such as constant population size followed by a bottleneck) to infer demographic histories. Therefore, a more detailed demographic history can be reflected in the model.

Table S2 List of primers used for amplification and sequencing.

Primer	Sequences	size (bp)	Source
Cyt *b*			
F2Dnihcob	5´– GTT TTA CTG ATA GGT TAT TTA AAC-3´	1100	Wicht et al. 2010
R2Dnihcob	5 – CAA TTT AAA AAA CGA GTT AAA GAT-3		
COBA	5- GTA TGT GGC TGA TTC AGA GTT GAG C-3	400	
COBB	5-TTC GAG CCC GAA GAA TGC AAG TAG-3		Bouzid et al. 2008
COI			
COIA2	5 -CAT ATG TTT TGA TTT TTT GG-3	400	
COIB2	5-AKA ACA TAA TGA AAA TGA GC-3		Bouzid et al. 2008
COIA2_L_Pavlovski	5 -CAT ATG TTC TGG TTC TTT GG-3	400	Present study
COIB2_L_Pavlovski	5 -ATT ACA TAG TGA AAG TGA GC-3		
ND1			
Spi-ND1F	5' – GGA GAA TAT TGG TTT GTC TAA CCA-3	1000	Eom et al. 2018
Spi-ND1R	5 – CCT TCT TAA CGT TAA CAG CAT TAC GAT-3		

Table S3 Results of BIC substitution model selection in PartitionFinder2 for different partitions of the mtDNA dataset.

	BEAST	IQtree	MrBayes		
			Substitution model	nst	rates
Cytb-codon1	TRN+I+G+X	TRN+I+G	GTR+I+G	6	Invgamma
Cytb-codon2	HKY+I+X	HKY+I	HKY+I	2	Propinv
Cytb-codon3	GTR+G+X	GTR+G	GTR+G	6	Gamma
COI-codon1	TRN+I+G+X	TRN+I+G	GTR+I+G	6	Invgamma
COI-codon2	HKY+I+X	HKY+I	HKY+I	2	Propinv
COI-codon3	GTR+G+X	GTR+G	GTR+G	6	Gamma
ND1-Codon1	GTR+G+X	GTR+G	GTR+G	6	Gamma
ND1-codon2	TRN+I+G+X	TRN+I+G	GTR+I+G	6	Invgamma
ND1-codon3	HKY+I+X	HKY+I	HKY+I	2	Propinv

Table S4. Maximum likelihood estimates for six biogeography scenarios indicate the DIVALIKE+ J showed the optimal range evolution (i.e. maximize likelihood). Abbreviations, DEC, dispersal–extinction–cladogenesis; ln LnL, log-likelihood; K, parameters; d, dispersal; l, extinction; j, cladogenesis per-event weights; AIC, Akaike information criterion, AIC-wt: Weighted Akaike information criterion.

Data set	Model	LnL	numparams	d	e	j	AIC	AIC_wt
mtDNA	DEC	-15.81	2	1.00E-12	1.00E-12	0	35.63	0.069
	DEC+J	-13.47	3	1.00E-12	1.00E-12	0.11	32.94	0.26
	DIVALIKE	-17.52	2	0.027	1.00E-12	0	39.04	0.012
	DIVALIKE+J	**-12.69**	**3**	**1.00E-12**	**1.00E-12**	**0.1**	**31.37**	**0.58**
	BAYAREALIKE	-24.68	2	0.035	0.36	0	53.36	9.70E-06
	BAYAREALIKE+J	-14.65	3	1.00E-07	1.00E-07	0.13	35.3	0.081
ddRAD	DEC	-20.28	2	0.018	0.0044	0	44.55	0.001
	DEC+J	-13.35	3	1.00E-12	1.00E-12	0.088	32.71	0.37
	DIVALIKE	-18.36	2	0.03	1.00E-12	0	40.73	0.0066
	DIVALIKE+J	**-13.05**	**3**	**1.00E-12**	**1.00E-12**	**0.09**	**32.1**	**0.5**
	BAYAREALIKE	-24.12	2	0.039	0.34	0	52.24	2.10E-05
	BAYAREALIKE+J	-14.39	3	1.00E-07	1.00E-07	0.099	34.78	0.13

Table S5. Host ancestral reconstruction . AIC and BIC for each model

	mtDNA		ddRAD	
	AIC	BIC	AIC	BIC
ER	**35.264**	**35.662**	**26.806**	**27.204**
SYM	52.389	56.369	36.606	38.993
ARD	66.283	74.240	43.863	48.638

Table S6 Output of Dtrios command: results of D statistics tests for 11 lineages. Trios with Z > 3 are shown in the table.

P1	P2	P3	Dstatistic	Z-score	p-value	f4-ratio	BBAA	ABBA	BABA
L. alternans	LineageA	L. pavlovskii	0.492497	9.65808	0	0.05994	49.7675	44.6564	15.1848
LineageA	LineageB	L. pavlovskii	0.353104	8.08807	6.66E-16	0.051701	50.1847	45.8005	21.8964
Australia	LineageB	LineageA	0.326624	6.69574	2.15E-11	0.155122	132.357	25.9863	13.1903
NewZealand	LineageB	LineageA	0.321417	6.55221	5.67E-11	0.151427	131.57	25.5616	13.1266
L. pavlovskii	LineageB	China	0.385941	6.28436	3.29E-10	0.099393	67.3717	57.2107	25.3479
China	NewZealand	L. pavlovskii	0.379399	5.99123	2.08E-09	0.071975	66.2316	62.8912	28.2952
China	Australia	L. pavlovskii	0.374515	5.86748	4.42E-09	0.071631	66.4985	63.1818	28.7515
LineageA	L. alternans	Canada	0.443981	5.86069	4.61E-09	0.059504	83.4059	38.5717	14.8524
Canada	L. alternans	EAR	0.461102	5.67512	1.39E-08	0.05135	66.5512	55.097	20.3214
Canada	L. alternans	CSA	0.454264	5.54966	2.86E-08	0.053939	65.35	56.7708	21.3042
EAR	LineageA	Canada	0.399662	5.47819	4.30E-08	0.053472	67.0639	39.4172	16.9067
China	LineageA	L. pavlovskii	0.278626	5.33177	9.73E-08	0.037697	48.1374	41.576	23.4563
LineageB	L. alternans	Canada	0.424636	5.18445	2.17E-07	0.064532	70.1938	43.424	17.5375
LineageB	LineageA	L. alternans	0.345835	4.92105	8.61E-07	0.044535	58.3266	35.1494	17.0849
CSA	LineageA	Canada	0.364392	4.89599	9.78E-07	0.048537	67.813	38.0544	17.7278
China	LineageA	Ethiopia	0.345126	4.63539	3.56E-06	0.016629	210.466	21.4168	10.4268
L. alternans	China	LineageB	0.309172	4.4105	1.03E-05	0.147855	70.4412	56.5223	29.8259
EAR	China	Canada	0.360416	4.38773	1.15E-05	0.069736	60.55	55.4054	26.0482
L. pavlovskii	L. alternans	Canada	0.337482	4.36498	1.27E-05	0.06389	60.9341	50.6826	25.1055
L. alternans	China	NewZealand	0.333782	4.16317	3.14E-05	0.067358	72.6845	64.8393	32.3869
L. alternans	China	Australia	0.331862	4.12352	3.73E-05	0.070369	73.0595	65.025	32.6202
China	LineageB	EAR	0.323237	4.09997	4.13E-05	0.03395	88.0403	44.6698	22.8461
EAR	LineageB	Canada	0.319196	4.08641	4.38E-05	0.048324	78.0203	42.0382	21.6949
China	NewZealand	EAR	0.354538	3.96094	7.47E-05	0.043836	96.9184	53.8291	25.6505
CSA	China	Canada	0.334574	3.9394	8.17E-05	0.064886	60.3399	54.1952	27.022
China	LineageB	CSA	0.312282	3.86498	0.000111	0.03337	87.1603	43.8233	22.9661
China	LineageA	CSA	0.294843	3.86012	0.000113	0.026913	77.0632	36.8179	20.0506
Australia	L. alternans	Canada	0.379397	3.81367	0.000137	0.062719	72.5625	45.6167	20.5233
China	LineageA	EAR	0.273006	3.80504	0.000142	0.024392	78.8571	36.5	20.8446
L. pavlovskii	NewZealand	LineageA	0.21155	3.7911	0.00015	0.162041	46.019	45.396	29.5427
China	Australia	EAR	0.336092	3.75798	0.000171	0.041517	97.3969	52.9648	26.3184
EAR	L. pavlovskii	Canada	0.289487	3.74482	0.000181	0.049059	79.3299	45.9975	25.3449
L. pavlovskii	Australia	LineageA	0.205122	3.71137	0.000206	0.158352	46.3283	45.5099	30.0176
NewZealand	L. alternans	Canada	0.375746	3.67872	0.000234	0.061676	72.2292	45.1417	20.4833
China	NewZealand	CSA	0.335035	3.64698	0.000265	0.042571	96.631	52.9286	26.3631
L. alternans	LineageA	Ethiopia	0.274298	3.5445	0.000393	0.011789	205.605	18.0065	10.2546
Australia	LineageA	L. alternans	0.279811	3.54036	0.0004	0.044203	51.2906	40.9094	23.021
NewZealand	LineageA	L. alternans	0.28029	3.51274	0.000444	0.043464	51.2734	40.5312	22.7845
CSA	LineageB	Canada	0.282448	3.51161	0.000445	0.043362	77.2417	41.226	23.0667
China	Australia	CSA	0.317076	3.49414	0.000476	0.040299	97.1167	52.1452	27.0381
EAR	NewZealand	Canada	0.324139	3.47698	0.000507	0.051242	85.1696	44.0607	22.4893
L. pavlovskii	LineageB	L. alternans	0.261776	3.45404	0.000552	0.039019	76.6668	39.8094	23.2912
EAR	Australia	Canada	0.313295	3.44283	0.000576	0.050208	84.3125	44.3007	23.1643
L. alternans	LineageB	EAR	0.27597	3.4073	0.000656	0.033589	66.4929	49.8188	28.269
L. alternans	NewZealand	EAR	0.3165	3.34234	0.000831	0.043427	68.6726	58.0357	30.131
CSA	L. pavlovskii	Canada	0.264706	3.33406	0.000856	0.044101	78.3521	44.1198	25.651
L. alternans	LineageA	EAR	0.244146	3.31422	0.000919	0.023975	80.0984	39.1916	23.81

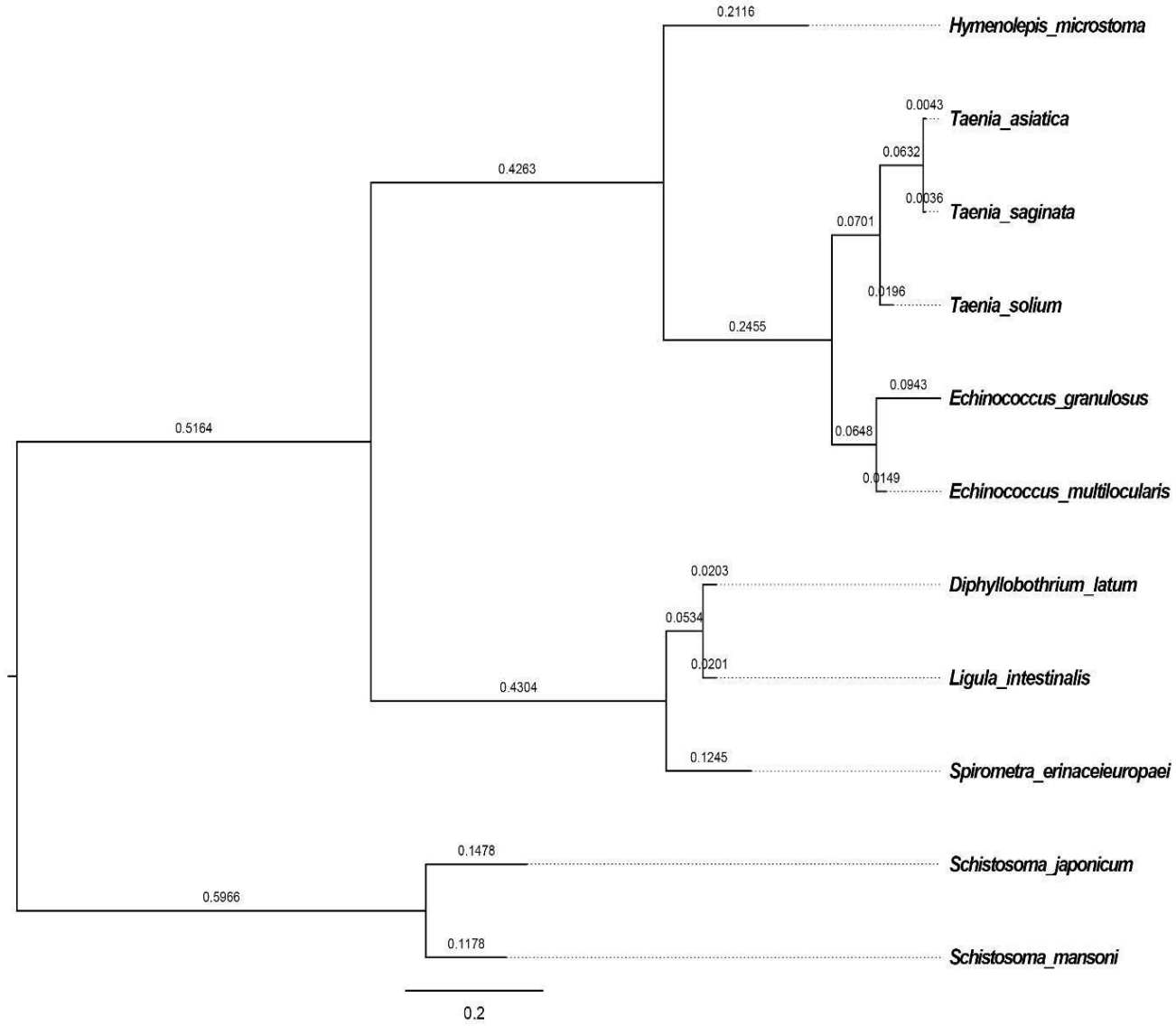


Fig. S1 phylogeny of 11 species using maximum likelihood inferred from the coding DNA sequences (CDS). Substitution rates are shown above the branches.

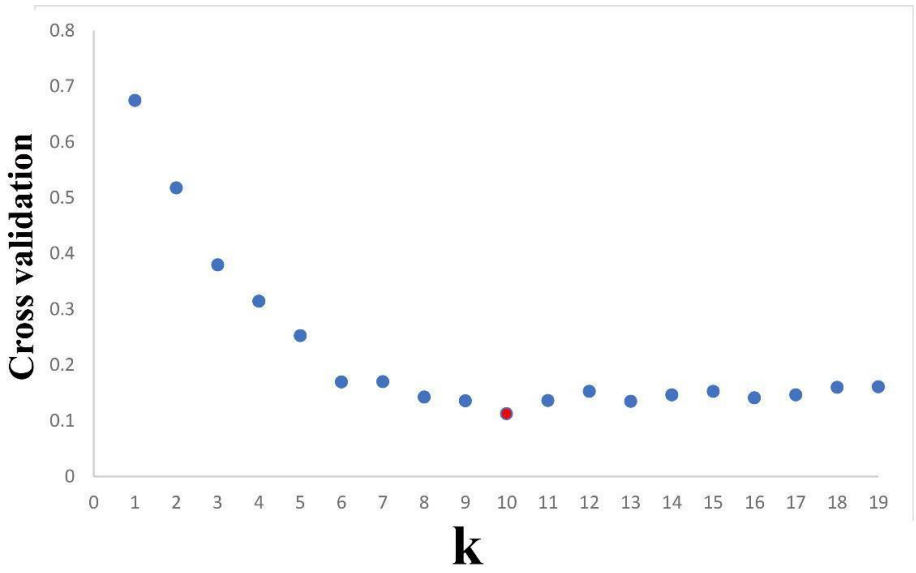


Fig. S2 Admixture cross validation from K=1 to K-19. K=10 was identified as the best predictive accuracy.

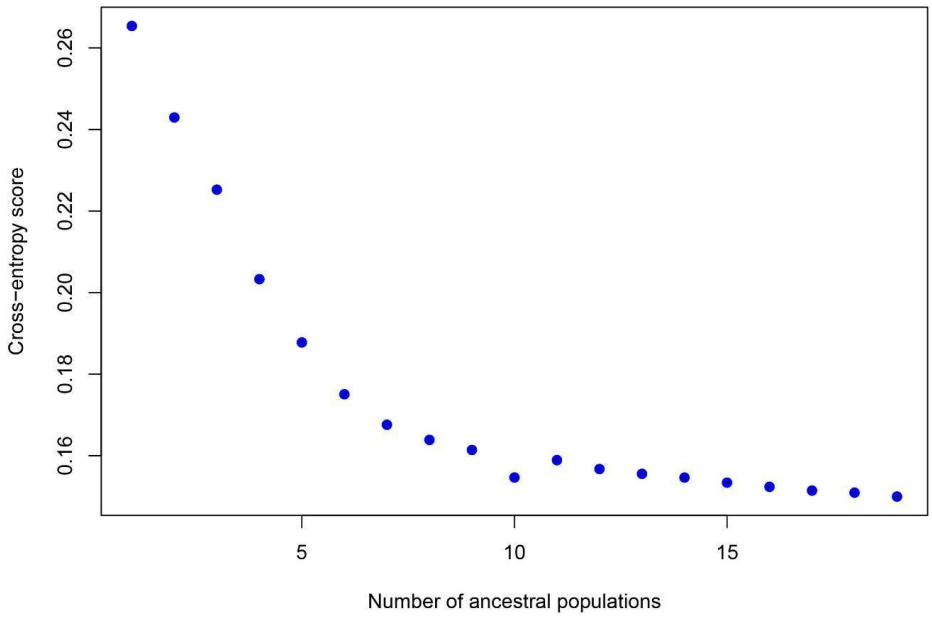
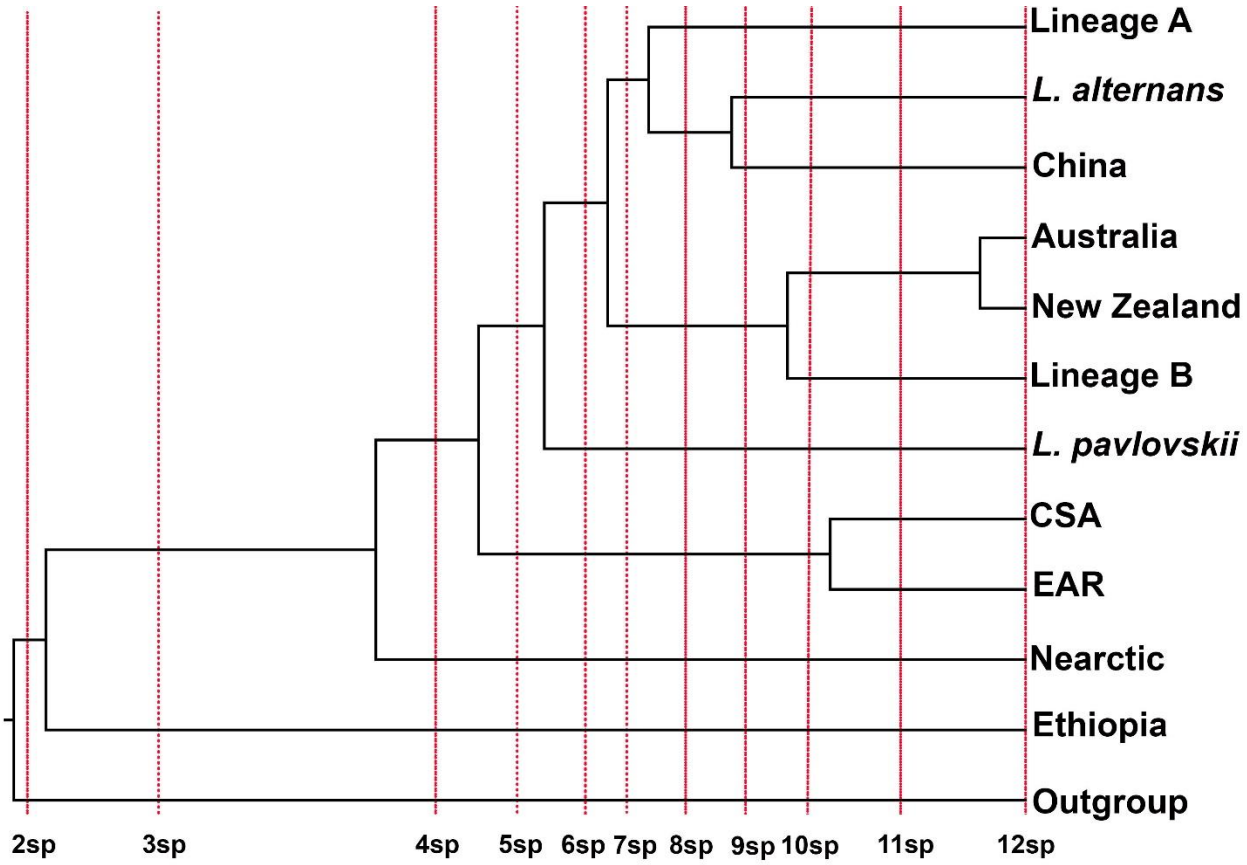


Fig. S3 Cross-entropy plot from the LEA r package revealed 10 genetic clusters for the *Ligula* complex species.



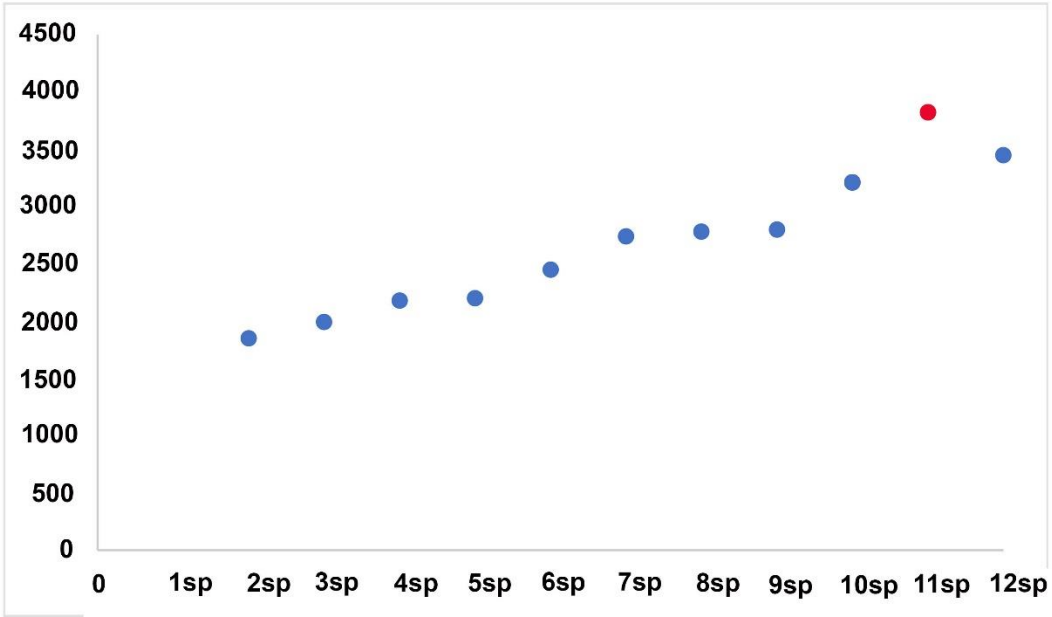


Fig. S4 Marginal-likelihood results for 12 species hypotheses based on the Bayes Factor Delimitation analysis.

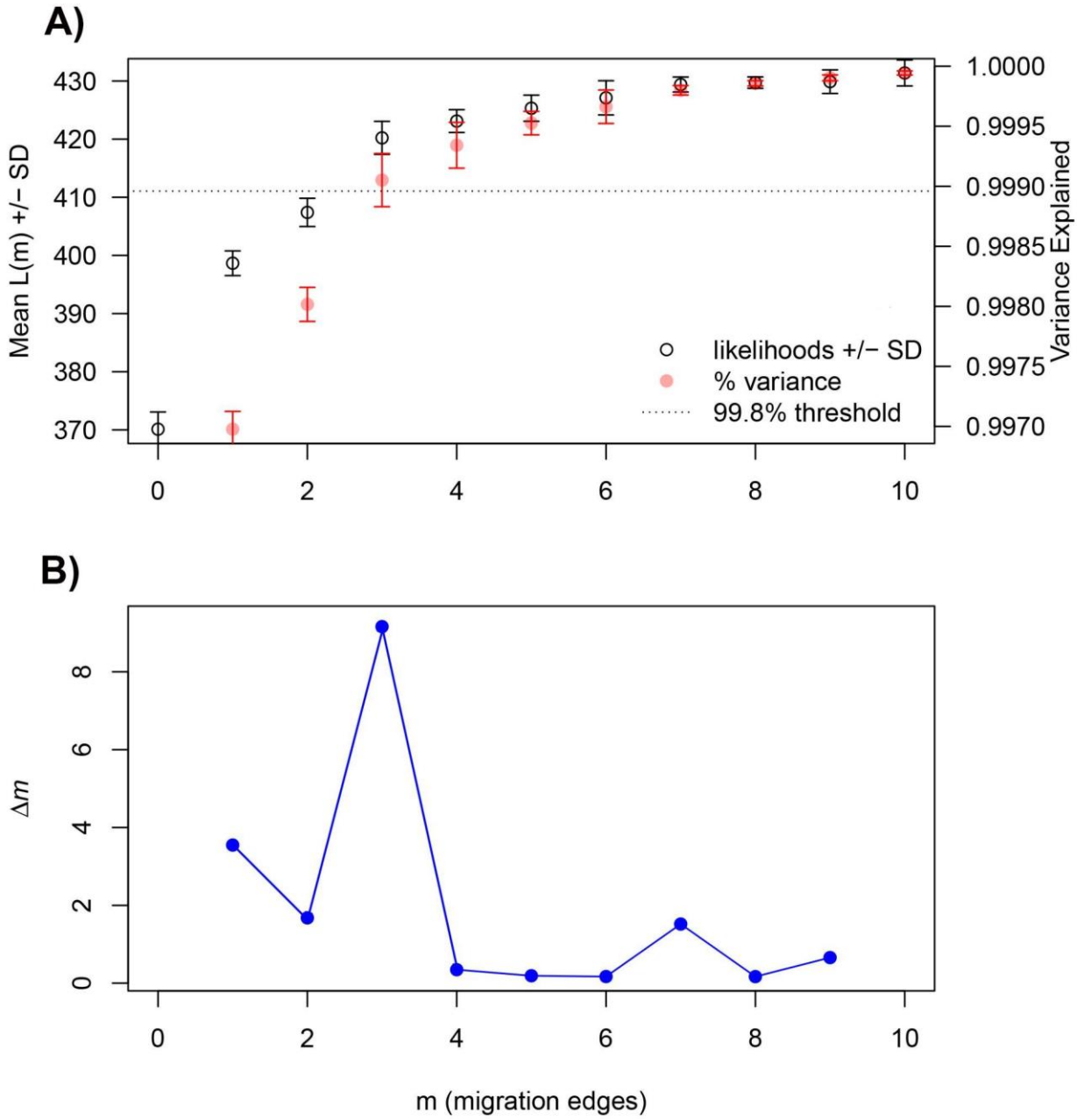


Fig. S5 Results of OptM reveal that the simplest model in TreeMix comprises three migration routes. A) the black dotted line shows 99.8% variance explained by m=3 (it is recommended by Pickrell and Pritchard, 2012. B) The second-order rate of change (Dm) demonstrated the maximum value (9.3208) for delta at m = 3 edges.

References

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinform. 12, 1–6. https://doi.org/ 10.1186/1471-2105-12-246.

Bianchini, G., Sánchez-Baracaldo, P., 2021. sMap: Evolution of independent, dependent and conditioned discrete characters in a Bayesian framework. Methods Ecol. Evol. 12, 479–486. https://doi.org/ 10.1111/2041-210X.13540.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. https://doi.org/ 10.1371/journal.pcbi.1003537.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring Species Trees Directly from Biallelic Genetic Markers: Bypassing Gene Trees in a Full Coalescent Analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086

Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent Model. Bioinform. 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinform. 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. https://doi.org/10.1093/molbev/mss075

Fonseca, E.M., Duckett, D.J., Carstens, B.C., 2021. P2C2M. GMYC: An R package for assessing the utility of the Generalized Mixed Yule Coalescent model. Methods Ecol. Evol. 12, 487–493.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinform. 35, 4453–4455. https://doi.org/10.1111/2041-210X.13541.

Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–542. https://doi.org/10.1126/sciadv.160313.

Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., Gascuel, O., 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556, 452–456. https://doi.org/10.1038/s41586-018-0043-0.

Liu, X., Fu, Y.-X., 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21, 1–9. https://doi.org/10.1186/s13059-020-02196-9.

Malinsky, M., Matschiner, M., Svardal, H., 2021. Dsuite-Fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21, 584–595. https://doi.org/10.1111/1755-0998.13265.

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: population inference from RADseq data. Mol. Biol. Evol. 35, 1284–1290.

https://doi.org/10.1093/molbev/msy023.

Mussmann, S.M., Douglas, M.R., Chafin, T.K., Douglas, M.E., 2020. A dmix P ipe: population analyses in A dmixture for non-model organisms. BMC Bioinform. 21, 1–9. https://doi.org/10.1186/s12859-020-03701-4.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7, e37135. https://doi.org/10.1371/journal.pone.0037135.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P., 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Syst. Biol. 55, 595–609. https://doi.org/10.1080/10635150600852011

Puillandre, N., Brouillet, S., Achaz, G., 2021. ASAP: assemble species by automatic partitioning. Mol. Ecol. Resour. 21, 609–620. https://doi.org/10.1371/journal.pgen.1002967.

Rambaut, A., Drummond, A., 2010. FigTree v1. 3.1 Institute of Evolutionary Biology. Univ. Edinburgh.

Rambaut, A., Drummond, A.J., 2015. LogCombiner v1. 8.2. LogCombinerv1 8, 656.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904. https://doi.org/10.1093/sysbio/syy032

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28, 4737–4754. https://doi.org/10.1111/mec.15253.

RStudio Team, n.d. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.

Swofford, D.L., Sullivan, J., 2003. Phylogeny inference based on parsimony and other methods using PAUP*. phylogenetic Handb. a Pract. approach to DNA protein phylogeny, cáp 7, 160–206.

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. Bioinform. 29, 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

Chapter III

Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *Ligula intestinalis*

Proceedings of the Royal Society B
DOI:10.1098/rspb.2023.2563

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Nazarizadeh M, Nováková M, Drábková M, Catchen J, Olson PD, Štefka J. 2024 Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *Ligula intestinalis*. *Proc. R. Soc. B* **291**: 20232563.
https://doi.org/10.1098/rspb.2023.2563

Received: 16 November 2023
Accepted: 2 January 2024

Subject Category:
Genetics and genomics

Subject Areas:
genomics, developmental biology, evolution

Keywords:
Cestoda, reference genome, transcriptome, life cycle, parasite

Author for correspondence:
Jan Štefka
e-mail: jan.stefka@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.7021305.

THE ROYAL SOCIETY PUBLISHING

Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *Ligula intestinalis*

Masoud Nazarizadeh[1,2], Milena Nováková[2], Marie Drábková[1], Julian Catchen[3], Peter D. Olson[4] and Jan Štefka[1,2]

[1]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
[2]Institute of Parasitology, Biology Centre CAS, České Budějovice, Czech Republic
[3]Department of Evolution, Ecology and Behavior, University of Illinois, Urbana-Champaign, IL 61801, USA
[4]Life Sciences Department, Natural History Museum, London, UK

MD, 0000-0003-1389-3605; PDO, 0000-0002-4183-4533; JŠ, 0000-0002-1283-9730

Ligula intestinalis (Cestoda: Diphyllobothriidae) is an emerging model organism for studies on parasite population biology and host–parasite interactions. However, a well-resolved genome and catalogue of its gene content has not been previously developed. Here, we present the first genome assembly of *L. intestinalis*, based on Oxford Nanopore Technologies, Illumina and Omni-C sequencing methodologies. We use transcriptome profiling to compare plerocercoid larvae and adult worms and identify differentially expressed genes (DEGs) associated with these life stages. The genome assembly is 775.3 mega (M)bp in size, with scaffold N50 value of 118 Mbp and encodes 27 256 predicted protein-coding sequences. Over 60% of the genome consists of repetitive sequences. Synteny analyses showed that the 10 largest scaffolds representing 75% of the genome display high correspondence to full chromosomes of cyclophyllidean tapeworms. Mapping RNA-seq data to the new reference genome, we identified 3922 differentially expressed genes in adults compared with plerocercoids. Gene ontology analyses revealed over-represented genes involved in reproductive development of the adult stage (e.g. sperm production) and significantly enriched DEGs associated with immune evasion of plerocercoids in their fish host. This study provides the first insights into the molecular biology of *L. intestinalis* and provides the most highly contiguous assembly to date of a diphyllobothriid tapeworm useful for population and comparative genomic investigations of parasitic flatworms.

1. Introduction

Parasites with multi-stage life cycles have evolved a number of mechanisms to facilitate transmission between the intermediate and definitive hosts and subsequent reproduction [1]. To understand parasite–host interactions on a molecular scale, it is necessary to develop genomic and transcriptomic resources that can elucidate the genetic underpinnings of the mechanisms associated with parasite adaptation. The genome of a parasite with a complex life cycle must contain genetic information that enables the parasite to survive and reproduce in vastly different environments with respect to metabolic requirements and strategies to evade the host immune system [2–6]. Therefore, comparative genomic analyses across several parasites can shed further light on the evolutionary adaptations [2–4,7,8]. Moreover, understanding the patterns of gene expression can provide key insights into the mechanisms underlying the regulation of physiological functions related to parasitic infection, within-host survival and reproduction [7]. Therefore, an integrated approach combining

both genomic and transcriptomic analyses promises to provide important insights into parasite biology and host–parasite interactions.

Significant strides in genomic research over the past decade have elucidated the genomic architecture of a broad spectrum of key species from the order Cyclophyllidea, predominantly the parasites of human or domestic animals [2,5,6,8–12]. Genomic and transcriptomic studies can reveal gene expression specific to each developmental stage in order to begin to understand the genes involved in parasitism and the features of parasitic life [3,5–7,13]. To date, genomic data exist for at least 17 cyclophyllidean tapeworms (National Center for Biotechnology Information (NCBI); https://www.ncbi.nlm.nih.gov/ and WormBase ParaSite), five of which have highly contiguous, chromosome-level or fully complete assemblies [8–10,12,14]. By contrast, genomic resources for diphyllobothriidean tapeworms remain limited, with only four draft genome assemblies currently available [4,15,16]. Despite these disparities, both tapeworm orders share several life-history characteristics, such as hermaphroditic adults parasitizing the intestines of vertebrates, and larval stages infecting invertebrate hosts. However, a stark ecological divergence is noticeable; whereas cyclophyllidean life cycles are typically bound to terrestrial ecosystems and hosts, the majority of diphyllobothriidean life cycles are associated with aquatic environments [17].

Ligula intestinalis is a common diphyllobothriidean cestode that has a nearly global distribution [18]. The parasite requires three hosts to complete its life cycle (electronic supplementary material, figure S1): a freshwater copepod as the first intermediate host, a wide array of fish species of the families Osmeriformes, Cypriniformes, Gobiiformes and Galaxiiformes as the second intermediate host, and a piscivorous bird as the definitive host [18]. The free-swimming coracidial larval stage of the parasite is consumed by planktonic copepods and develop into procercoid forms upon entering the haemocoel [19]. Once the copepod is consumed by fish, the procercoids mature into plerocercoid larvae in the abdominal cavity [18]. The parasites become sexually matured in the enteric tract of birds which consume the infected fish [18]. Compared with other tapeworms, its adult stage is very short and only lasts for several days during which the eggs are dispersed by infected birds. Larval worms can be induced to develop sexually *in vitro* [18]. The tapeworm has been extensively used in endocrinological, ecological, population genetic, phylogeographic and immunological studies [20–23].

The current study primarily aimed to provide the first highly contiguous genome assembly of *L. intestinalis*. Furthermore, we sought to investigate RNA transcription patterns to uncover the regulation of biological functions during the tapeworm's life cycle, by establishing a comprehensive reference transcriptome for both plerocercoid and adult stages of *L. intestinalis*. This research enriches our understanding of *L. intestinalis*' molecular biology, offering a platform for further exploration in the field of population genomics for this species, as well as in comparative genomic studies of parasitic flatworms.

2. Material and methods

(a) Sampling and culturing

A total of 16 plerocercoids were collected from 15 roach (*Rutilus rutilus*), caught from water bodies Želivka, Most and Klíčava in the Czech Republic according to previously described methods [23] (electronic supplementary material, table S1). To maintain uniform conditions and minimize variability in transcriptome expression, only 1-year old fish infected with a single tapeworm were selected from the Most and Klíčava water bodies (electronic supplementary material, figure S1). Whole-genome sequencing was conducted on one sample obtained from a roach carrying two *L. intestinalis* parasites. Detailed methodologies, including sample preparation and *in vitro* culturing conditions used to obtain adult tapeworms are available in the electronic supplementary material.

(b) DNA extraction and sequencing

High molecular weight (HMW) genomic DNA was extracted from plerocercoid larvae using a DNeasy tissue kit (Qiagen, Germany), with the DNA quantity measured by a Qubit DNA BR Assay Kit fluorometer (Invitrogen, USA) and integrity assessed via gel electrophoresis. For genome assembly, three strategies were employed. First, Oxford Nanopore sequencing yielded 7 574 939 reads (average read length 5699 bp, max. 123 333 bp, min. 5402 bp), totalling 43 giga (G)bp. Second, a library was constructed for shotgun sequencing with a Hyper Library kit (Kapa Biosystems, Roche, Basel, Switzerland) and sequenced using Illumina NovaSeq, resulting in 485 493 102 paired-end 150 bp reads, totalling 146 Gbp (details in electronic supplementary material, table S2). Finally, a chromosome conformation capture library was prepared, producing 106 Gbp by Illumina NovaSeq 150 bp sequencing, at approximately 24× coverage (see electronic supplementary material, table S2 for more details).

(c) RNA extraction and sequencing

We used RNA-seq to facilitate annotation of the coding regions of the genome and to investigate differences between plerocercoid and adult stages of the life cycle, as well as among three different regions of the worms. Samples were taken from the anterior, middle and posterior regions of three adult and three plerocercoid worms (18 samples in total). Total RNA extraction was done according to Chomczynski & Sacchi [24] by the acid guanidinium thiocyanate-phenol-chloroform method using Trizol reagent (Invitrogen, Carlsbad, CA, USA). Integrity and yield were, respectively, evaluated in an Agilent Bioanalyzer 2100 (Agilent Technologies, USA) and a spectrophotometer (NanoDrop 1000; Thermo Scientific, USA). RNA samples were commercially processed into complementary (c)DNA libraries and sequenced using Illumina Novaseq 150 PE read technology (Novogene, UK and GeneWiz, USA). The results were exported as FASTQ files.

(d) Genome assembly

As a first step in the genome assembly, all FASTQ ONT files were merged into a single file and Flye v2.4 [25] was used to assemble Oxford Nanopore Technologies (ONT) sequences. A total of 137 scaffolds were generated during the assembly process. The largest scaffold was 2 329 270 bp, the scaffold N50 was 320 313 bp and the mean coverage was 55×. The assembly was polished with Illumina 150 bp PE reads (60× coverage) using ntEdit [26]. Then, HiRise scaffolding (commercial service by Dovetail Genomics LLC) allowed for improvement of the scaffold contiguity using the Omni-C library. Prior to annotation, we excluded scaffolds with coverage < 10× and length < 500 bp that constituted less than 1% of our data. Moreover, their exclusion did not omit any contigs containing hits with the transcripts from *L. intestinalis*. We used Phyloligo version 1.0 [27] to check for contamination in the genome assembly. To evaluate the quality of the genome assembly, we used GenomeQC [28] and QUAST v4 [29]. Genome completeness was estimated via BUSCO v5 (Benchmarking Universal Single-Copy Orthologs), as well as searching for BUSCO groups

in the Metazoan database [30]. The whole-genome sequencing data from our study are accessible on GenBank, associated with the Bioproject numberPRJNA1055111.

(e) Prediction of coding genes

Repeat families were first identified and classified using RepeatModeler v2.0.1 [31], a package that uses RepeatScout v1.0.6 and RECON v1.08 to detect repeated sequences in genomes, with all analyses run using default parameters. The output from RepeatModeler was then processed using RepeatMasker v 4.1.0, allowing the repeated sequences to be identified, classified and masked in the assembly file. Upon completion of the repeat masking, an initial *ab initio* gene prediction model for *L. intestinalis* was generated, using coding sequences from *Dibothriocephalus latus*, *Schistocephalus solidus*, *Sparganum proliferum*, *Spirometra erinaceieuropaei* and *Taenia solium* as references. This gene prediction model was optimized through six rounds of prediction adjustments, made in AUGUSTUS v2.5.5 [32], using default parameters. Concurrently, a separate *ab initio* gene prediction model was trained using the same set of coding sequences in SNAP v2006-07-28 [33].

To further improve the accuracy of the gene predictions, RNA sequences were aligned to the assembled genome using the STAR aligner software v2.7 [34]. We used the two-pass mode in STAR for enhanced detection of splice junctions. Intron hints were also obtained using the bam2hints tool from AUGUSTUS, which were used to refine the RNA sequence alignment to our genome. Gene prediction was then performed in this repeat-masked reference genome, using a combination of AUGUSTUS, SNAP and MAKER. These programs used the intron-exon boundary hints from the RNA sequences to guide the prediction process. Swiss-Prot peptide sequences from the UniProt database and from *D. latus*, *S. solidus*, *S. proliferum*, *S. erinaceieuropaei* and *T. solium* were incorporated into the MAKER pipeline to inform peptide generation.

Genes that were not predicted by both AUGUSTUS and SNAP were excluded from the final gene set. The MAKER pipeline assigned Annotation Edit Distance (AED) scores to the predicted genes, with genes having an AED score below 0.5 deemed as highly supported predictions. Subsequently, BLAST searches were conducted in the UniProt database to gather further information about the function of our predicted genes. In parallel, tRNA prediction was carried out using tRNAscan-SE v 2.05 [35]. Lastly, TransposonPSI [36] was used to identify potential transposons within the predicted protein set and to discover regions of transposon homology within the genome sequence.

(f) Functional annotation of the gene models

We carried out the functional annotation of the predicted genes using the OmicsBox v3.1 software [37], through the BLASTx function [38]. We set an E-value cut-off of 10^{-5} and cross-referenced with multiple databases including EggNOG (orthologous groups of genes; [39], InterProScan [40], Swiss-Prot [41] and the NCBI non-redundant protein sequences database (Nr; [42])). Furthermore, to enhance the quality of our gene annotations, we individually compared the protein sequences using the Pfam database [43], setting a HMMER E-value cut-off at 10^{-5}.

(g) Mitochondrial genome

Reconstruction of the mitochondrial genome was done using Illumina reads in MITObim v.1.67 [44]. *Ligula intestinalis* (NC039445) mitochondrial genes were used as queries by BLASTX to identify mitochondrial fragments in the assembly. To extend the fragments, we mapped the Illumina short reads to the assembly in MITObim. The MITOS web server was used to annotate protein-coding, ribosomal RNA and tRNA genes. The Artemis genome annotation tool [45] was used to manually curate annotations and assemblies based on sequence similarity to other published mitogenomes.

(h) Orthologue identification, functional annotation and gene family evolution

Orthologous and duplicate genes were identified via phylogenetic analysis using OrthoFinder v2.5.2 [46]. Then, the proteomes predicted in the *L. intestinalis* genome were compared with those in other species of Diphyllobothriidea (*S. solidus* (NCBI, BioProject: PRJNA576252), *S. erinaceieuropaei* (NCBI, BioProject: PRJEB35375), *D. latus* (NCBI, BioProject: PRJEB1206)) and Cyclophyllidea (*Taenia multiceps* (NCBI, BioProject: PRJNA307624), *T. saginata* (NCBI, BioProject: PRJNA71493), *Echinococcus granulosus* (NCBI, BioProject: PRJNA754835), *Hymenolepis microstoma* (NCBI, BioProject: PRJEB124), *H. nana* (NCBI, BioProject: PRJEB508) and *H. diminuta* (NCBI, BioProject: PRJEB30942). For increased accuracy, DIAMOND blast (E-value $< 1 \times 10^{-5}$) in OrthoFinder was used [46]. The Markov Cluster algorithm (MCL) was used for sequence similarity and clustering with the default inflation rate. MAFFT v.7 [47] was employed for multiple protein sequence alignment using default settings (including the FFT-NS-2 algorithm for general-purpose alignments and a standard scoring matrix like BLOSUM62 for amino acid substitutions), and FastTree2 v2.1 [48] was applied for inference of maximum-likelihood gene trees for each orthogroup. OrthoFinder used a concatenated alignment of single copy orthogroups to infer a species tree with at most one gene per species. The tree was reconstructed using the Species Tree Root Inference from Duplication Events (STRIDE) algorithm [49].

To infer the putative functions of each orthologue, we extracted protein sequences from each orthologous gene cluster and performed a comprehensive functional annotation in the OmicsBox software. The BLASTp-fast algorithm was employed to identify the best matches within the NCBI's non-redundant (NR) database, and the InterProScan algorithm was used to predict potential protein domains and motifs. To ensure the reliability of our functional inferences during the annotation processes, we adhered to a stringent E-value cutoff of 10^{-5}. These results were subsequently merged with the gene ontology (GO) annotation and mapping data in OmicsBox. Consequently, we retrieved GO terms, classifying the predicted functions into biological processes, molecular functions and cellular components. These classifications were then mapped to their respective orthologous clusters.

The output from OrthoFinder was analysed to identify gene families. Using the previously generated species tree, we constructed a calibrated tree using the penalized likelihood method and the TN algorithm in the r8s software [50]. For time calibration, *T. multiceps* and *T. saginata* were chosen as calibration points, with a median time of 4.09 million years ago (Mya), as sourced from (http://www.timetree.org). We employed CAFE v5 [51] to investigate gene family expansion and contraction throughout the phylogeny. Furthermore, gene gain/loss rates were assessed by adjusting the lambda (representing the maximum-likelihood value for the birth and death parameter) for each branch. The optimal lambda was determined via iterative calculations. After analysing 1000 random samples with a *p*-value threshold set at 0.05, we observed significant variation in gene family expansions and contractions. Branches with considerable deviations were determined using the Viterbi algorithm in CAFE, also using a *p*-value threshold of 0.05.

(i) Synteny analyses

We compared synteny between the *Ligula* genome and two chromosome-level tapeworm assemblies available: *H. microstoma* (BioProject: PRJEB124) and *E. granulosus* (BioProject: PRJNA754835). Due to the fragmented state of the other Diphyllobothriidea

genomes, synteny analysis was performed with only the two most contiguous assemblies of *S. erinaceieuropaei*, and *S. proliferum* (table 1) using SyMAP v5.1.0 (Synteny Mapping and Analysis Program) [52,53] with default visualization settings. For the synteny analysis, we omitted all scaffolds shorter than 1 Mbp and then aligned sequences using the promer algorithm from MUMmer [54]. The raw anchors generated by MUMmer were pooled into (putative) gene anchors, followed by filtering with a reciprocal top-2 filter (top_n = 2).

(j) Differentially expressed genes

A total of 18 RNA-seq samples were used to compare transcriptomic differences between plerocercoid and adult worms, and among the anterior, middle and posterior regions of both life stages (three replicates for each body part and life stage). Low-quality reads were initially discarded from all samples using Trimmomatic v0.33 [55] with a quality threshold of 15, a sliding window of 4 : 20 and the provided Illumina adapters. Following trimming, paired-end reads (electronic supplementary material, table S1) were mapped to the genome of *L. intestinalis* using STAR in a two-pass mode to improve alignment around splice junctions. Using Samtools v1.10 [56], reads with a mapping quality below 20 that mapped to multiple loci were eliminated. The filtered Sequence Alignment Map (SAM) files were sorted and converted to Binary Alignment Map (BAM) format and the mapped reads counted for genomic features using featureCounts v1.06 in the Subread package [57]. Electronic supplementary material, table S3, provides mapped reads per sample replicate and gene model. The Trimmed Mean of M values (TMM) method was applied to normalize the raw counts. We used DiCoExpress R script [58] to estimate differential expression of genes between two life stages (plerocercoid and adult) and categorized them based on the contrasts between the different regions of the body. The R script uses generalized linear models in the package edgeR [59]. To identify differentially expressed genes (DEGs), we used criteria of p-value < 0.05 and absolute log-fold change > 1. In order to evaluate the general similarities and differences among all the transcriptomes, plotMDS function in the edgeR was used to plot the first two principal components.

(k) Gene ontology and enrichment analysis

The clusterprofiler package [60] in R was used to perform a hypergeometric test for functional annotation of GO. The Benjamini–Hochberg false discovery rate adjustment of the p-value was applied before comparisons were made in the enrichment analysis with a threshold value of 0.05. Dot plots and enrichment maps for over- and under-represented genes were then generated using the DOSE R package [61] to visualize the results. The outcome of the enrichment analysis was juxtaposed with the findings from the DicoExpress analysis. GOs were summarized using REVIGO for sets of DEGs with a threshold of 0.4 and SimRel as a measure of similarity [62].

3. Results

(a) Genome metrics

We sequenced genomic DNA and first generated a nanopore-based 778 Mbp assembly, which was then improved with Omni-C data (electronic supplementary material, table S2). Using this combination of sequencing technologies, a 775.3 Mbp draft genome was assembled, representing one of the largest genome estimates yet known for diphyllobothriidea tapeworms, which range from 531 to 796 Mbp. Table 1 compares the assembly and genome metrics of *Ligula* with four other genomes in the same tapeworm order (*S. proliferum*, *S. erinaceieuropaei*, *S. solidus* and *D. latus*) and two representative chromosome assemblies of Cyclophyllidea (*H. microstoma* and *E. granulosus*). The new *Ligula* genome had the best results to date for scaffold N50 (61 Mbp) and L50 (5) in comparison to the other four species, respectively. The GC content of the draft assembly is 44%, which is similar to those species (43%–45%). Out of 954 BUSCOs, 613 (64.2%) were complete, 89 (9.3%) were fragmented, 24 (2.5%) complete and duplicated, and 252 (26.5%) were missing from the assembled genome. A total of 61.1% of the genome was found to comprise repetitive sequences, including 7921 unique transposon protein sequences (table 1; electronic supplementary material, table S4).

(b) Orthologous genes and gene duplications

A total of 198 304 proteins from 11 cestode species were placed into 165 590 (83.15%) gene families in orthogroups and 32 714 (16.5%) unassigned singletons. Figure 3 depicts a species tree for five diphyllobothriidean and six cyclophyllidean species and demonstrates gene duplication events for each node and tree branch. *Ligula intestinalis* separated from *D. latus* with high support value and 767 gene duplication events. The *L. intestinalis* proteome (27 256 proteins) was encoded by 19 104 (70.8%) gene families (figure 1*a*). Among the proteins, 6677 (24.5%) were specific to the species or singletons. Also, the highest number of species-specific duplication and species-specific orthogroups were observed in the *Ligula* genome (figure 1*a*). Even though the *Ligula* genome shared 1829 orthogroups with all diphyllobothriid and cyclophyllid species, it is distinguished by its own set of orthogroups, exhibiting comparatively smaller intersections with other species (electronic supplementary material, figure S2). The unique orthologous cluster genes in *Ligula* were mostly annotated as DNA integration, Retrovirus-related Pol polyprotein from transposon, DNA binding and Probable RNA-directed DNA polymerase from transposon X-element (electronic supplementary material, table S5).

Gene family evolution analysis indicates that the *L. intestinalis* genome gained 2028 gene families while losing a significantly higher number of 3543 families. Additionally, the new genome has featured an expansion of 39 families and a contraction of 406 families. By contrast, *D. latus* and *S. solidus* demonstrated a heightened propensity for gene acquisition, reporting gains of 1546 and 1583 families, respectively, contrasted with fewer losses of 244 and 103 families. Similarly, both *S. erinaceieuropaei* and *S. proliferum* have exhibited significant gains, with respective numbers of 1301 and 950 gene families, tempered by losses of 264 and 289 gene families (figure 1*b*).

(c) Comparative synteny analysis between *Ligula* and other genomes

We analysed chromosome-level synteny among the 10 longest scaffolds of the *Ligula* assembly with the six chromosomes of *H. microstoma* and nine chromosomes of *E. granulosus*. Extensive preservation of synteny is evident when the first six scaffolds are compared with the chromosome assemblies of the two species. Large regions of these scaffolds are aligned with single chromosomes in both species. However, scaffolds 8 to 10 were predominantly mapped to chromosome six in *H. microstoma* and to

Table 1. Genome assembly metrics of *L. intestinalis* and four other tapeworms in the order Diphyllobothriidea, along with two representative complete genomes from the order Cyclophyllidea. n.a. data not analysed. Data for the new genome are in bold.

	Diphyllobothriidea					Cyclophyllidea	
assembly feature	*Sparganum proliferum*	*Spirometra erinaceieuropaei*	*Schistocephalus solidus*	*Diphyllobothrium latus*	*Ligula intestinalis*	*Hymenolepis microstoma*	*Echinococcus granulosus*
assembly version	v2.2	v2.0	NST_G2_0011_upd	Geneva_0011_upd	**V1**	NHM_v.3	Eg-G1s
source	Kikuchi *et al*. [4]	International Helminth genomes [15]			**present study**	Olson *et al*. [10]	Korhonen *et al*. [14]
genome size (Mbp)	653.4	796	539.4	531.4	**775.3**	168.9	173
GC content (%)	45	45	43	43	**44**	36	42
total scaffolds (excluding mitochondrial (mt)DNA)	7388	5723	56 778	140 294	**49 061**	6	31
longest scaffold (Mbp)	8.09	5.4	0.59	0.08	**118**	43	38.9
number of scaffolds>1 Mbp	203	200	0	0	**10**	6	11
total number of Ns (percentage)	7.80%	9.8%	4.09%	7.22%	**3.4%**	3.8%	2.32%
scaffold N50 (Mbp)	1.24	0.82	0.031	0.0067	**61**	25.6	18.6
scaffold N75 (Mbp)	0.56	0.38	0.014	0.0033	**18**	25	14
scaffold L50	146	271	4833	21 785	**5**	3	4
scaffold L75	333	625	11 174	50 033	**10**	5	6
number of gene models	22 427	28 290	20 228	19 966	**27 663**	10 139	9985
number of transposon proteins	4124	6128	1905	318	**7921**	n.a.	n.a.
average gene length (bp)	a	a	6612.8	2721.8	**4766.1**	1930	5727
number of exons	a	a	148 972	50 732	**101 613**	90 693	70 115
number of exons per gene model	a	a	7.4	2.8	**3.7**	7.9	7
average exon length	a	a	266.9	227.6	**302.7**	217	221
complete BUSCOs (single-copy and duplicated)	63.4% (61.4% and 2.0%)	62% (55.6% and 6.4%)	61.1% (60% and 1.1%)	34.7% (34.2% and 0.5%)	64.2% (61.7% and 2.5%)	77% (78% and 2.7%)	69.8% (69% and 0.8%)

[a]Data mis-reported in Kikuchi *et al*. [4] due to missing intron information.

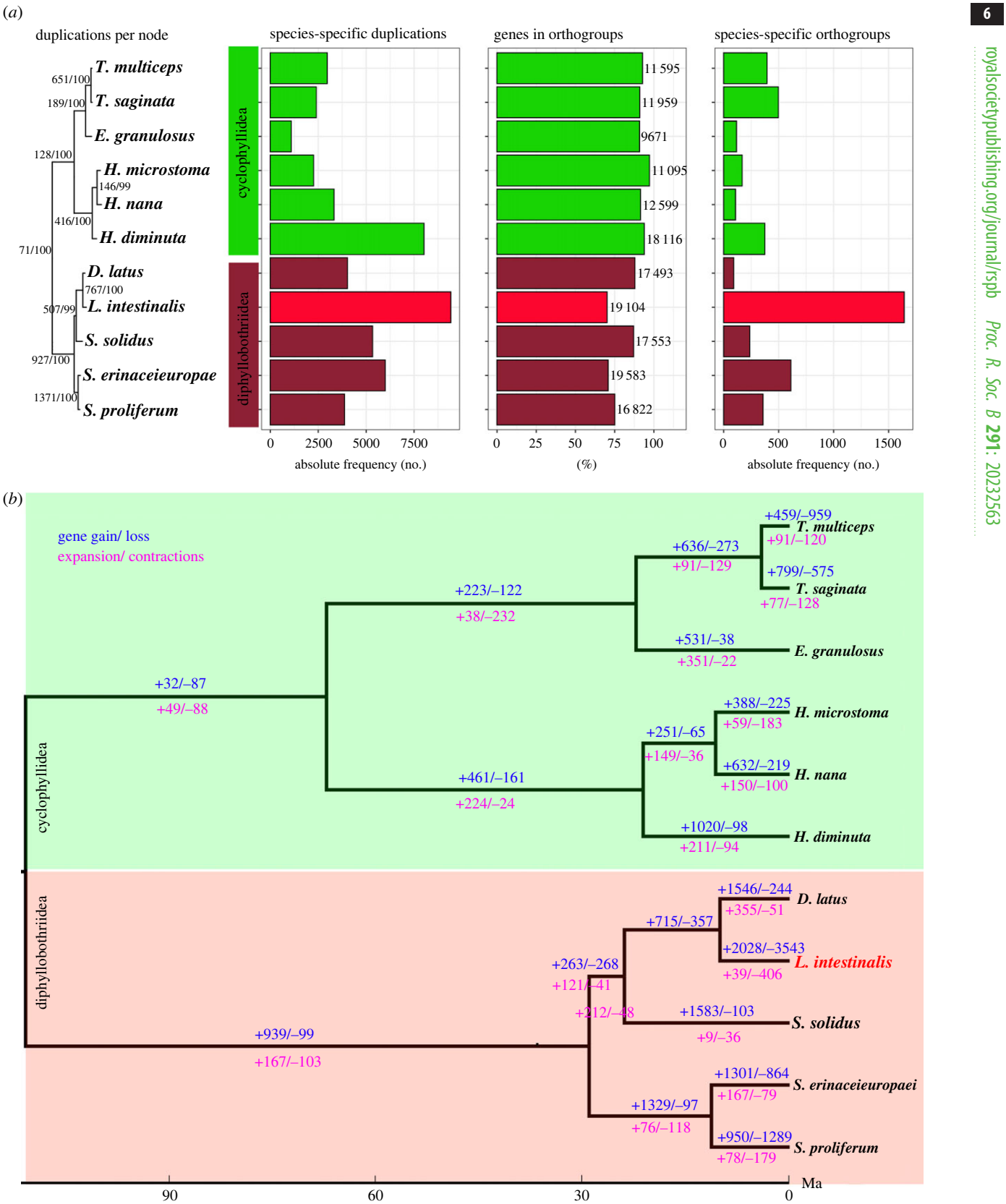


Figure 1. Gene duplication and orthology in diphyllobothriidean and cyclophyllidean cestodes. (*a*) Summary of OrthoFinder analysis on five available genomes of diphyllobothriidean species and six representative genomes of cyclophyllidean tapeworms. The species tree was reconstructed using a concatenated alignment of 226 single copy orthogroups, with the number of top branches corresponding to gene duplication events and bootstrap value. The horizontal bar plots show gene duplication events per species, percentage of genes in orthogroups, and the number of species-specific orthogroups. The tree is rooted with a cyclophyllidean tapeworm *E. granulosus*. (*b*) Gene family evolution highlighting gene gains and losses (above the branches), as well as expansions and contractions (below the branches) across the Cyclophyllidea and Diphyllobothriidea taxa.

chromosomes eight and nine in *E. granulosus* (figure 2*a*). We also mapped the first 30 scaffolds of *S. erinaceieuropaei* and *S. proliferum* to the 10 scaffolds of *L. intestinalis* (figure 2*b*). The 10 longest scaffolds cover 75% of the *Ligula* genome, whereas in *Sparaganum* and *Spirometra* only 60% and 45% genes were covered by their 10 longest scaffolds, respectively.

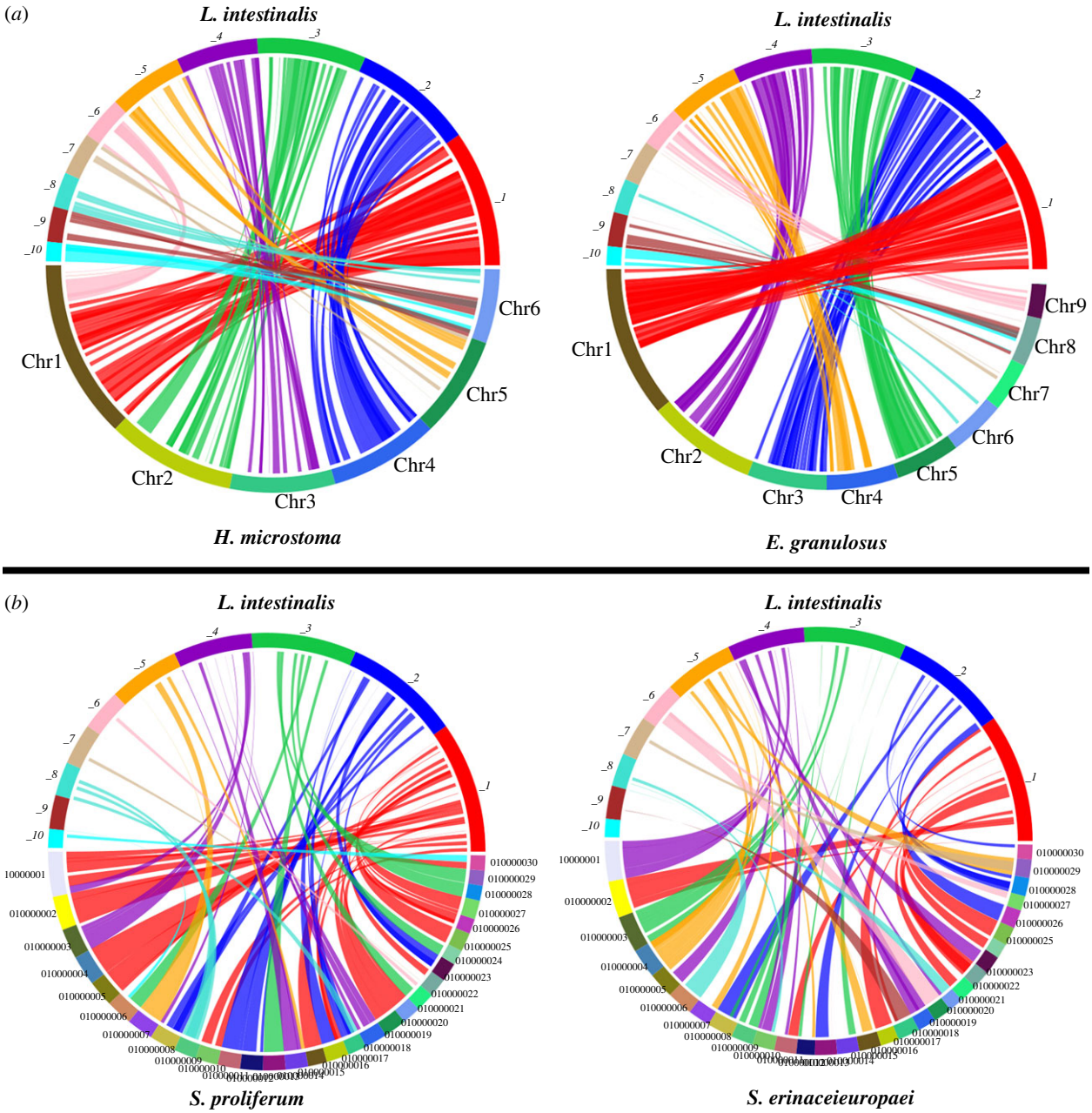


Figure 2. Syntenies of *L. intestinalis* and cestodes of the cyclophyllidean and diphyllobothriidean orders. (*a*) Chromosomal synteny between *L. intestinalis* and two representative complete genomes in Cyclophyllidea (*H. microstoma* and *E. granulosus*) is highly conserved. (*b*) Synteny plots between *L. intestinalis* and *S. proliferum* and *S. erinaceieuropaei* genomes demonstrate higher contiguity of the *L. intestinalis* genome.

We also compared the 10 longest scaffolds of the *Ligula* genome with 200 and 203 scaffolds of *S. erinaceieuropaei* and *S. proliferum*, respectively (electronic supplementary material, figure S3, all scaffolds longer than 1 Mbp). Analysis of pairwise collinearity of orthologous blocks on the scaffolds showed a stronger similarity between *L. intestinalis* and *S. proliferum* (342 blocks and 16 089 anchors) than between *L. intestinalis* and *S. erinaceieuropaei* (240 blocks and 10 657 anchors). The longest scaffold (118 Mbp) in the *L. intestinalis* genome corresponds to 41 and 42 scaffolds in *S. erinaceieuropaei* and *S. proliferum*, respectively.

(d) Differential gene expression and gene ontology patterns

Our transcriptome analyses identified 3922 DEGs between the two life stages (electronic supplementary material, figure S4, figure 3*a* and table S6), comprising 2301 upregulated and 1621 downregulated genes in the plerocercoids. Interestingly, RNA expression patterns revealed that larvae and adults clearly diverged from each other by the first dimension of PCA (52.76% variance explained), and the central part of larval worms separated from its posterior and anterior parts by the second dimension of PCA (9.9%) (figure 3*c*). Moreover, in analysis of the relationship between different parts of the body and life stage, the central and posterior parts of larvae and adults demonstrated 1856 DEGs (1270 upregulated and 586 downregulated) and 1573 DEGs (945 upregulated and 628 downregulated), respectively (electronic supplementary material, figure S4 and table S6).

Upregulated and downregulated DEGs were analysed using the GO enrichment method to identify the characteristics of the DEGs between adults and plerocercoids. Results using both DiCoExpress and clusterprofiler were similar, however, DiCoExpress not only encompassed all the enriched genes identified by clusterprofiler but also provided additional

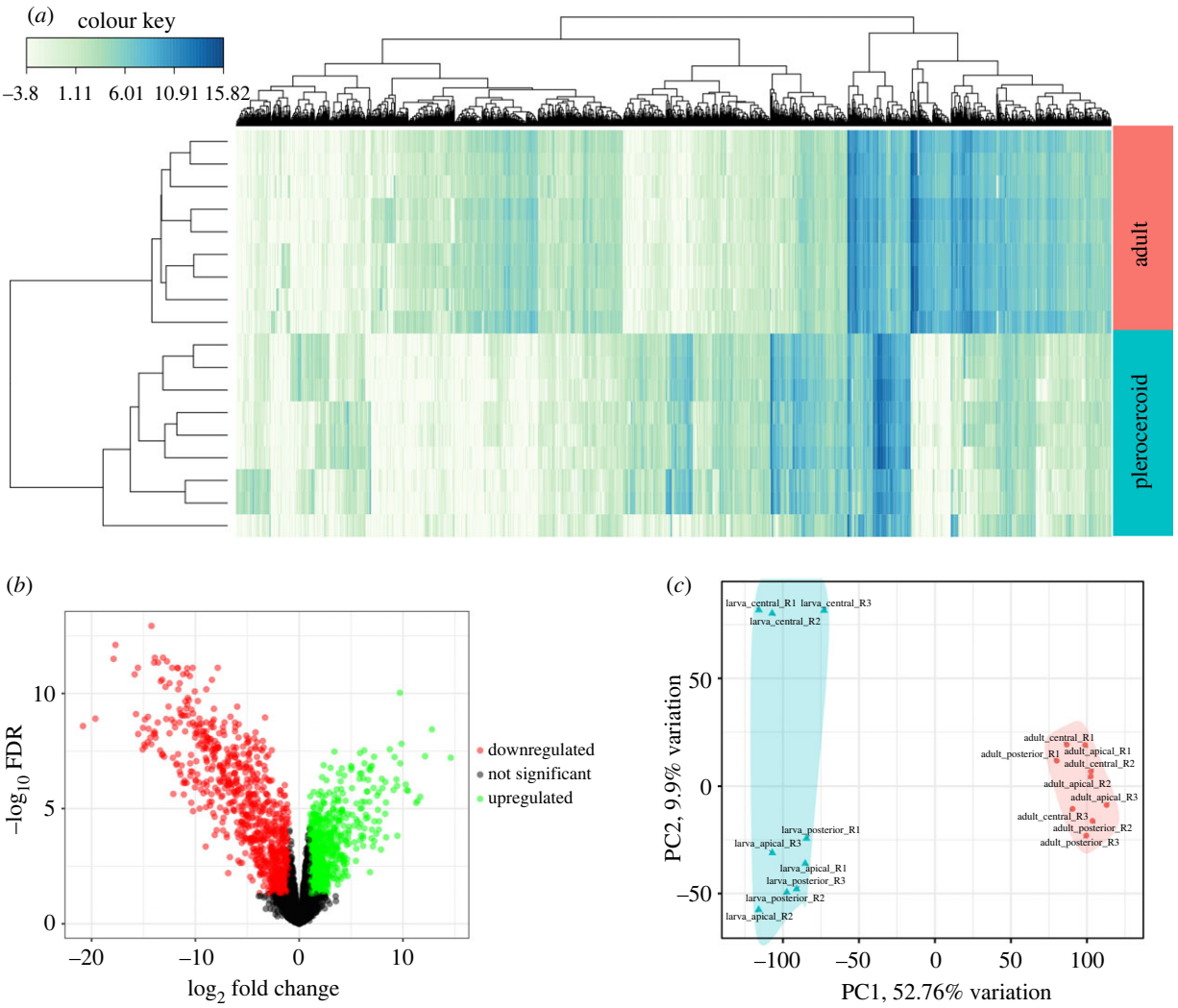


Figure 3. Differential gene expression in *L. intestinalis*. (*a*) Heatmap of the differentially expressed (DE) genes between plerocercoids and adults. (*b*) Volcano plot of upregulated/downregulated and non-DE transcripts between plerocercoids and adults. Red and green dots correspond to genes with significantly differentially expressed transcripts, while transcripts whose expression levels were not significant are depicted as blue dots. (*c*) Principal component analysis for clustering of sample replicates with relative variances across PC1 and PC2. Larval and adult samples are colour coded as in plot *a*.

insights by revealing an increased quantity of over-represented and under-represented genes (electronic supplementary material, table S7 and table S8).

The enrichment analysis from clusterprofiler revealed that the biological process of the 275 over-represented DEGs were mainly enriched in the cytoskeleton organization (94 DEGs), cilium organization (82 DEGs), microtubule-based movement (70 DEGs), microtubule-based process (12 DEGs), axoneme assembly (9 DEGs) and organelle assembly (8 DEGs) (figure 4). Moreover, the cellular components of the upregulated genes revealed 262 DEGs corresponding to cilium (104 DEGs), cytoskeleton (79 DEGs), cytoplasm (55 DEGs), microtubule cytoskeleton (16 DEGs) and cytoplasmic region (8 DEGs). In addition, molecular function was associated with 123 upregulated DEGs, which mainly included protein binding (62 DEGs), cytoskeletal motor activity (26 DEGs), structural constituent of cytoskeleton (11 DEGs), calcium ion binding (12 DEGs) and guanosine-5′-triphosphate (GTP) binding (12 DEGs) (figure 4). The 59 downregulated DEGs in biological processes were mainly involved in monoatomic cation transmembrane transport (8 DEGs). The corresponding cellular component of downregulated DEGs included extracellular space (28 DEGs), caveola (7 DEGs), membrane (54 DEGs) and cellular anatomical entity (16 DEGs). Finally, the molecular function of downregulated DEGs revealed over-represented genes in actin binding, transmembrane signalling receptor activity, serine-type endopeptidase activity, protein binding, monoatomic ion transmembrane transporter activity, inorganic anion transmembrane transporter activity and gated channel activity (figure 4).

4. Discussion

Hybrid assembly facilitates more accurate and contiguous genome assemblies, which allow for improved estimates of the genome content and the ability to analyse syntenic relationships and other features of genome architecture [10]. Recent genomic studies on cestodes have used multiple sequencing technology platforms to improve previously published genomes (e.g. [7,13,14,16,22]). In this study, the *L. intestinalis* genome was constructed using a hybrid assembly, and comparative genomics was conducted using all other available genomes representing the Diphyllobothriidea and six genomes representing the Cyclophyllidea. This study revealed substantial conservation of synteny between the first six scaffolds of the *Ligula* genome- and chromosome-level assemblies of *H. microstoma* and *E. granulosus*. In addition, the 10 longest

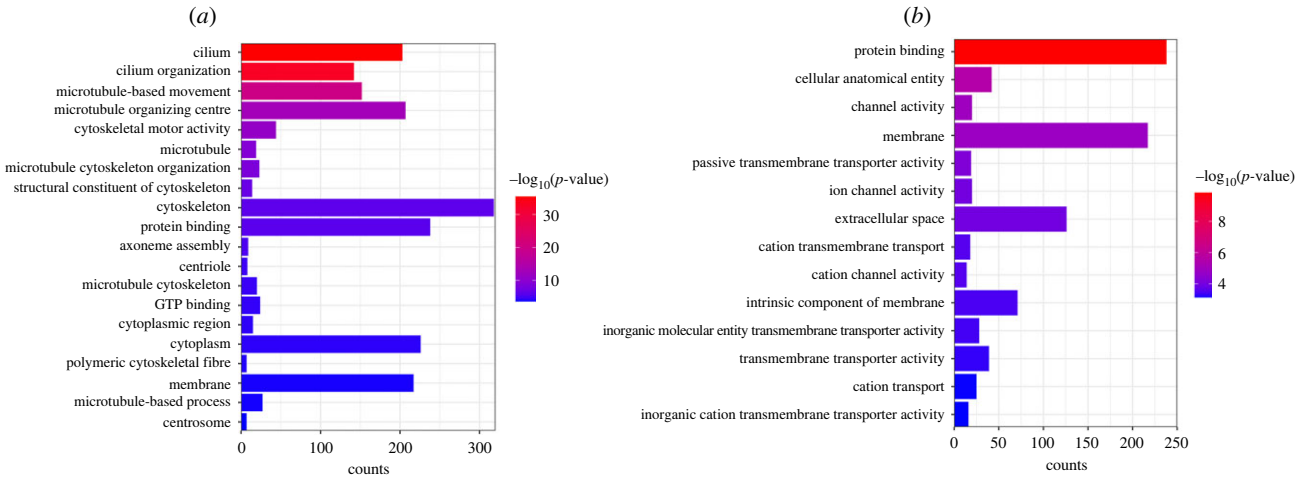


Figure 4. GO enrichment analysis of DEGs in larvae and adults. The bar charts present the most significant GO terms (with an false discovery rate (FDR) less than 0.05) associated with genes over-represented in adults (*a*) and under-represented in larvae (*b*).

scaffolds constitute 75% of the total genome assemblies and 66% of the total number of genes. Assembly quality metrics (NX and LX) indicate high contiguity for the new *Ligula* genome in comparison to other diphyllobothriidean genomes. Consequently, the *L. intestinalis* genome acquired in this study allows for a more reliable estimate of the genome's content and size for this tapeworm.

(a) Genome size, repetitive content and gene duplications

In the newly assembled genome, we discovered a substantial proportion of repetitive sequences (electronic supplementary material, table S4; 61.1%), which aligns with previous findings of high repetitive sequence content in three other diphyllobothriidean species (*D. latus*, *S. proliferum* and *S. erinaceieuropaei*). In addition, sequenced diphyllobothriidean species consistently have a larger size and number of species-specific genes compared with sequenced cyclophyllidean tapeworms [4,10]. In addition to the increased repeat content inflating the genome size in Diphyllobothriidea, our study and the previous studies have shown that diphyllobothriideans have experienced more expansions or fewer contractions of gene families compared with cyclophyllidean tapeworms [4,11]. These factors could explain the observed differences in genome size between these two major tapeworm groups and suggest that cyclophyllidean genomes have become streamlined.

Reconstruction of the phylogenetic relationships among the five diphyllobothriidean species recovered a generally well-supported tree that is congruent with a previous genomic study not including *L. intestinalis* [4]. The finding of *L. intestinalis* and *D. latus* being recovered as sister taxa to each other and in turn sister clades to *S. solidus* is in line with a study based on more samples but fewer genetic loci [63]. In accordance with the phylogeny, the highest number of genes in orthologues was found between *L. intestinalis*, *D. latus* and *S. solidus*. Higher numbers of orthologous genes indicate a more recent common ancestor and a shorter divergence time, whereas lower numbers indicate a more ancient common ancestor and a longer divergence time [64]. The results indicated a relatively high frequency of gene duplication during the divergence of *L. intestinalis* and *D. latus*, as well as between *S. proliferum* and *S. erinaceieuropaei*. Such frequent gene duplication events possibly played a major role in the speciation processes [65]. Additionally, *L. intestinalis* exhibits a high number of species-specific duplications (figure 1). Although the implication of these duplications remains to be tested, it could be suggested that it allows *L. intestinalis* a greater potential for functional diversification and host specialization [66] compared with the other four species.

The results of the current study reveal a distinct genomic evolutionary pattern for *L. intestinalis* within the Diphyllobothriidea. Interestingly, the newly sequenced *Ligula* genome lacks orthogroups commonly found in other tapeworms, instead possessing its own set of gene orthogroups. This observation is in line with findings from gene family evolution, where significant gene loss was noted in this genome. Such gene loss is not only a random by-product of evolutionary drift; it can also act as a significant driver for species diversification and adaptation [67]. Additionally, although some level of gene loss could be attributed to potential incomplete annotation (and we acknowledge that further work on the *Ligula* genome is desirable), we also found that the genome contains a notable set of unique orthologous related to DNA integration and retrovirus-related Pol polyprotein, distinguishing it from other members of both the Diphyllobothriidea and Cyclophyllidea. This finding is in line with the high volume of repetitive DNA in the *Ligula* genome, and it is further supported by our transcriptome data, which indicates that DNA integration is significantly over-represented in the larval stage of the tapeworm (electronic supplementary material, table S7). Occurrence of retrovirus genes might be a consequence of horizontal gene transfer events from its cypriniform fish hosts, which are known to have genomes significantly invaded by retrovirus genes [68]. In a similar way, virus or virus-like particles resembling RNA viruses have been observed in various parasitic groups, from protozoans to nematodes [69,70]. Retroviruses are highly abundant in the genomes of eukaryotes [71] and can add novel genes and regulatory elements to the host's repertoire. For example, endogenous retroviruses may have shaped aspects of the innate immune system, such as the regulation of interferons [72]. How *Ligula* has been able to withstand the pressure on its genome integrity from such massive and ongoing integration of mobile elements is a question that merits further investigation

(b) Transcriptomics and stage-specific gene expression

We conducted the first comprehensive analysis of the plerocercoid and adult stages of *L. intestinalis* using RNA sequencing methods. Results show that more than 80% of all transcriptome reads from both life stages could be assigned to the new reference genome. This indicates that the RNA-seq data obtained are of high quality and can accurately quantify the expression level of genes [73]. Moreover, the results revealed different transcriptome patterns between the two life stages and identified numerous potential DEGs associated with various biological processes, cellular components and molecular functions. Interestingly, enrichment results showed that although *L. intestinalis* has the shortest adult life stage (3–7 days) among pseudophyllidean tapeworms, enabled by a plerocercoid stage that develops to the size of an adult inside the fish intermediate host [18,19]; it then undergoes significant physical and biochemical changes during the transition from the larval to adult phase, similar to other tapeworm species (e.g. [6,7,10,17,74]). These changes include modulation of gene expression associated with various biological tasks such as reproduction, development and host survival.

Transcriptome results also revealed that the central body part of the larva possesses a relatively distinct transcriptome profile compared with its apical and posterior parts. Further enrichment analysis unveiled an abundance of genes related to essential metabolic processes, such as monosaccharide metabolism, pyruvate metabolism, glycolysis and energy reserve metabolism (electronic supplementary material, table S8). These processes are crucial for energy production and maintenance [75] and are probably associated with energy accumulation of the larval stage. The larval body is filled with glycogen, required for a subsequent very fast transformation into an adult in its final host (hours to days) and the production of gametes. Variations observed within the GO processes mentioned above might suggest differential energy needs or nutrient absorption efficiencies across different regions of the tapeworm.

The results of our enrichment analysis revealed an abundance of over-represented genes associated with cilia and microtubules in adult tapeworms (electronic supplementary material, tables S7, S8 and figure 4). In agreement with the present study, Preza *et al*. [13] showed significantly over-represented cilia and flagellum genes in a rodent tapeworm (*Hymenolepis microstoma*). These structures are essential for sperm production. Cilia, hair-like structures, facilitate sperm movement within the reproductive system [76], while microtubules provide structural support and contribute to the formation and organization of the sperm tail or axoneme [77]. Correspondingly, our results also revealed several significantly enriched genes directly related to adult tapeworm sperm production (e.g. axoneme assembly, sperm motility, microtubule-based movement and microtubule-based processes) (figure 4; electronic supplementary material, table S7). Spermatogenesis is characterized by the formation of a flagellum that is integrated into the cytoplasmic extension of the spermatid, eventually leading to a mature spermatozoon. The spermatozoon has unique features, such as a single spiral anterior body, a cortical microtubule, and two axonemes with different lengths (the 9 + '1' and 9 + 9 + 2 patterns), which distinguish the mature spermatozoon of *L. intestinalis* from other tapeworm species [78,79]. Therefore, studies of the candidate genes involved in axoneme assembly and microtubule function in *L. intestinalis* may provide valuable insights for evolutionary developmental relationships among tapeworms.

In addition, we discovered a prominent pattern characterized by a significant enrichment of the cytoplasmic region (electronic supplementary material, table S7 and figure 5; cytoplasm, cytoskeleton, Golgi cisternae membrane, dynein complex) as a cellular component during the development of adult tapeworms. It is known that cytoplasm plays a crucial role in reproduction, nutrient uptake and energy production in tapeworms [74,80–82]. Consistent with our findings, high levels of expressed transcripts have been reported for cytoskeletal proteins in multiple cestode species, particularly in diphyllobothriidean cestodes [6,8,83,84]. The cytoplasm of vitellocytes houses organelles such as ribosomes, the rough endoplasmic reticulum, and the Golgi apparatus, which are involved in the synthesis, modification and packaging of yolk proteins. In accordance with that, Yoneva *et al*. [74] demonstrated that vitellocytes undergo significant cytoplasmic changes during vitellogenesis in *L. intestinalis* as they produce yolk components.

The present study also identified several over-represented genes associated with osmoregulatory balance in adult *L. intestinalis* (electronic supplementary material, tables S7–S8). Osmoregulation is critical for many parasitic flatworms that live in their host's intestine because they lack a digestive system and must therefore absorb nutrients from the intestinal contents of the host [85]. However, *L. intestinalis* uses the host intestine mainly as a means of sexual maturation and egg dispersal, while it is not thought to absorb nutrients during its short-lived adult stage [18]. On the other hand, possible differences in osmotic pressure between the intermediate host environment (plerocercoids in fish body cavity) and the bird intestine could explain the transcriptional differences between plerocercoids and adults even without the need for nutritional intake of the adult. Lastly, we should note that the *in vitro* conditions used to raise adults in this study may have resulted in an unnatural osmotic stress in the tapeworm. Consequently, further research is needed to conclusively determine the osmotic balance of *Ligula* tapeworms, with an emphasis on understanding the molecular and cellular mechanisms involved in osmoregulation.

Recent transcriptome analyses of several tapeworm species have identified an increased number of upregulated genes related to immune evasion in the larval stage compared with the adult [4–6,84]. Similarly, we identified DEGs significantly enriched in immune evasion of the plerocercoid larvae (electronic supplementary material, table S8). Consequently, combating the host immune system is critical for the parasite's survival and establishment in the fish intermediate host. Phylogeographic studies have revealed several evolutionary lineages of *L. intestinalis*, with each lineage exhibiting a high degree of host specificity for particular fish species groups [20,22,86]. Although the precise immune response of *L. intestinalis* is still unclear, strategies to evade the immune system may play a crucial role in the host specificity of different *Ligula* lineages and thus may evolve via directional selection. Focusing on sequential differences between evolutionary *Ligula* lineages in the DEGs identified here could help understanding the process of host-specialization assumed to drive speciation in *Ligula* [20,22,86].

In conclusion, the high-quality genome and transcriptomic profiling of the larvae and adults provided by this study can serve as a starting point for research targeted at developmental biology, host–parasite coevolution and immunogenetics in *L. intestinalis* and other diphyllobothriidean tapeworms. Although the current assembly with its 10 longest scaffolds covering 75% of the genome length represents the most contiguous genome available for any non-cyclophillidean tapeworm, focusing on a full chromosome-level assembly in future studies can potentially enhance the quality of the genome.

Ethics. Parasite samples were obtained from fish sampled in the frame of a hybrobiological research approved by the Animal Welfare Committee of the Biology Centre CAS according to § 16a of the Act No. 246/1992 Coll., on the protection of animals against cruelty, as amended.

Data accessibility. The whole-genome sequencing data from the study are accessible on GenBank, under the Bioproject no. PRJNA1055111.

Supplementary material is available online [87].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. M.N.: conceptualization, formal analysis, investigation, methodology, visualization, writing—original draft, writing—review and editing; M.N.: formal analysis, investigation, writing—review and editing; M.D.: formal analysis, investigation, writing—review and editing; J.C.: investigation, supervision, writing—review and editing; P.D.O.: methodology, supervision, writing—review and editing; J.S.: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. We are grateful for the computational resources provided by the e-INFRA CZ project (ID: 90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic. This research was made possible through the support of a grant from the Czech Science Agency (grant no. GA19-04676S).

Acknowledgements. The authors thank Anna Mácová and Roman Hrdlička for their invaluable assistance during the fieldwork. We thank anonymous reviewers for their comments and suggestions.

Reference

1. Auld SKJR, Tinsley MC. 2015 The evolutionary ecology of complex lifecycle parasites: linking phenomena with mechanisms. *Heredity (Edinb).* **114**, 125–132. (doi:10.1038/hdy.2014.84)
2. Wang S *et al.* 2016 Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. *Nat. Commun.* **7**, 1–12. (doi:10.1038/ncomms12845)
3. Olson PD *et al.* 2018 Genome-wide transcriptome profiling and spatial expression analyses identify signals and switches of development in tapeworms. *Evodevo* **9**, 1–29. (doi:10.1186/s13227-018-0110-5)
4. Kikuchi T *et al.* 2021 Genome of the fatal tapeworm *Sparganum proliferum* uncovers mechanisms for cryptic life cycle and aberrant larval proliferation. *Commun. Biol.* **4**, 649. (doi:10.1038/s42003-021-02160-8)
5. Li W-H, Yang Y, Zhang N-Z, Wang J-K, Liu Y-J, Li L, Yan H-B, Jia W-Z, Fu B. 2021 Comparative transcriptome analyses of the developmental stages of *Taenia multiceps*. *Front. Vet. Sci.* **8**, 677045. (doi:10.3389/fvets.2021.677045)
6. Liu SN, Su XY, Chen WQ, Yu JW, Li JR, Jiang P, Cui J, Wang ZQ, Zhang X. 2022 Transcriptome profiling of plerocercoid and adult developmental stages of the neglected medical tapeworm *Spirometra erinaceieuropaei*. *Acta Trop.* **232**, 106483. (doi:10.1016/j.actatropica.2022.106483)
7. Hébert FO, Grambauer S, Barber I, Landry CR, Aubin-Horth N. 2016 Transcriptome sequences spanning key developmental states as a resource for the study of the cestode *Schistocephalus solidus*, a threespine stickleback parasite. *Gigascience* **5**, s13742-016. (doi:10.1186/s13742-016-0128-3)
8. Tsai IJ *et al.* 2013 The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **496**, 57–63. (doi:10.1038/nature12031)
9. Li W *et al.* 2018 The genome of tapeworm *Taenia multiceps* sheds light on understanding parasitic mechanism and control of coenurosis disease. *DNA Res.* **25**, 499–510. (doi:10.1093/dnares/dsy020)
10. Olson PD, Tracey A, Baillie A, James K, Doyle SR, Buddenborg SK, Rodgers FH, Holroyd N, Berriman M. 2020 Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. *BMC Biol.* **18**, 1–16. (doi:10.1186/s12915-020-00899-w)
11. Liu G-H *et al.* 2021 *Dipylidium caninum* draft genome - a new resource for comparative genomic and genetic explorations of flatworms. *Genomics* **113**, 1272–1280. (doi:10.1016/j.ygeno.2021.02.019)
12. Pu L *et al.* 2022 A chromosome-level genome assembly for the rabbit tapeworm *Taenia pisiformis*. *Gene* **834**, 146650. (doi:10.1016/j.gene.2022.146650)
13. Preza M, Calvelo J, Langleib M, Hoffmann F, Castillo E, Koziol U, Iriarte A. 2021 Stage-specific transcriptomic analysis of the model cestode *Hymenolepis microstoma*. *Genomics* **113**, 620–632. (doi:10.1016/j.ygeno.2021.01.005)
14. Korhonen PK *et al.* 2022 Chromosome-scale *Echinococcus granulosus* (genotype G1) genome reveals the Eg95 gene family and conservation of the EG95-vaccine molecule. *Commun. Biol.* **5**, 199. (doi:10.1038/s42003-022-03125-1)
15. International Helminth Genomes Consortium. 2017 Comparative genomics of the major parasitic worms International Helminth Genomes Consortium. *bioRxiv* 51, 236539. (doi:10.1101/236539)
16. Berger CS, Laroche J, Maaroufi H, Martin H, Moon K-M, Landry CR, Foster LJ, Aubin-Horth N. 2021 The parasite *Schistocephalus solidus* secretes proteins with putative host manipulation functions. *Parasit. Vectors* **14**, 436. (doi:10.1186/s13071-021-04933-w)
17. Caira JN, Jensen K. 2017 *Planetary biodiversity inventory (2008–2017): tapeworms from vertebrate bowels of the earth*. Natural History Museum, University of Kansas, pp. 463. Lawrence, KS, USA.
18. Dubinina MN. 1980 Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR. New Delhi, India: Amerind Publishing Company.
19. Loot G, Francisco P, Santoul F, Lek S, Guégan J-F. 2001 The three hosts of the *Ligula intestinalis* (Cestoda) life cycle in Lavernose-Lacasse gravel pit, France. *Fundam. Appl. Limnol.* **152**, 511–525. (doi:10.1127/archiv-hydrobiol/152/2001/511)
20. Nazarizadeh M *et al.* 2023 Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite – The *Ligula intestinalis* species complex. *Mol. Phylogenet. Evol.* **180**, 107677. (doi:10.1016/j.ympev.2022.107677)
21. Bouzid W, Lek S, Mace M, Ben Hassine O, Etienne R, Legal L, Loot G. 2008 Genetic diversity of *Ligula intestinalis* (Cestoda: Diphyllobothriidea) based on analysis of inter-simple sequence repeat markers. *J. Zool. Syst. Evol. Res.* **46**, 289–296.
22. Štefka J, Hypša V, Scholz T. 2009 Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: microsatellite study of *Ligula intestinalis* (Cestoda). *Mol. Ecol.* **18**, 1187–1206. (doi:10.1111/j.1365-294X.2008.04074.x)
23. Nazarizadeh M *et al.* 2022 Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies. *Folia Parasitol. (Praha).* **69**, 18. (doi:10.14411/fp.2022.018)
24. Chomzynski P. 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162**, 156–159. (doi:10.1006/abio.1987.9999)
25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019 Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546. (doi:10.1038/s41587-019-0072-8)

26. Warren RL *et al.* 2019 ntEdit: scalable genome sequence polishing. *Bioinformatics* **35**, 4430–4432. (doi:10.1093/bioinformatics/btz400)
27. Mallet L, Bitard-Feildel T, Cerutti F, Chiapello H. 2017 PhylOligo: a package to identify contaminant or untargeted organism sequences in genome assemblies. *Bioinformatics* **33**, 3283–3285. (doi:10.1093/bioinformatics/btx396)
28. Manchanda N, Portwood JL, Woodhouse MR, Seetharam AS, Lawrence-Dill CJ, Andorf CM, Hufford MB. 2020 GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. *BMC Genomics* **21**, 193. (doi:10.1186/s12864-020-6568-2)
29. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013 QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075. (doi:10.1093/bioinformatics/btt086)
30. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021 BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654. (doi:10.1093/molbev/msab199)
31. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020 RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457. (doi:10.1073/pnas.1921046117)
32. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006 AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439. (doi:10.1093/nar/gkl200)
33. Korf I. 2004 Gene finding in novel genomes. *BMC Bioinf.* **5**, 1–9. (doi:10.1186/1471-2105-5-59)
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. (doi:10.1093/bioinformatics/bts635)
35. Chan PP, Lin BY, Mak AJ, Lowe TM. 2021 tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096. (doi:10.1093/nar/gkab688)
36. Haas B. 2007 *TransposonPSI: an application of PSI-blast to mine (retro-) transposon ORF homologies*. Cambridge, MA: Broad Institute.
37. BioBam. 2019 OmicsBox-Bioinformatics made easy. *March* **3**, 2019, https://www.biobam.com/omicsbox.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
39. Huerta-Cepas J *et al.* 2019 EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314. (doi:10.1093/nar/gky1085)
40. Jones P *et al.* 2014 InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240. (doi:10.1093/bioinformatics/btu031)
41. Bateman A *et al.* 2023 UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531. (doi:10.1093/nar/gkac1052)
42. Pruitt KD, Tatusova T, Maglott DR. 2007 NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65. (doi:10.1093/nar/gkl842)
43. Finn RD *et al.* 2016 The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285. (doi:10.1093/nar/gkv1344)
44. Hahn C, Bachmann L, Chevreux B. 2013 Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129. (doi:10.1093/nar/gkt371)
45. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012 Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469. (doi:10.1093/bioinformatics/btr703)
46. Emms DM, Kelly S. 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. (doi:10.1186/s13059-019-1832-y)
47. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
48. Price MN, Dehal PS, Arkin AP. 2010 FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490. (doi:10.1371/journal.pone.0009490)
49. Emms DM, Kelly S. 2017 STRIDE: species tree root inference from gene duplication events. *Mol. Biol. Evol.* **34**, 3267–3278. (doi:10.1093/molbev/msx259)
50. Sanderson MJ. 2003 r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302.
51. Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020 CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518.
52. Soderlund C, Nelson W, Shoemaker A, Paterson A. 2006 SyMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Res.* **16**, 1159–1168. (doi:10.1101/gr.5396706)
53. Soderlund C, Bomhoff M, Nelson WM. 2011 SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* **39**, e68. (doi:10.1093/nar/gkr123)
54. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004 Versatile and open software for comparing large genomes. *Genome Biol.* **5**, 1–9. (doi:10.1186/gb-2004-5-2-r12)
55. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. (doi:10.1093/bioinformatics/btu170)
56. Danecek P *et al.* 2021 Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008. (doi:10.1093/gigascience/giab008)
57. Liao Y, Smyth GK, Shi W. 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. (doi:10.1093/bioinformatics/btt656)
58. Lambert I, Paysant-Le Roux C, Colella S, Martin-Magniette M-L. 2020 DiCoExpress: a tool to process multifactorial RNAseq experiments from quality controls to co-expression analysis through differential analysis based on contrasts inside GLM models. *Plant Methods* **16**, 68. (doi:10.1186/s13007-020-00611-7)
59. Robinson MD, McCarthy DJ, Smyth GK. 2010 edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. (doi:10.1093/bioinformatics/btp616)
60. Wu T *et al.* 2021 clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141. (doi:10.1016/j.xinn.2021.100141)
61. Yu G, Wang L-G, Yan G-R, He Q-Y. 2015 DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609. (doi:10.1093/bioinformatics/btu684)
62. Supek F, Bošnjak M, Škunca N, Šmuc T. 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800. (doi:10.1371/journal.pone.0021800)
63. Waeschenbach A, Brabec J, Scholz T, Littlewood DTJ, Kuchta R. 2017 The catholic taste of broad tapeworms – multiple routes to human infection. *Int. J. Parasitol.* **47**, 831–843. (doi:10.1016/j.ijpara.2017.06.004)
64. Koonin EV. 2005 Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**, 309–338. (doi:10.1146/annurev.genet.39.073003.114725)
65. Zhang J. 2003 Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18**, 292–298. (doi:10.1016/S0169-5347(03)00033-8)
66. Lynch M, Conery JS. 2000 The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155. (doi:10.1126/science.290.5494.1151)
67. Albalat R, Cañestro C. 2016 Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391.
68. Wang J, Han G-Z. 2023 Genome mining shows that retroviruses are pervasively invading vertebrate genomes. *Nat. Commun.* **14**, 4968. (doi:10.1038/s41467-023-40732-w)
69. Mueller JF, Strano AJ. 1974 *Sparganum proliferum*, a Sparganum infected with a virus? *J. Parasitol.* **60**, 15. (doi:10.2307/3278671)
70. Esch GW, Fernández JC. 1993 *A functional biology of parasitism*. Dordrecht, Netherlands: Springer Netherlands. (doi:10.1007/978-94-011-2352-5)
71. Finnegan DJ. 2012 Retrotransposons. *Curr. Biol.* **22**, R432–R437. (doi:10.1016/j.cub.2012.04.025)
72. Chuong EB, Elde NC, Feschotte C. 2016 Regulatory evolution of innate immunity through co-option of

endogenous retroviruses. *Science* **351**, 1083–1087. (doi:10.1126/science.aad5497)

73. Conesa A *et al.* 2016 A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13. (doi:10.1186/s13059-016-0881-8)
74. Yoneva A, Scholz T, Bruňanská M, Kuchta R. 2015 Vitellogenesis of diphyllobothriidean cestodes (Platyhelminthes). *C. R. Biol.* **338**, 169–179. (doi:10.1016/j.crvi.2015.01.001)
75. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J. 1999 Glycobiology of Protozoal and Helminthic Parasites. In *Essentials of glycobiology*. New York, NY: Cold Spring Harbor Laboratory Press.
76. Biserova NM, Gordeev II. 2010 Fine structure of nervous system in plerocercoid *Ligula intestinalis* (Cestoda: Diphyllobothriidea). *Invertebr. Zool.* **7**, 133–154. (doi:10.15298/invertzool.07.2.05)
77. Justine J-L, Briand MJ, Bray RA. 2012 A quick and simple method, usable in the field, for collecting parasites in suitable condition for both morphological and molecular studies. *Parasitol. Res.* **111**, 341–351. (doi:10.1007/s00436-012-2845-6)
78. Ehlers U. 1984 *Phylogenetisches system der Plathelminthes*. Munich, Germany: Urban & Fischer.
79. Levron C, Sitko J, Scholz T. 2009 Spermiogenesis and spermatozoon of the yapeworm *Ligula intestinalis* (Diphyllobothriidea): phylogenetic implications. *J. Parasitol.* **95**, 1–9. (doi:10.1645/GE-1646.1)
80. Smyth JD, McManus DP. 1989 *The physiology and biochemistry of cestodes.* Cambridge, UK: Cambridge University Press. (doi:10.1017/CBO9780511525841)
81. Yoneva A, Kuchta R, Scholz T. 2014 First study of vitellogenesis of the broad fish tapeworm *Diphyllobothrium latum* (Cestoda, Diphyllobothriidea), a human parasite with extreme fecundity. *Parasitol. Int.* **63**, 747–753. (doi:10.1016/j.parint.2014.07.002)
82. Yoneva A, Scholz T, Młocicki D, Kuchta R. 2015 Ultrastructural study of vitellogenesis of *Ligula intestinalis* (Diphyllobothriidea) reveals the presence of cytoplasmic-like cell death in cestodes. *Front. Zool.* **12**, 35. (doi:10.1186/s12983-015-0128-7)
83. García-Montoya GM, Mesa-Arango JA, Isaza-Agudelo JP, Agudelo-Lopez SP, Cabarcas F, Barrera LF, Alzate JF. 2016 Transcriptome profiling of the cysticercus stage of the laboratory model *Taenia crassiceps*, strain ORF. *Acta Trop.* **154**, 50–62. (doi:10.1016/j.actatropica.2015.11.001)
84. Basika T *et al.* 2019 Transcriptomic profile of two developmental stages of the cestode parasite *Mesocestoides corti*. *Mol. Biochem. Parasitol.* **229**, 35–46. (doi:10.1016/j.molbiopara.2019.02.006)
85. Dold C, Holland C V. 2011 Ascaris and ascariasis. *Microbes Infect.* **13**, 632–637. (doi:10.1016/j.micinf.2010.09.012)
86. Bouzid W, Štefka J, Hypša V, Lek S, Scholz T, Legal L, Hassine OB, Loot G. 2008 Geography and host specificity: two forces behind the genetic structure of the freshwater fish parasite *Ligula intestinalis* (Cestoda: Diphyllobothriidae). *Int. J. Parasitol.* **38**, 1465–1479. (doi:10.1016/j.ijpara.2008.03.008)
87. Nazarizadeh M, Nováková M, Drábková M, Catchen J, Olson J, Štefka J. 2024 Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *ligula intestinalis*. Figshare. (doi:10.6084/m9.figshare.c.7021305)

Supporting Information

Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *Ligula intestinalis*

Masoud Nazarizadeh[1,2], Milena Nováková[2], Marie Drábková[1], Julian Catchen[3], Peter D. Olson[4], Jan Štefka[1,2*]

[1]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
[2]Institute of Parasitology, Biology Centre CAS, České Budějovice, Czech Republic
[3]Department of Evolution, Ecology and Behavior, University of Illinois, Urbana-Champaign, IL 61801, USA
[4]Life Sciences Department, Natural History Museum, London, United Kingdom

Corresponding author: jan.stefka@gmail.com

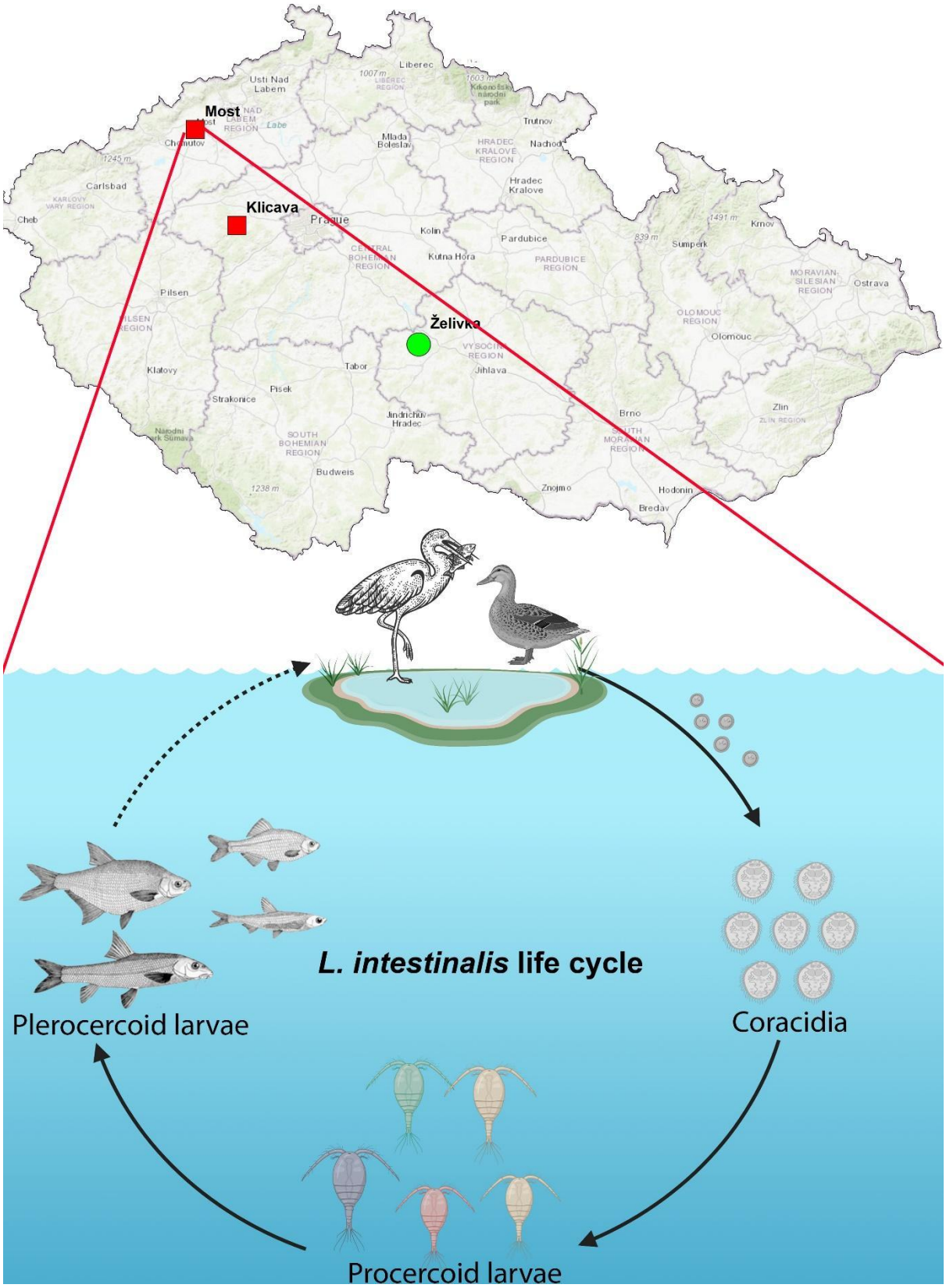


Figure S1 Sampling localities in the Czech Republic and the life cycle of *L. intestinalis*

Tapeworm sampling and *in vitro* culturing

We sampled 16 plerocercoids from 15 roach (*Rutilus rutilus*) captured from water bodies Želivka, Most, and Klíčava, Czech Republic. In order to ensure consistent transcriptome expression, our selection focused on 1-year old fish, each infected with a single tapeworm. The sample from a roach in the Želivka dam, carrying *Ligula intestinalis*, was specifically used for whole genome resequencing. This sample was carefully rinsed with physiological saline solution (0.9% sodium chloride) and subsequently stored at -80 °C. We performed transcriptome sequencing on sections taken from the anterior, central, and posterior parts of three living plerocercoids. In an aseptic environment, these worms were removed from the fish, rinsed with Ultra Pure DNAse/RNase-free water (Ambion, Austin, Texas, USA), and then sectioned into 5x5 mm squares. These sections were immersed in RNAlater (Ambion, Austin, Texas, USA) at 4 °C overnight and preserved at −80 °C prior to RNA extraction. For culturing adult worms in the lab, 12 plerocercoids were transferred in physiological saline solution. The worms were paired and housed in six bottles, each containing 250 ml of a specially prepared culture medium. This medium comprised 25 mM HEPES buffer (Sigma, Taufkirchen, Germany), Minimum Essential Medium with Earle's salts, additives (1 ml/litre antibiotic/antimycotic solution (Sigma), 6.5 g D-glucose), and L-glutamine. We adjusted the medium's pH to 7.5 using NaOH. The cultures were incubated at 40 °C with constant agitation at 70 rpm until all six pairs of parasites successfully produced eggs. Three adult parasites were then randomly selected from different pairs, and samples from their anterior, central, and posterior parts were taken for RNA analysis, following the protocol used for plerocercoids.

Table S1 Samples utilised for transcriptome data.

Sample ID	Locality	Life Stage	Part of Body	Number of Reads	Fq Accession number
9KLA	Klíčava	Larva	Anterior	31,770,990	SAMN38982002
9KLC			Central	36,322,323	SAMN38982003
9KLP			Posterior	32,642,797	SAMN38982004
22KLA			Anterior	28,399,001	SAMN38982005
22KLC			Central	37,722,618	SAMN38982006
22KLP			Posterior	34,884,550	SAMN38982007
35MA	Most		Anterior	27,541,654	SAMN38982008
35MC			Central	31,174,775	SAMN38982009
35MP			Posterior	27,251,031	SAMN38982010
14M21A		Adult	Anterior	42,185,310	SAMN38982011
14M21C			Central	33,164,231	SAMN38982012
14M21P			Posterior	30,911,458	SAMN38982013
13M21A			Anterior	29,430,746	SAMN38982014
13M21C			Central	36,953,181	SAMN38982015
13M21P			Posterior	28,870,701	SAMN38982016
12M21A			Anterior	30,109,372	SAMN38982017
12M21C			Central	29,886,337	SAMN38982018
12M21P			Posterior	25,167,799	SAMN38982019

High molecular weight DNA extraction and sequencing

The methodology for extracting HMW genomic DNA from plerocercoid larvae followed manufacturer instructions using a DNeasy tissue kit. The amount of extracted DNA was measured by a fluorometer (Qubit DNA BR Assay Kit, Invitrogen, USA), and DNA integrity assessed via gel electrophoresis. For Oxford Nanopore sequencing, the preparation and commercial sequencing of the library was performed at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, which covered converting HMW gDNA into four Nanopore libraries using the 1D (SQK-LSK109) library kit, loading onto a FLO-MIN106 R9.4.1 SpotON flow cell, and sequencing in the GridIONx5 sequencer (Oxford Nanopore Technologies). The albacore algorithm (read_fast5_basecaller.py, ONT Albacore Sequencing Pipeline Software v2.3.1) was used for base calling all FAST5 files, resulting in 7,574,939 reads (average read length 5,699 bp, max. 123,333 bp, min. 5,402 bp) totalling 43 Gb of data (electronic supplementary material, table S2). To correct ONT reads another library using the same HMW sample as above was constructed for shotgun sequencing with the Hyper Library construction kit from Kapa Biosystems (Roche, Basel, Switzerland) and sequencing was performed using an Illumina NovaSeq at the Roy J. Carver Biotechnology Center of the University of Illinois at Urbana-Champaign. This sequencing yielded 485,493,102 paired-end 150 bp reads for a total of 146 Gbp (electronic supplementary material, table S2). To further improve genome contiguity, we produced a chromosome conformation capture library. An Omni-C proximity ligation library was prepared commercially (Dovetail Genomics, USA) using a flash frozen *L. intestinalis* sample. The library was sequenced using Illumina NovaSeq 150bp (106 Gbp, approximately 24X coverage) by the same commercial laboratory (electronic supplementary material, table S2).

Table S2 Statistics of sequencing data generated in this study.

	Nanopore	Novaseq	Omni-C
Total number of reads	7,574,939	485,493,102	106,428,637
Average read length	5,699	150	150
GC %	45	44	44
Assembly size,bp	780,205,246	471,887,469	778,975,446
Depth of Coverage (X)	55	60	54
Number of scaffolds	137	81,578	61,514
Longest Scaffold, bp	2,329,270	143,614	1,726,504
Scaffold N50 size, bp	320,313	9,840	146,819

Table S4 Analysis of repeat content in the *Ligula* genome.

Category	Number of elements	Length occupied	percentage of sequences
Retroelements	**450036**	**264202183 bp**	**34.07 %**
Penelope	227627	86470978 bp	11.15 **%**
LINEs	397285	233051484 bp	30.05 %
L2/CR1/Rex	124069	112872190 bp	14.55%
R1/LOA/Jockey	475	511392 bp	0.07 %
R2/R4/NeSL	66	94388 bp	0.01 %
RTE/Bov-B	33424	23638860 bp	3.05 %
LTR elements	52751	31150699 bp	4.02%
BEL/Pao	5776	3525641 bp	0.45 %
Gypsy/DIRS1	42744	27127907 bp	3.50 %
DNA transposons	**51119**	**14165910 bp**	**1.83%**
hobo-Activator	3898	1759147 bp	0.23%
Tc1-IS630-Pogo	14059	4797296 bp	0.62 %
Small RNA	**15184**	**2946277 bp**	**0.38%**
Low complexity	**3913**	**233542 bp**	**0.03%**
Simple repeat	**53777**	**3099135 bp**	**0.40%**
Unclassified	**823742**	**189184304 bp**	**24.40%**
Total	1397771	473831351 bp	61.1%

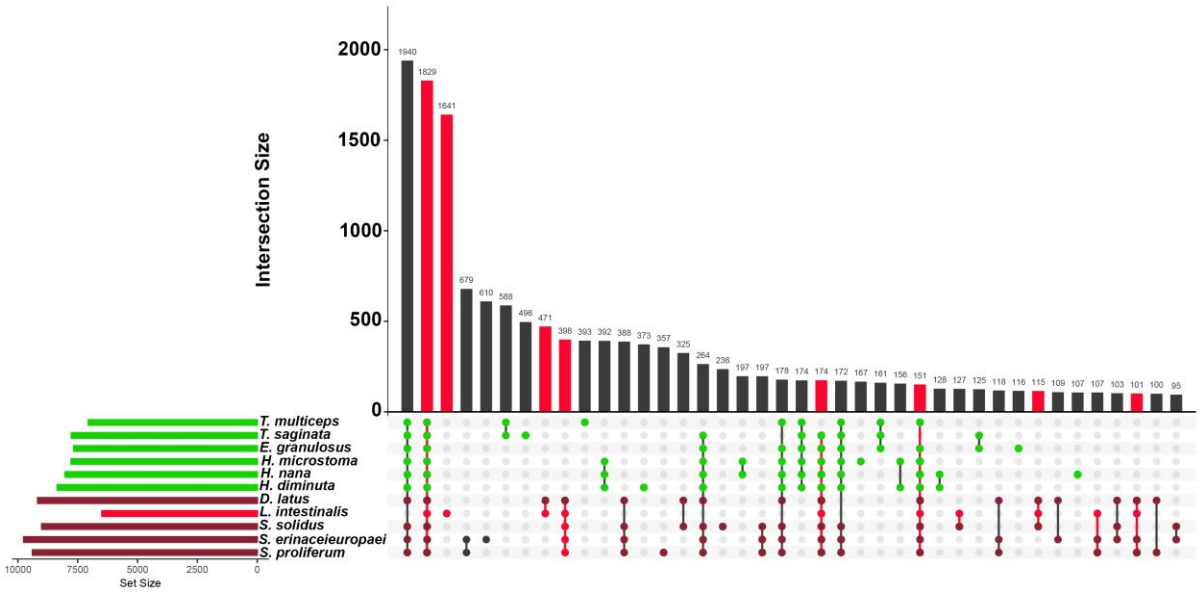


Figure S2 UpSet plot of the intersection of gene families in diphyllobothriidean and cyclophyllidean tapeworms.

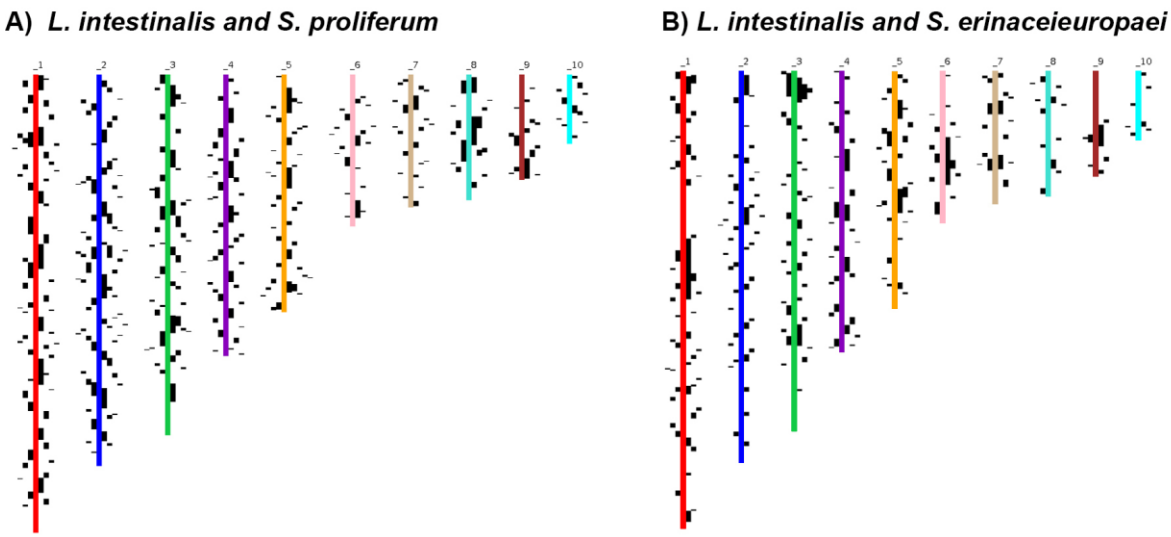


Figure S3 Synteny and collinearity bar plot between 10 longest scaffolds of *Ligula* genome with 200 and 203 scaffolds of *S. erinaceieuropaei* and *S proliferum*, respectively. Collinear gene blocks generated using Symap between genome scaffolds (>1 Mb) represent 18424, 8994, and 9324 genes for *L. intestinalis*, *S. proliferum* and *S. erinaceieuropaei*, respectively.

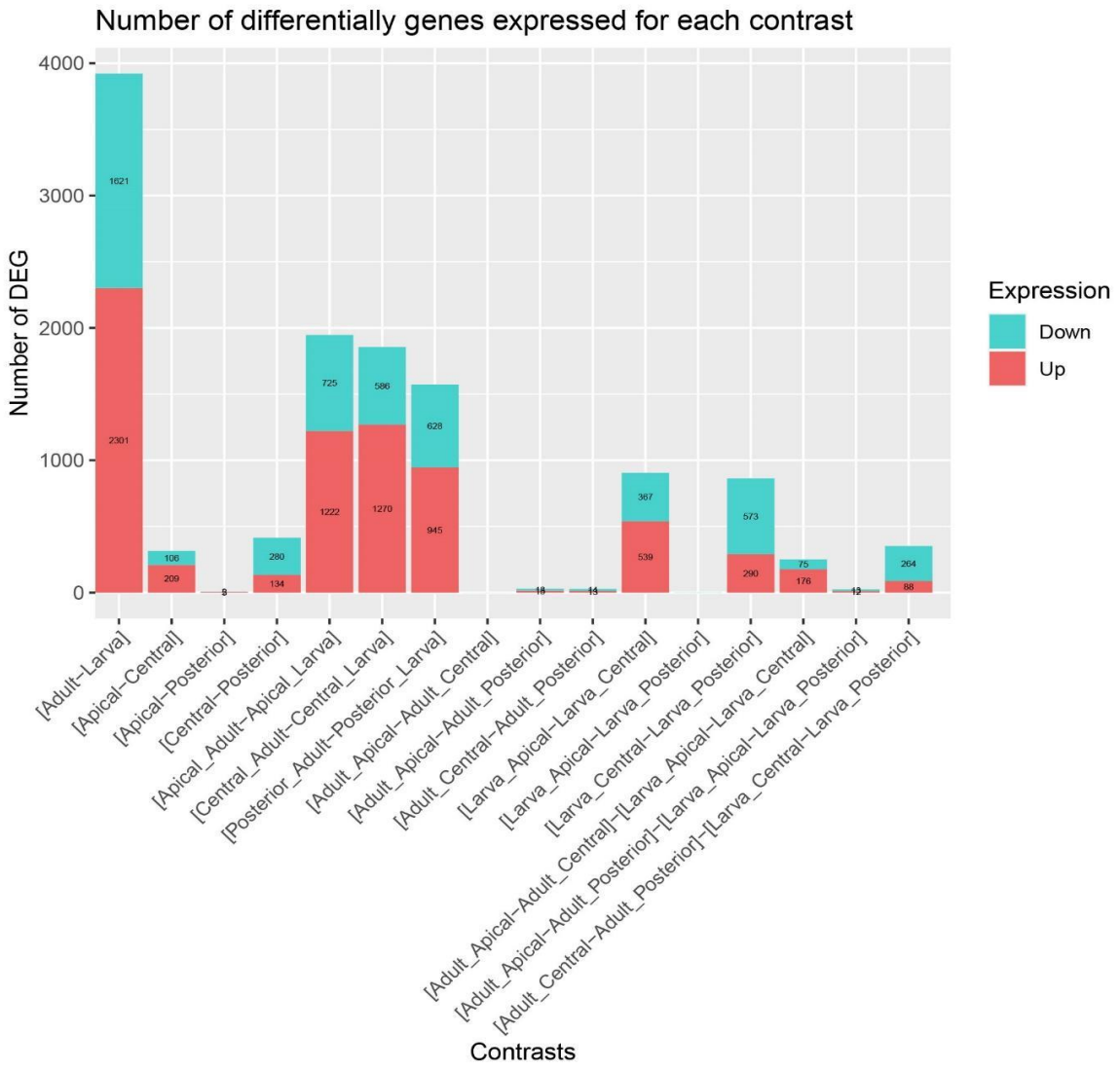


Figure S4 the number of up and down regulated genes between life stage (Adult and Larva) and different parts of body (Apical, central, and posterior).

Chapter IV

Host-associated genetic differentiation in the face of ongoing gene flow: ecological speciation in a pathogenic parasite of freshwater fish

The Manuscript is available as a preprint form
(DOI: 10.1101/2024.01.27.577373)

Host-associated genetic differentiation in the face of ongoing gene flow: ecological speciation in a pathogenic parasite of freshwater fish

Masoud Nazarizadeh[1,2], Milena Nováková[2], Jakub Vlček[2], Jan Štefka[1,2*]

[1] Faculty of Science, University of South Bohemia in České Budějovice, CZ

[2] Institute of Parasitology, Biology Centre, CAS, České Budějovice, CZ

[*]corresponding author: jan.stefka@gmail.com

Abstract

Adaptive evolution in response to varying environments, which leads to population divergence, is one of the key processes of natural selection. However, its effectiveness amidst ongoing gene flow remains controversial. Our study explores this phenomenon by focusing on a tapeworm parasite (*Ligula intestinalis*), which is capable of parasitising a wide spectrum of fish species, overcoming their immunological defences, and having a highly pathogenic impact. We analysed the population genetic structure, the degree of gene flow, and the level of genomic divergence between sympatrically occurring parasites from different cyprinid fish hosts. Utilising genome-wide Single Nucleotide Polymorphisms (SNPs) and transcriptome data, we investigated whether individual host species impose selection pressures on the parasite populations. Genetic clustering analyses indicated a divergence between the parasites infecting breams and those in roaches, bleaks, and rudds. Historical demography modelling suggested that the most plausible scenario for this divergence is isolation with continuous gene flow. Selection analysis identified 896 SNPs under selection, exhibiting higher nucleotide diversity and genetic divergence compared to neutral loci. Transcriptome profiling corroborated these results, revealing distinct gene expression profiles for the two parasite populations. An in-depth examination of the selected SNPs and differentially expressed genes revealed specific genes and their physiological functions, as candidates for the molecular mechanisms of immune evasion and, thus, for driving ecological speciation in the parasite. This study exhibits the interplay between host specificity, population demography, and disruptive selection in ecological speciation. By dissecting the genomic factors at play, we gain a better understanding of the mechanisms facilitating population divergence in the presence of gene flow.

Keywords: Ecological speciation, Isolation with continuous gene flow, differential gene expression, *Ligula intestinalis*

1 Introduction

The study of ecological speciation facilitates a comprehensive understanding of the complex interplay between evolutionary mechanisms and ecological interactions (Feder et al., 2012; Nosil et al., 2012; Rundle and Nosil, 2005). This form of speciation occurs when divergent selection pressures, associated with environmental factors, induce reproductive isolation (Nosil, 2012). In allopatric situations, reproductive isolation may arise due to accumulation of genetic variation, independently of selection pressures and adaptation to local conditions (Howard and Berlocher, 1998; Orr and Smith, 1998; Schluter, 1996). Conversely, in sympatric scenarios where populations occupy the same geographic area, reproductive isolation typically stems from reinforcement processes dampening gene flow between emerging population lineages (Bush, 1994).

Despite growing research in ecological speciation, our understanding of this process in parasites with complex life cycles remains limited (Brunner and Eizaguirre, 2016; Henrich and Kalbe, 2016; Poulin, 2011). These parasites present unique research opportunities due to their multi-host life cycles, which expose them to a variety of ecological pressures (Combes, 2001; Poulin, 2011). The complexity in their life cycles arises as these parasites must adapt to different hosts, each presenting a distinct set of ecological challenges, such as varied immune responses, habitats, and behaviors (Auld and Tinsley, 2015). Interestingly, the unique ecological conditions of parasites, such as their strict habitat selection and potential for intra-host speciation, might make sympatric speciation a more common mode of speciation for parasites compared to free-living animals (Le Gac and Giraud, 2004; McCoy, 2003). Recent studies revealed that speciation by disruptive selection for host choice plays a crucial role in sympatric speciation in parasites (de Meeûs et al., 1995; Duffy et al., 2008; Thaenkham et al., 2022). Over time, this selective pressure could result in the emergence of specialized subpopulations of parasites, each being adapted to a distinct host. Eventually, these subpopulations may become reproductively isolated, leading to sympatric speciation (McCoy, 2003). Furthermore, a hard selective sweep occurs in response to various selective pressures, such as host immune responses, drug treatments, or competition among parasites for host resources (Ebert and Fields, 2020; Le Pennec et al., 2023). When a beneficial mutation arises that enhances the parasite's ability to cope with one of these pressures, that allele might rapidly become predominant in the population. This rapid adaptation could lead to the formation of a specialized subpopulation of parasites, attuned to a specific host or set of conditions (Ebert and

Fields, 2020). Despite advances in genomic tools, a gap persists in our understanding of the interplay of these mechanisms in the context of complex life cycles and the precise genomic factors driving parasite diversification.

Ligula intestinalis (Cestoda) is a widespread diphyllobothriidean tapeworm and a quintessential model organism for studying the complex dynamics of ecological speciation (Bouzid et al., 2008; Hoole et al., 2010; Nazarizadeh et al., 2023, 2022; Štefka et al., 2009). It possesses a complex life cycle, beginning with the hatching of eggs into coracidia larvae in freshwater. These larvae then infect copepods and transform into procercoid larvae. Subsequently, after being ingested by their second intermediate hosts, the planktivorous fish, the larvae undergo further development and eventually mature into the plerocercoid stage within the fish's gut. The life cycle culminates in a piscivorous bird, the final host, where the tapeworm reaches maturity (Dubinina, 1980). The plerocercoid stage is a critical phase in the parasite's life cycle, marking a period of significant growth and development. By infecting numerous species of freshwater fish, plerocercoids encounter a range of ecological variables and host-specific selective pressures, such as different immune responses and hormonal manipulation of the host, which is castrated by the parasite (Dubinina, 1980; Halimi et al., 2013; Williams and Hoole, 1995). Previous genetic studies found different species of the tapeworm, sometimes co-occurring in the same water bodies (Bouzid et al., 2008; Nazarizadeh et al., 2023; Olson et al., 2002; Štefka et al., 2009). For instance, Olson et al. (2002) identified significant genetic variations in *Ligula* populations between two sympatric fish hosts in Lough Neagh, Northern Ireland. Furthermore, recent phylogeographic research has revealed that this parasite has undergone multiple modes of speciation, including both sympatric and allopatric, resulting in its diversification into at least ten distinct evolutionary lineages across various biogeographical realms (see Figure S1 and figure 2 in Nazarizadeh et al., 2023). This ecological complexity and diverse host range provide an ideal setting for studying how speciation occurs in response to ecological factors, particularly given the parasite's ability to adapt to different host species.

Of all ten known evolutionary lineages, Lineage A is particularly noteworthy due to its wide distribution in western Palaearctic and wide fish host spectrum, including cyprinids such as freshwater bream (*Abramis brama*), white bream (*Blicca bjoerkna*), roach (*Rutilus rutilus*), rudd (*Scardinius erythrophthalmus*), bleak (*Alburnus alburnus*), minnow *(Phoxinus phoxinus*), chub (*Squalius cephalus*), and crucian carp (*Carassius carassius*) (Bouzid et al., 2008; Štefka

et al., 2009; Nazarizadeh et al., 2023). Interestingly, it has been documented that host preferences vary considerably in different water bodies (Zhokhov and Pugacheva, 2012; Loot et al., 2002; Nazarizadeh et al., 2022). For instance, a study in south-western France observed that L. intestinalis predominantly infects roach populations, even in the presence of other potential hosts (Loot et al., 2002). This variation in host selection may indicate the development of host-specific races, suggesting adaptive differentiation within Lineage A. Given its ecological diversity and the observed patterns of host specificity, Lineage A provides a valuable model for investigating the mechanisms of divergent selection pressures and ecological isolation, contributing to our understanding of parasite evolution and speciation.

Here, we aim to elucidate the mechanisms underlying possible ecological speciation in *L. intestinalis*, particularly focusing on Lineage A populations in areas where all hosts coexist in sympatry. We intend to examine the potential divergent selection pressures and ecological isolation mechanisms that might influence parasite adaptations and diversity in sympatric settings. Utilising a comprehensive approach that involves the analysis of genome-wide SNPs and transcriptome data, we strive to: (a) ascertain whether host specificity in Lineage A can potentially influence the population structure of parasites in the absence of geographical separation, and (b) identify potential genomic signatures indicative of host specialisation within identified *Ligula* populations through selection analyses. Moreover, we aspire to expand our understanding of the complex dynamics of parasite evolution, especially in the context of ecological speciation under varied host pressures. Finally, we aim (c) to study RNA transcription patterns to identify differentially expressed genes (DEGs) related to host specialisation in parasite populations by establishing a comprehensive reference transcriptome sequence.

2 Material and Methods

2.1 Sample collection

We collected 83 plerocercoid samples from five prevalent host species in Czechia: *R. rutilus*, *A. alburnus*, *B. bjoerkna*, *S. erythrophthalmus*, and *A. brama* (Nazarizadeh et al., 2022). These specimens were sourced from 10 different freshwater ecosystems throughout Czechia (see Table S1 and Figure 1) using gillnets, in the frame of hydrobiological research (see Nazarizadeh et al., 2022 for sampling details). The samples were fixed in 96% ethanol and subsequently stored at a refrigerated temperature prior to DNA extraction. For the purpose of transcriptome

sequencing, we specifically selected one-year-old fish harbouring a single tapeworm to standardize conditions and minimize transcriptome expression variability. This led to 14 samples from three key host species (*R. rutilus*, *B. bjoerkna*, and *A. brama*), ensuring a reduction in sampling bias and an accurate reflection of host-parasite interactions across broader ecological settings. The central segment of each tapeworm was sectioned into several 5x5 mm squares, which were then submerged in RNAlater (Ambion, Austin, Texas, USA) and maintained at 4 °C overnight to stabilise and protect the RNA, and then stored at −80 °C until RNA extraction.

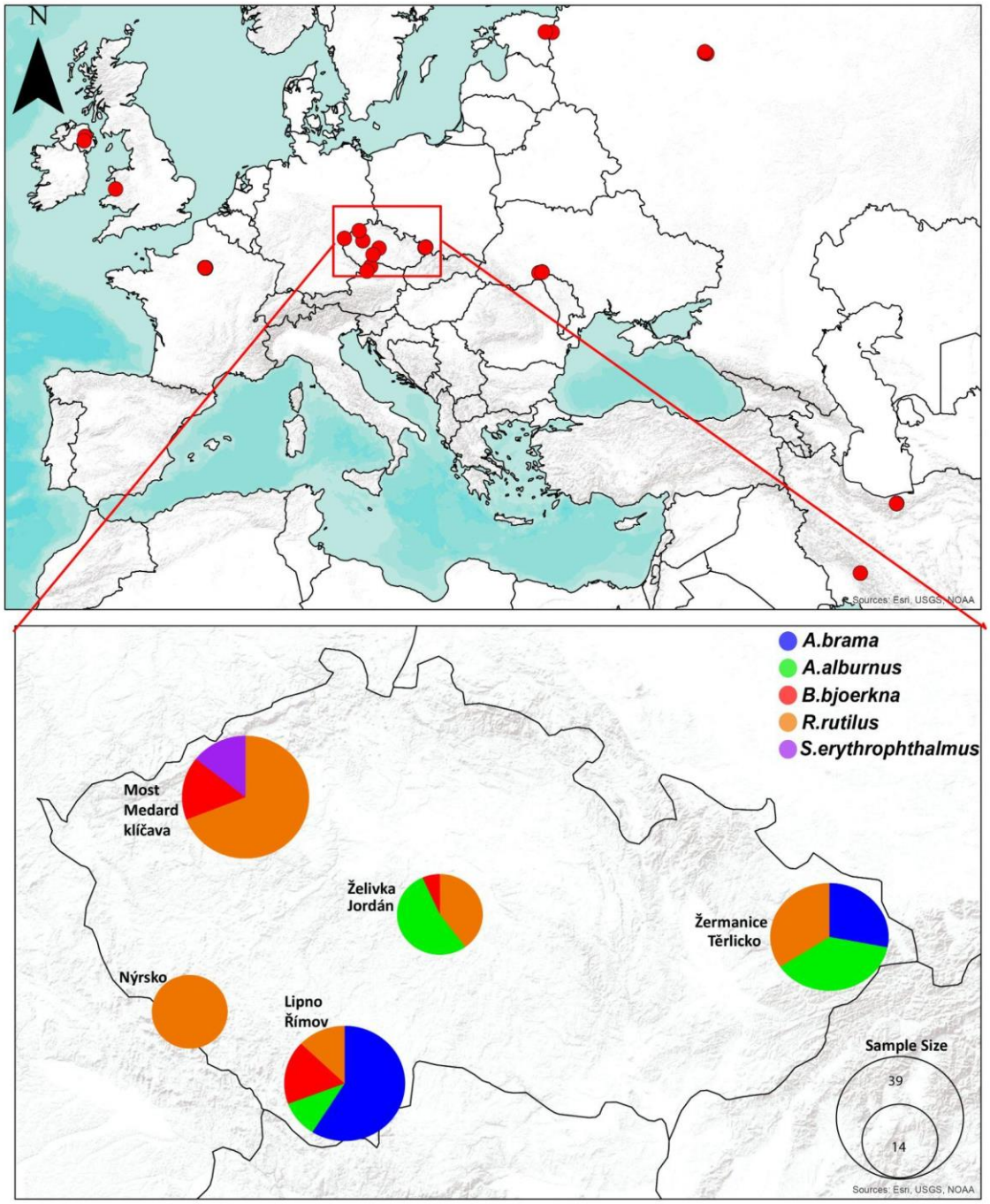


Figure 1. The distribution of Lineage A in *L. intestinalis* across the Palearctic (from Nazarizadeh et al., 2023, also analysed here). Inset: Sampling localities in Czechia, colour-coded to reflect the five most prevalent host species. Sampling sites in proximity to each were amalgamated into a single circle. The size of each circle correlates with the sample size.

2.2 DNA extraction and ddRAD library preparation

Total genomic DNA was extracted from specimens preserved in 96% ethanol using the DNeasy Blood and Tissue Kit (Qiagen). The quality and quantity of the DNA samples were verified

using a 0.8% agarose gel and a Qubit 2.0 Fluorometer, respectively. Double digest restriction-site associated DNA (ddRAD) libraries were generated in accordance with the modified ddRAD protocol established by Peterson et al., (2012). We utilised the same restriction enzymes (nspI + mluCI) had been previously employed to generate ddRAD data for *L. intestinalis* (Nazarizadeh et al., 2023). All 83 samples were pooled into two libraries, which were then sequenced in two lanes to maintain uniform coverage. This process generated paired-end reads of 150 bp on an Illumina NovaSeq (Novogene UK), yielding approximately 6.7 million paired-end reads per sample. The procedures for assembling multiplexed ddRAD-seq libraries for each barcoded individual sample (Table S1), as well as the detailed purification processes, were adopted from Nazarizadeh et al. (2023).

2.3 Transcriptome sequencing

To compare the transcriptome profiles among different parasite populations, we extracted RNA from all 14 plerocercoid samples across three species of fish hosts (*R. rutilus*, *A. brama*, and *B. bjoerkna*) (Table S1). RNA isolation followed the method described by Chomczynski and Sacchi (1987), utilising the acid guanidinium thiocyanate-phenol-chloroform procedure with reagents from Invitrogen (Carlsbad, CA, USA). The RNA yield was estimated using the Qubit RNA Broad Range Assay Kit (Thermo Fisher), and its integrity was assessed with an Agilent Bioanalyzer 2100 (Agilent Technologies, USA). Subsequently, the RNA samples underwent commercial processing to be converted into cDNA libraries, followed by sequencing using the Illumina Novaseq 150 PE read technology facilitated by Novogene (UK).

2.4 ddRAD data assembly and SNP calling

We applied the process_radtags program from Stacks v.2.5.6 to demultiplex and filter out raw reads with low-quality or unidentified bases from 83 new samples (Rochette et al., 2019). To enhance our dataset, we integrated 61 ddRAD datasets from Lineage A, derived from an earlier population genomic study on *L. intestinalis* (Nazarizadeh et al., 2023; see Table S1). Subsequently, all paired-end reads were aligned to the *Ligula* reference genome (Nazarizadeh et al., 2024; BioProject PRJNA1055111) using Bowtie v2.5.0's default settings (Langmead and Salzberg, 2012). We then sorted and converted all alignments to BAM format using Samtools v1.18 (Danecek et al., 2021) and conducted reference assembly for genotype calling and locus assembly using the ref_map.pl wrapper in Stacks. Different filtering options were applied to the genotype calling process (detailed in the supporting information) resulting in the generation

of four distinct SNP matrices in VCF format: **Dataset A**, which incorporates all samples from Lineage A (145 samples), featuring 334,469 SNPs with an average locus coverage of 21x; **Dataset B**, containing 39,127 SNPs/loci after LD filtering across all Lineage A individuals, with an average locus coverage of 18x; **Dataset C**, developed exclusively for samples from Czechia (118 individuals), including 333,497 SNPs with a 21x average locus depth of coverage; and **Dataset D**, comprising only unlinked SNPs from the Czech samples (72,696 SNPs/loci) with an average locus coverage of 17x.

2.5 Genome-wide diversity

We used the R package snpR (Hemstrom and Jones, 2023) to compute several metrics of genetic diversity - including nucleotide diversity (Pi), standardised individual heterozygosity (Hs), expected heterozygosity (He), observed heterozygosity (Ho), the inbreeding coefficient (Fis), and the proportion of polymorphic loci (P) - across different parasite populations, utilising Dataset A which contains 331,589 SNPs.

2.6 Population Genetic Structure

To understand the impact of host specificity on the population genetic structure of parasites, we employed three genetic clustering methods on two datasets: Lineage A samples (Dataset B) and Czechia-specific samples (Dataset D). Firstly, Discriminant Analysis of Principal Components (DAPC) was used to identify and visualize genetic clusters through Principal Component Analysis (PCA). Secondly, evolutionary clustering with ADMIXTURE determined the optimal number of clusters based on cross-validation errors, with results visualized using CLUMPAK. Lastly, we used fineRADstructure to examine haplotype-based population structures. Detailed methodologies for these analyses are provided in the supporting information.

2.7 Coalescent analysis of Speciation modelling

To better understand the impact of isolation and migrations on the population dynamics of Lineage A, we utilised a coalescent-based approach examining gene flow and demographics. Four speciation models were proposed, and their corresponding demographic models were applied to two parasite populations, as defined by the analyses of population genetic structure. The first speciation scenario was defined as allopatric speciation, characterised by complete geographic isolation without gene flow. The second speciation model accounted for post-

divergence gene flow through isolation after migration (primary contact), while the third speciation model tested post-divergence with secondary contact, which led to recent gene flow between the two parasite populations. Finally, the fourth scenario hypothesised isolation with continuous gene flow between parasite populations (Figure 2). See the supporting information for a detailed explanation of the methodologies employed in this analysis.

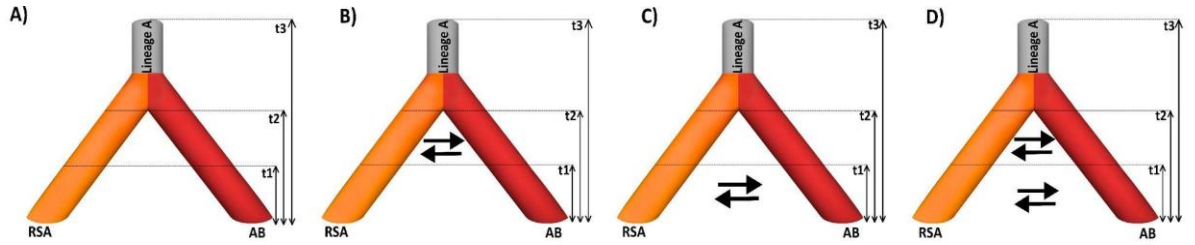


Figure 2. The four speciation scenarios testing the gene flow and demography patterns in Lineage A. A) Represents the allopatric speciation model with complete geographic isolation indicated by no gene flow between parasite populations from *R. rutilus*, *S. erythrophthalmus*, and *A. alburnus* (RSA) and parasite populations from *A.brama* and *B. bjoerkna (*AB). B) Depicts the model with post-divergence gene flow through primary contact. C) Shows post-divergence speciation with secondary contact leading to recent gene flow. D) Illustrates the scenario with continuous gene flow, as indicated by multiple gene flow arrows between populations over time.

2.8 Isolation by Distance

To investigate a potential correlation between genetic distance and geographic distance among parasite populations, we considered Dataset B to perform an isolation by distance (IBD) analysis using the Mantel test in the adegenet package in R (Jombart, 2008; Mentel, 1967). This analysis required calculating the Nei genetic distances and the geographic distances (measured in kilometres between population locations), which were then correlated through the mantel.randtest function. To reduce the impact of host specificity on genetic distance, we extended the analysis to various genetic structures, reinforced by a Monte Carlo simulation with 999 permutations for stronger statistical inference. Following this, we sought to clarify the nature of the observed correlations, determining if they illustrated a continuous or a patchy distant cline of genetic differentiation. To do this, we employed a 2-dimensional Kernel density estimator through the kde2d function found in the MASS package (Ripley et al., 2013) in R v.3.6.2, enhancing our grasp of the spatial distribution of genetic variation across the area.

2.9 Gene flow among parasite populations

To investigate gene flow among sympatric host parasite populations, we analyzed SNP data from Czechia (Dataset D) using Treemix v.1.12 to create a Maximum Likelihood (ML) tree based on allelic frequencies, thereby identifying historical migration events. Furthermore, we assessed contemporary gene flow using Bayesian inference with BA3SNP software, conducting a 10 million iteration analysis and discarding the initial 1 million as burn-in. Parameter adjustments were made for optimal acceptance rates, and cross-validation was employed to ensure model accuracy. Detailed methodologies are provided in the supporting information.

2.10 Genomic signatures of host specific selection

To test for evidence of natural selection associated with host specificity, we analysed Dataset C to identify loci under selection between two distinct parasite populations in sympatric hosts. These populations exhibit marked host specificity, as evidenced by their genetic structure. To accurately identify loci under directional selection, we used multiple selection analyses and adopted a strategy that relies on the convergence of results from several analytical methods (Tsumura et al., 2012). We employed four well-established tests to detect outlier loci: the PCA-based method outlined in pcadapt v4.3.3 (Privé et al., 2020), the FST frequency method detailed in Outflank v2 (Whitlock and Lotterhos, 2015), the Bayesian method that focuses on allele frequency variations as detailed in BayeScan v2.1 (Foll and Gaggiotti, 2008), and the variance analysis in haplotype frequencies conducted using the hapFLK v1.4 (Detailed methodologies are in the supporting information).

Finally, we estimated FST, Tajima's D, per locus absolute divergence (dXY), and nucleotide diversity (pi) using the R package PopGenome (Pfeifer et al., 2014). We then plotted both FST and dXY values against pi. We expect dXY to be significantly higher for outlier loci compared to neutral loci in a scenario of divergence with gene flow, assuming a sufficient amount of time has elapsed since the initial divergence (Cruickshank and Hahn, 2014). Moreover, dXY is expected to elevate in highly differentiated regions, resulting in a positive correlation between FST and dXY. Furthermore, we employed the “KStest” function from the GSAR R package (Rahmatallah et al., 2017) to perform the Kolmogorov–Smirnov D test, comparing Tajima’s D values between outlier and neutral loci. This helped ascertain if the distributions of Tajima's D values from putative outlier loci were statistically different from those of putative non-outlier

loci. A significant difference ($p < 0.05$) in distributions would imply that outlier loci have been subjected to distinct evolutionary pressures (i.e., selection, genetic drift, gene flow) compared to the rest of the genome.

2.11 Differentially Expressed Genes (DEGs) associated with host specificity

14 RNA-seq samples were used to compare transcriptome differences between two parasite populations detected by population genetic structure. Trimmomatic v0.33 (Bolger et al., 2014) was used to discard low quality reads from all samples. Following trimming, paired-end reads (Table S2) were mapped to the *L. intestinalis* reference genome using STAR (Dobin et al., 2013). To eliminate reads that mapped to multiple loci and also to sort and convert the filtered SAM files to BAM format, we used Samtools v1.18 (Danecek et al., 2021). Furthermore, the mapped reads counted for genomic features using featureCounts v1.06 in the Subread package (Liao et al., 2014). Table S3 presents mapped reads per sample replicate and gene model. To normalise the raw counts, we applied the Trimmed Mean of M values (TMM) method. The edgeR (Robinson et al., 2010) was used to estimate differential expression of genes between parasite populations based on their host specificity. Additionally, we applied the criteria of an absolute log-fold change greater than 1 or less than -1 and a p-value less than 0.05 to identify DEGs. Finally, using the plotMDS function in edgeR, we plotted the first two principal components to evaluate the general similarities and differences among all the transcriptomes.

2.12 Functional annotation of loci under selection and DEGs

We implemented a two-step strategy to annotate the selected loci. In the first step, we identified outliers in the *Ligula* reference genome and retrieved the relevant annotation details using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011). Next, for SNPs located in non-gene regions of the *Ligula* genome, we took into account a 201 bp sequence - comprising 100 bp both upstream and downstream of the SNP from the reference genome sequences. Using OmicsBox v3 (BioBam, 2019), we performed a comprehensive functional annotation of these sequences. Initially, nucleotide sequences (201 bp) were blasted to the NCBI non-redundant nucleotide (nt) database Pruitt et al., 2007) through BLASTX. For each search, we annotated the primary hit with the highest total score, maintaining an E-score of 10-5 for BLASTX and 10-15 for BLASTN, and insisting on a search coverage of more than 70%. We further enriched our annotations using the integrated InterProScan (Jones et al., 2014) module for domain-based

information and mapped our data to the EggNOG database to provide broader functional insights (Huerta-Cepas et al., 2019).

Gene Ontology (GO) annotations for *L. intestinalis* genes were sourced from Nazarizadeh et al., 2024. We conducted functional enrichment analysis using the GSEA function in the clusterprofiler R package (Wu et al., 2021), applying the Benjamini-Hochberg false discovery rate adjustment with a 0.05 threshold. We used the DOSE R package (Yu et al., 2014) to create dot plots and enrichment maps for the highlighted genes. GOs were summarized via REVIGO (revigo.irb.hr) for DEG sets, adopting a 0.4 threshold and using SimRel for similarity (Supek et al., 2011).

3 Results

3.1 Genomic diversity

We compared the genetic characteristics of various parasite populations from different fish host species within *L. intestinalis* Lineage A (Table 1). Populations found in *R. rutilus* located in Czechia, France, Ireland, and Ukraine exhibited the highest nucleotide diversity (Pi = 0.013) and expected heterozygosity (He = 0.131). Conversely, the lowest nucleotide diversity was observed in parasites from *A. brama* in Czechia, Estonia, and Russia, as well as *A. alburnus* hosts in Ukraine and Czechia (Pi = 0.012). Notably, parasites in *S. orientalis* in Iran showed the highest observed heterozygosity (Ho = 0.172), which was significantly higher than their expected heterozygosity. In contrast, the lowest Ho was observed in *B. bjoerkna* in Czechia (Ho = 0.089). Moreover, the highest proportion of polymorphic loci (P) was found in parasite populations of *R. rutilus* hosts (P = 0.93), whereas *S. orientalis* exhibited the lowest proportion (P = 0.35) (Table 1).

Table 1. Genetic diversity metrics of parasite populations in sampled hosts and localities. The table shows nucleotide diversity (Pi), standardized individual heterozygosity (Hs), expected heterozygosity (He), observed heterozygosity (Ho), Inbreeding coefficient (Fis), and the proportion of polymorphic loci (P).

Parasite population	Hosts	Locality	N	Pi	Hs	He	Ho	Fis	P
Lineage A	*R. rutilus*	Czechia,France,Ireland, Ukraine	73	0.013	1.02	0.131	0.118	0.106	0.93
	A. brama	Czechia, Estonia, Russia,	28	0.012	0.944	0.120	0.115	0.116	0.65
	A alburnus	Ukraine, Czechia	21	0.012	1.00	0.126	0.116	0.107	0.65
	B. bjoerkna	Czechia	10	0.011	0.997	0.119	0.108	0.089	0.48
	S. erythrophthalmus	Czechia	5	0.012	1.02	0.117	0.118	0.102	0.40
	S. orientalis	Iran	4	0.012	0.89	0.110	0.103	0.172	0.35
	C. carassius	Ukraine	1	-		-	-	-	-
	P. Phoxinus	United Kingdom	1	-		-	-	-	-

3.2. Population genetic structure

The DAPC analysis for Dataset B (spanning all geographic locations) and Dataset D (specific to Czechia) consistently identified two primary genetic clusters within Lineage A. This finding was reinforced by the K-means method, with the best partitioning at a K value of 2, as indicated by the lowest BIC score for these datasets (Figure S2A and B). In both datasets, the primary DAPC axis, DAPC 1, accounts for 39.3% and 45.1% of the genetic variance, respectively. This axis divides the populations into two distinct clusters: the first includes parasite populations from *A. brama* and *B. bjoerkna*, while the second comprises populations from a variety of other species, including *S. erythrophthalmus*, *R. rutilus*, *S. orientalis*, *P. phoxinus*, and *C. Carassius*. The secondary axis, contributing to a smaller 8.9% of the total variance, distinguishes the parasite populations in *S. orientalis* and *A. alburnus* from those in *R. rutilus* and *S. erythrophthalmus* (Figure 3A and B). The results of admixture analysis from Datasets B and D align well with the DAPC analysis, identifying an optimal K value of 2, as indicated by the cross-validation error (CV, Figure S2C and D). On a large scale, parasite populations in Lineage

A are divided into two genetic clusters, mirroring the local scale where all host species coexist (Figure 3A and 3B). The first cluster contains parasite populations from *A.brama* and *B. bjoerkna*, while the second groups together populations from *R. rutilus*, *A. alburnus*, and *S. erythrophthalmus*. Furthermore, the analysis showed that parasite populations from *B. bjoerkna* are significantly admixed, with a 46%–90% assignment to the *R. rutilus* parasite population (Figure 3C).

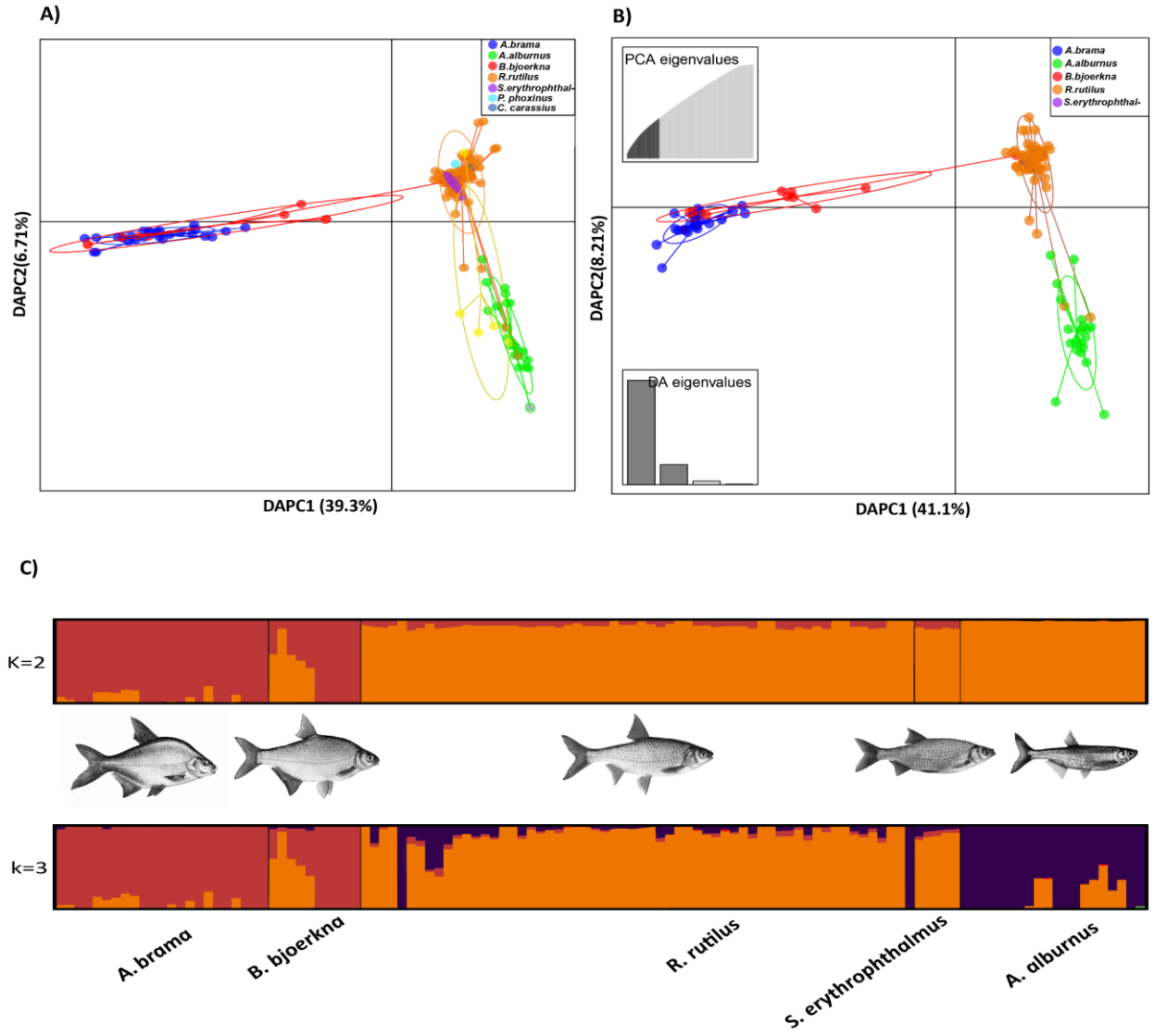


Figure 3. Genetic structure of parasite populations within Lineage A across different host species. A) DAPC analysis of datasets B (global, 148 samples) and D (Czechia-specific, 118 samples) reveals two primary genetic clusters, with primary axis (DAPC 1) accounting for 39.3% and 45.1% of variance, respectively. This axis predominantly segregates populations from *A. brama* and *B. bjoerkna* from other species such as *S. erythrophthalmus*, *R. rutilus*, *S. orientalis*, *P. phoxinus*, and *C. carassius*. The secondary axis highlights

differentiation between *S. orientalis* and *A. alburnus* versus *R. rutilus* and *S. erythrophthalmus*, contributing to 8.9% of total variance. B) Admixture analysis is consistent with DAPC findings, particularly noting the significant admixture in *B. bjoerkna* populations, shown with a 46%–90% genetic overlap with the *R. rutilus* parasite population.

Additionally, the fineRADstructure analysis validated the clustering results with K=2, demonstrating significant support in the dendrogram. This analysis identified the parasite population in *A. alburnus* as a sub-cluster stemming from the populations in *R. rutilus* and *S. erythrophthalmus*. It also highlighted a notable level of shared co-ancestry between the parasite populations in *A. brama* and *B. bjoerkna*. However, this genetic cluster exhibits a lower level of shared co-ancestry with other parasite populations, both on a broad and local scale (Figure S3 and Figure 4). In line with the admixture results, one sample from *B. bjoerkna* was clustered with the parasite populations in *R. rutilus*. The Mantel test results further confirmed the absence of a significant correlation between genetic distance and geographic location within Lineage A parasite populations. This suggests that geographical distances do not hinder gene flow at the intra-population level (Figure S4).

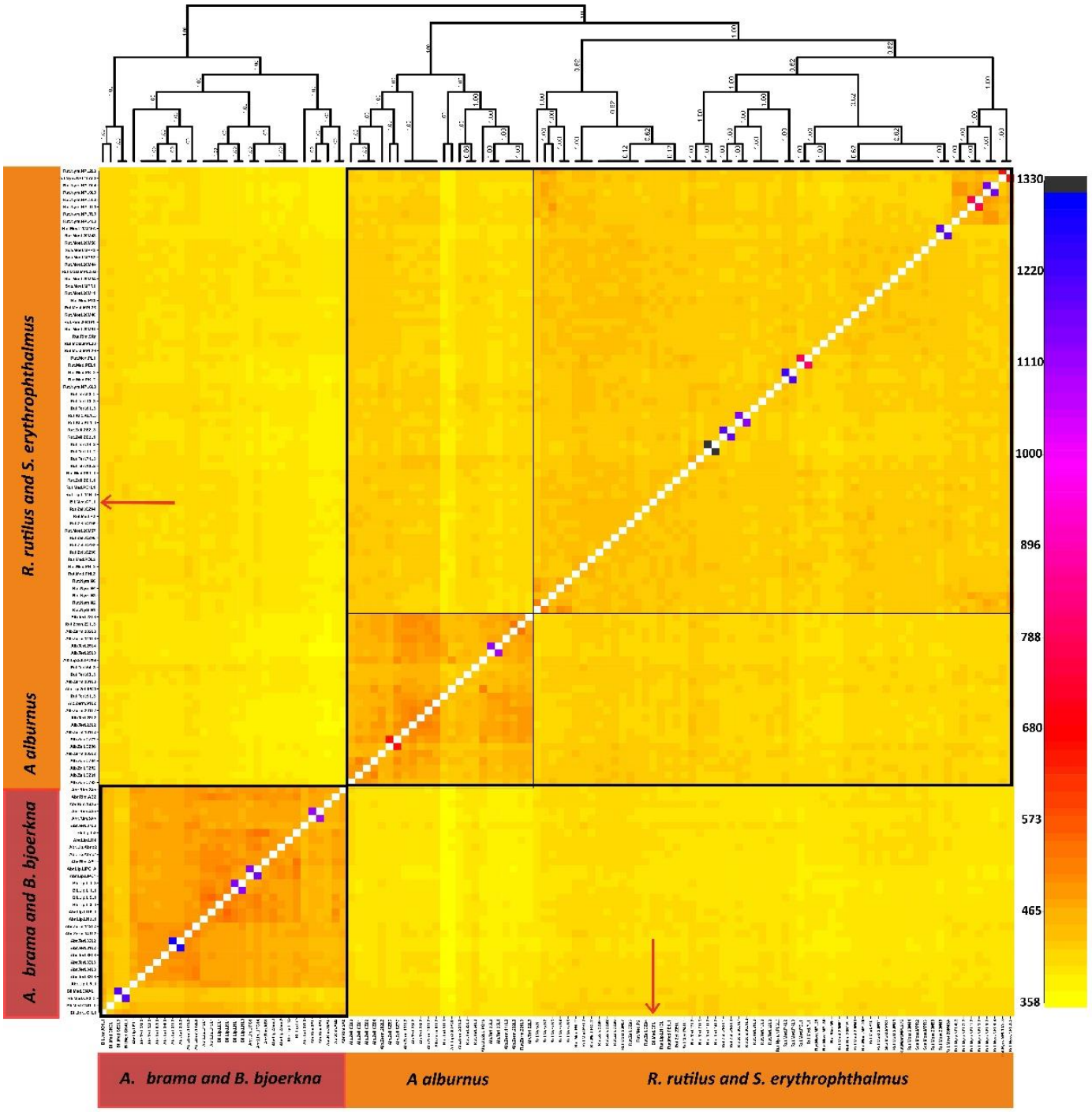


Figure 4. Co-ancestry plot and dendrogram derived from the fineRADstructure evaluation of *Ligula* populations in Czechia. Co-ancestry coefficients transition from low (yellow) to high (blue), indicating the extent of recent shared ancestry between a focal individual (on the vertical axis) and all other individuals included in the study (on the horizontal axis). The red arrow highlights a single sample of *Ligula* from the host *B. bjoerkna* (a host switch), which is nested in a genetic cluster from *R. rutilus* and *S. erythrophthalmus*.

3.2 The divergence modelling of parasite populations

Coalescent analysis for species modelling showed the most probable scenario to be Isolation with ongoing gene flow, where gene flow rates have recently risen. The change in the rate of gene flow was inferred to have occurred about 58000 generations ago. However, these time estimates should be interpreted with caution as we fixed the first split to the lower bound of 160,000 generations ago (corresponding roughly to the time since when two parasite populations probably diverged from each other, Nazarizadeh et al 2023). Splitting times may be even more recent if there was a significant time lag between gene tree and species tree. The second most likely model, with a ΔAIC of 28.1, suggested recent gene flow starting 9,100 generations ago, representing secondary contact. Models involving primary contact showed very close likelihoods (ΔAIC 25.4), while the model excluding gene flow was a poor match for the data (ΔAIC 96.4). These findings indicate speciation occurred amidst gene flow. However, having only two migration matrices is a strong simplification of the speciation process (Table S3).

3.3 Gene flow among parasite populations in different hosts

The unrooted species tree derived from the Treemix analysis identified two primary genetic groups, corroborating the findings from both the admixture and fineRADstructure analyses. Notably, the data are best explained by a model that considers two migration events. Parasite populations in *A. brama* and *B. bjoerkna* demonstrated a close genetic relationship, forming a sister lineage distinct from other parasite groups. The most prominent migration events were observed among the parasite communities in *A. alburnus*, *R. rutilus*, and *S. erythrophthalmus*. Additionally, a gene flow with a lower weight was detected between *B. bjoerkna* and *S. erythrophthalmus* populations (Figure 5A). The results of the contemporary gene flow indicated that most of the gene flow originated from populations within the same host species. While there was some gene flow between different hosts, populations in *A. brama* and *B. bjoerkna* symmetrically contributed to the pool of exogenous allelic variants. Similarly, a noticeable amount of balanced gene flow was observed from the parasite populations in *R. rutilus* to those in *A. alburnus* and *S. erythrophthalmus*. Moreover, gene flow was detected between populations in *A. brama* and *B. bjoerkna* and those in *A. alburnus*, *R. rutilus*, and *S. erythrophthalmus* (Figure 5B).

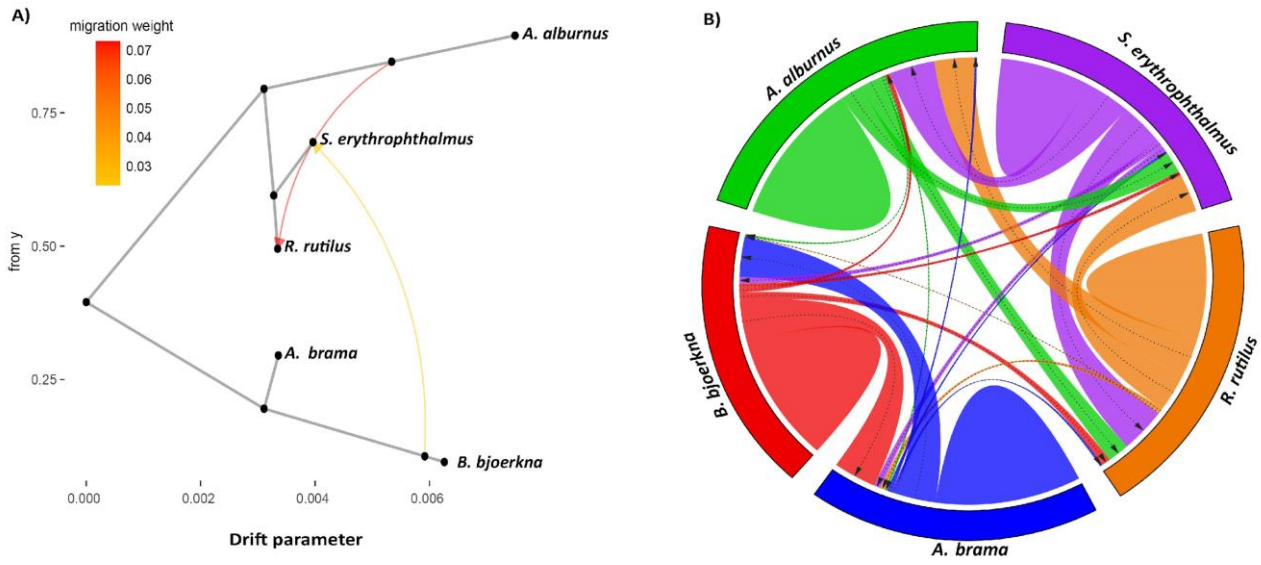


Figure 5. Gene flow analysis of *Ligula* populations in Czechia. (A) Species tree derived from Treemix shows two principal genetic clusters with significant gene flow among *A. alburnus*, *R. rutilus*, and *S. erythrophthalmus*. (B) analysis of BayesAss reveals contemporary gene flow among parasite populations in different hosts, particularly originating from *R. rutilus*, and underscores the contributions of *A. brama* and *B. bjoerkna* to various host populations.

3.4. Detection of SNPs under divergent selection

Selection analyses were conducted to identify loci potentially under divergent selection in two parasite populations structured within sympatric hosts. A total of 3,004 outlier loci, putatively under selection, were identified using PCAdapt, OutFLANK, BAYESCAN, and HapFLK. The least conservative method, BAYESCAN, identified the highest number of outliers with 2,367 SNPs, PCAdapt detected 1,743 SNPs, whilst OutFLANK and HapFLK were the most conservative, pinpointing 1,384 and 1,256 loci, respectively. Notably, all four methods collectively detected 896 SNPs (29.8% of the total SNPs), marking them as the putatively adaptive dataset (Figure S4). Most of the commonly detected loci across the methods were found in the 10 longest scaffolds of the *Ligula* genome (Figure 6). Out of 896 outlier loci identified through four methods, 156 were directly associated with the coding sections of 65 gene models in the *Ligula* genome. Considering potential gaps in the genome annotation, a second method was applied, extending the search to 100 bp before and after each outlier. This approach revealed 37 additional outliers, sharing homology with genes from closely related tapeworm species.

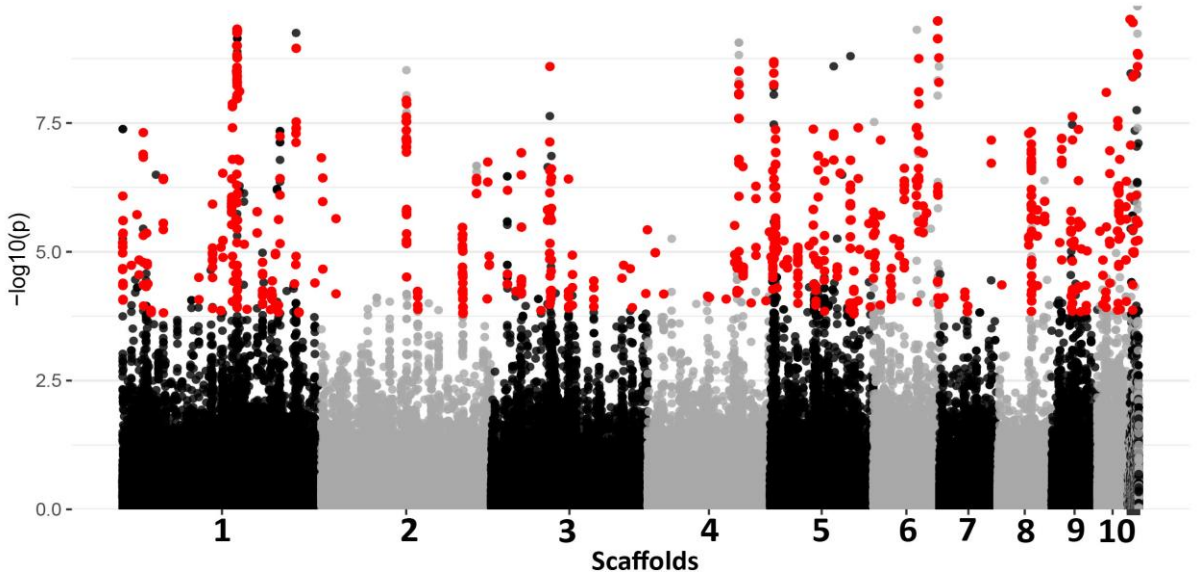


Figure 6. A consensus Manhattan plot displaying results from four genome-wide selection analyses: PCAdapt, Outflank, BAYESCAN, and HapFLK. This plot presents the 10 longest scaffolds of the *Ligula* genome (additional scaffolds are also included in the plot and are represented collectively due to their shorter lengths; Nazarizadeh et al., 2024). SNPs marked in red indicate those under selection between two parasite populations found in sympatric hosts. The y-axis represents -log10 (p) values.

A significant positive correlation was observed between the values of dXY and nucleotide diversity for both neutral and outlier loci. Outlier loci predominantly displayed a higher dXY compared to neutral loci and exhibited lower to moderate nucleotide diversity in contrast to neutral loci (Figure 7A). Similarly, dXY showed a positive correlation with FST, and outlier loci displayed a relatively higher FST compared to neutral loci (Figure 7B). Notably, the distributions of Tajima's D revealed significant negative values for putative outlier in comparison to non-outlier loci (P value < 0.01, Figure S5).

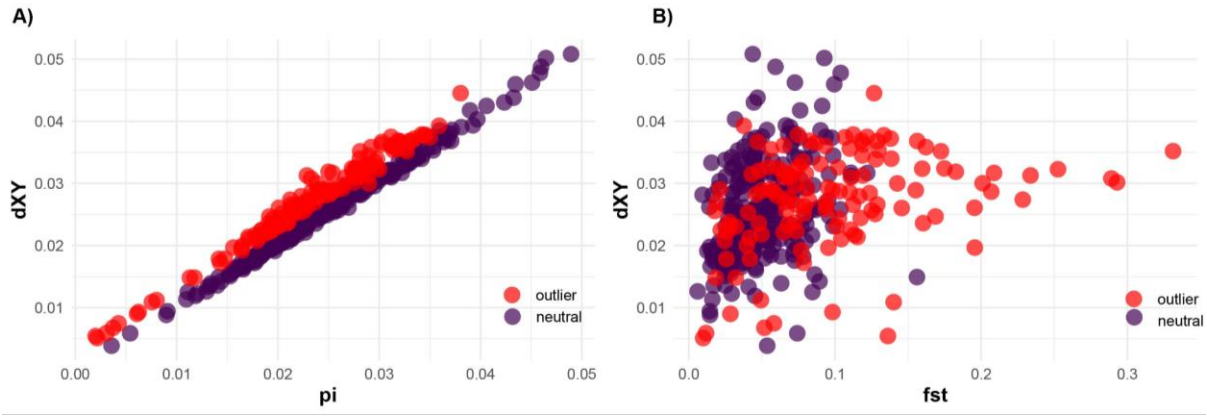


Figure 7. Comparative analysis of genome-wide variations between putative outlier and non-outlier loci. A) Correlation between dXY and pi distinguishing outlier from neutral loci. B) Patterns of dXY distribution in association with FST, highlighted for both outlier and neutral loci.

3.5 Differential gene expression

Gene expression analysis of 14 transcriptome samples derived from three different hosts in a sympatric setting, identified 993 DEGs: 556 were up-regulated and 367 down-regulated in the parasite populations of *R. rutilus* compared to *A. brama* and *B. bjoerkna* Figure (8A and 8B). This pattern is consistent with the results of the genetic structure of the parasite populations. The RNA expression profiles showed that parasites in two of the hosts, *A. brama* and *B. bjoerkna*, had similar transcriptome patterns. In contrast, parasites in *R. rutilus* showed a distinct transcriptomic profile (Figure 8A). Furthermore, the parasite population in *R. rutilus* differed from those in *A. brama* and *B. bjoerkna* in PCA, with the first dimension accounting for 32.11% of the observed variance (Figure 8C).

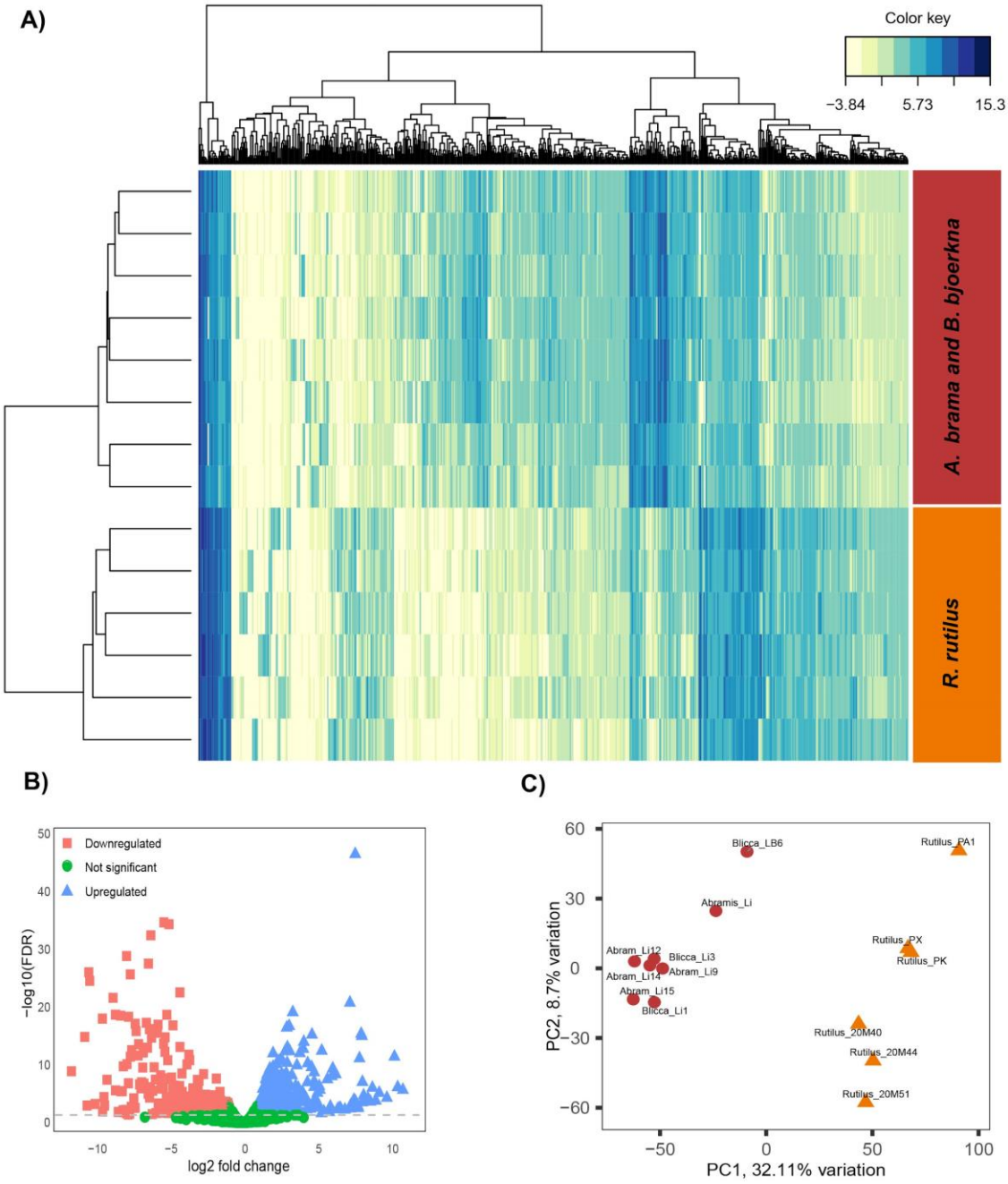
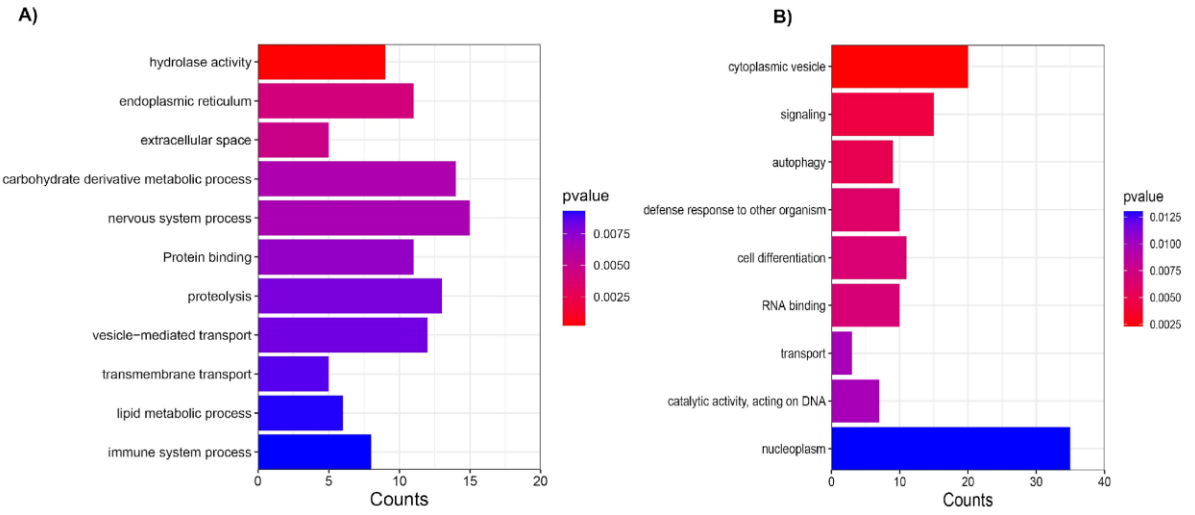


Figure 7. Differential gene expression in *L. intestinalis* samples obtained from three host species living in sympatry. (A) Heatmap illustrates the differential expression of genes across parasite populations in different host species. (B) Volcano plot displaying upregulated (red rectangles), downregulated (blue triangles), and non-differentially expressed transcripts (green dots) across hosts *(R. rutilus*, *A. brama,* and B. *bjoerkna*). (C) PCA showcasing the clustering of sample replicates, with relative variances detailed across PC1 and PC2.

3.6 Gene Ontology Patterns in DEGs and SNPs under selection

The results of the Gene Ontology analysis revealed significant variations in gene enrichment in parasite populations across different host environments. In the parasite population from *R rutilus*, the Biological Process (BP) ontology was highlighted by dominant terms such as "nervous system process" with 15 genes and "proteolysis" with 13 genes, indicating a possible focus on neural activities and protein degradation. Regarding Cellular Components (CC), genes associated with "extracellular space" and "endoplasmic reticulum" were particularly prominent, suggesting interactions between cellular organelles and the external environment. The Molecular Function ontology (MF) highlights the significant presence of "hydrolase activity" in this environment, indicating central enzymatic activities (Figure 9A). In contrast, in *L. intestinalis* populations from *A. brama* and *B. bjoerkna*, the BP ontology showed "signalling" with 15 genes, indicating important communication and regulatory pathways. The CC ontology was dominated by "nucleoplasm" with 35 genes, indicating potentially enhanced nuclear activities or regulations. "Cytoplasmic vesicles" also played an important role, indicating important intracellular transport mechanisms. Within the MF ontology, there was a notable focus on "RNA binding" and "Catalytic activity acting on DNA", indicating essential molecular interactions and potential DNA manipulations (Figure 9B).



Figre 9. Differential Gene Ontology (GO) term enrichment in *L. intestinalis* populations parasitising different hosts. (A) Overrepresented GO terms in the parasite population from *R. rutilus*, highlighting key biological processes, cellular components, and molecular functions with the respective gene counts. (B) Overrepresented GO terms in parasite populations from *A. brama* and *B. bjoerkna* hosts, showcasing the distinct adaptive molecular responses and interactions within these environments. The y-axis represents the number of genes associated with each GO term.

The ddRAD-derived putative outlier data revealed a comprehensive profile of GO term enrichment within *L. intestinalis* populations. Most notably, the BP ontology had 140 terms that emphasised processes such as the cell junction organisation, autophagy, DNA integration, response to abiotic stimulus, generation of precursor metabolites and energy, Regulation of signalling, immune system processes and ammonium ion metabolic process (Figure S6). The enrichment for cellular components included 74 terms that emphasised cellular structures such as the cytosol, lipid droplet, sarcolemma and plasma membrane (Figure S7). The ontology for molecular functions, on the other hand, was the most diverse with 52 terms, highlighting functionalities such as cAMP binding, ether hydrolase activity, protein binding and catalytic activity (Figure S8). A comparative analysis with previously evaluated transcriptome data revealed a common enrichment of GO. These common terms from proteolysis, nucleolysis, immune system processes, signalling, autophagy, cytosol, and hydrolase activity highlight the consistent molecular and cellular activities central to *L. intestinalis* populations across different analytical platforms and datasets. Overall, five genes were identified as common between the ddRAD selection analyses and differential gene expression analysis.

4 Discussion

The interplay between ecological pressures and natural selection in shaping host specificity is a critical factor in the evolutionary dynamics of parasites (Simmond et al., 2020; Nikolakis et al., 2022; Bellis et al., 2022). Ecological pressures, such as environmental conditions and the availability of hosts, can greatly influence the range of organisms a parasite can infect (Cháves-González et al 2022). This creates a selective environment where certain traits are favoured, leading to adaptations in host specificity (Clark and Clegg, 2017). In the present study, we have employed genome-wide SNPs and transcriptome data to unravel the complexities of the intraspecific population structure within a single evolutionary lineage of *L. intestinalis* tapeworm. This study places a special focus on the parasite's interactions with its second intermediate hosts, thus shedding light on how ecological factors play a significant role in driving speciation between closely related parasite populations. Our research was propelled by the intriguing observation of two distinct parasite populations coexisting in sympatry. Notably, these two clusters exhibited a high level of concordance with two different groups of host species, suggesting a subtle yet significant interplay between host specificity and the evolutionary trajectory of the parasite. Our study suggests that the genetic divergence between

the two parasite clusters is indicative of speciation with ongoing gene flow, where disruptive selection facilitates reproductive isolation. Direct association between ecological divergence and reproductive isolation is a keenly sought phenomenon in specialist organisms, including parasitic plants (Giraud et al., 2010) phytophagous ladybird beetles (Matsubayashi et al., 2017), parasitic wasps (Stelinski and Liburd, 2005), leaf mites (Skoracka et al., 2013) and phytophagous insects (Berlocher and Feder, 2002; Funk et al., 2002). However, only a few studies explored this link using genomic data and in a framework including ongoing gene flow (Hume et al., 2018; Mateus et al., 2013; Villacis-Perez et al., 2021). Below, we delve into the results of our investigation, specifically examining the role of ecological speciation in the evolution and diversification of this parasite.

4.1 Population genetic structure and divergence with ongoing gene flow

Analysis of genetic structure using genome-wide SNP data revealed that *L. intestinalis* Lineage A is divided into two primary clusters within its Palearctic distribution. At this intraspecies level, distinct differences were observed in the parasite populations of *A. brama* and *B. bjoerkna* compared to those in *R. rutilus*, *S. erythrophthalmus*, *A. alburnus*, *P. phoxinus*, and *C. carassius*. This pattern was also evident on a more local scale within Czechia, where these host species were sampled in sympatry, highlighting the role of host specificity in the parasite's evolution. These findings align with our previous study that revealed a non-random shared haplotype between five individuals of parasite populations in *A. brama* and *R. rutilus* (Nazarizadeh et al., 2022). Yet, earlier investigation involving more individuals showed a high number of shared mtDNA haplotypes and an indistinguishable population structure among all parasite populations in Lineage A (Bouzid et al., 2008). This result indicates that while mtDNA provides valuable insights into broad evolutionary patterns and phylogeography, it may be less effective in detecting subtle, microevolutionary changes and ecological divergence that occur over short timescales or in complex population structures. In contrast, genome-wide SNP and nuclear marker analyses offer a more comprehensive and detailed view of these processes.

Ecological divergence often interacts with gene flow, the extent of which varies with dispersal distance (Räsänen & Hendry, 2008). In our study, we observed genetic divergence among parasites from two related genetic clusters, yet we also identified notable gene flow between them. Our Treemix analysis indicated a significant migration event that suggests interbreeding between these groups. This was further supported by admixture analysis, which identified hybrid individuals among the two parasite clusters. Additionally, within one of these clusters,

encompassing the fish hosts *A. alburnus*, *R. rutilus*, and *S. erythrophthalmus*, we detected another significant gene flow event. Because these specialist parasite populations have demonstrated high divergence in host use, our results clearly demonstrate that the ecological divergence has been well maintained by natural selection even in the presence of gene flow. Our results are in line with the contemporary perspective of sympatric speciation in the genomic era, in which gene flow is almost pervasive between recently diverged sister lineages (Marques et al. 2019).

We focused on distinguishing between historical secondary admixture and ongoing gene flow in the *Ligula* system using two approaches, analysis of genetic diversity in outlier loci and demographic analysis of various admixture scenarios. Analysis of 896 outlier loci p revealed positive correlation between dXY and nucleotide diversity suggesting that as populations split, genetic variation within them increases, hinting at a continued gene flow (Cruickshank and Hahn, 2014). This is further corroborated by the finding that outlier loci, which are likely under strong selective pressure, exhibit a higher degree of divergence (as indicated by higher dXY values) compared to neutral loci. Additionally, these outlier loci displayed lower to moderate nucleotide diversity, suggesting that selection is acting on these loci to a greater extent, limiting their genetic variation (Nosil et al., 2009). The correlation of dXY with FST and the significant negative values of Tajima's D in outlier loci reinforce the notion that selection is driving specific parts of the genome to diverge, even as gene flow persists (Cruickshank and Hahn, 2014). This pattern aligns with the second approach, using coalescent species modelling, where gene flow does not preclude the emergence of distinct populations, as differential selection pressures can lead to significant divergence in key genomic regions, setting the stage for the evolution of new evolutionary species.

Recent studies in various host-parasite systems show complex patterns of gene flow and reproductive isolation in parasites (Leder et al., 2021; Momigliano et al., 2017; Simmonds et al., 2020; Teske et al., 2019). These studies highlight the potential influence of disruptive selection in the process of speciation, leading to the divergence of a single ancestral population into groups that specialise in different habitats and hosts. In the present study, the result of species modelling revealed that the most plausible evolutionary scenario for the divergence between parasite genetic clusters is speciation through isolation with migration. This involves recurrent migration after divergence, as illustrated in Figure 2a. This notion supports the increasing evidence that isolation with ongoing gene flow is a likely mechanism for speciation

during ecological divergence (Nosil, 2012; Rundle and Nosil, 2005). The divergence of these two parasite clusters likely occurred in the Chibanian period (Middle Pleistocene, according to Nazarizadeh et al., 2023), possibly in tandem with the colonisation of and host shifting among various cypriniform hosts. This period, characterised by significant climatic and environmental shifts, might have led to different subgroups within a population diverging due to varying selective pressures (Walker et al., 2019). Here, ecological disruptive selection is a key driver of ecological speciation, as it enhances the fitness of each population within its respective host groups. Notably, such divergence, driven by local adaptation, can occur even among populations in close geographical proximity, leading to sympatric speciation. In such scenarios, segments within a parasite population begin to specialise in exploiting different host species within the same ecosystem. For instance, one subgroup of a parasitic species might adapt to thrive with a specific host, exploiting unique physiological or immunological traits of that host, while another subgroup might specialise in a completely different host species.

4.2 Candidate loci and transcriptome genes involved in adaptation to host

The study of gene expression variation within and between species has long been a subject of interest, particularly in discerning the influences of gene plasticity and natural selection (Mathieu-Bégné et al., 2022; Romero et al., 2012). If significant genetic differentiation and selection-related outlier SNPs are detected, it implies that changes in gene expression might be attributed to evolutionary adaptations, rather than just to plastic responses to varying environments (Mathieu-Bégné et al., 2022). In this study, we explored the genomic differentiation between two parasite populations living in sympatric hosts over different time scales by comparing genomic outlier loci indicative of long-term adaptation with DEGs representing short-term acclimation. Although our transcriptome data did not cover all parasite populations from the five hosts studied, our results were consistent with genome-wide SNP analyses, revealing clear separate profiles of gene expression between parasite populations from *R. rutilus* and those in *A. brama* and *B. bjoerkna*. These differences in gene expression profiling between the two host groups are associated with various biological tasks. We will further explore possible explanations for the observed DEG patterns in relation to the parasite's adaptation to different hosts. However, the application of enrichment analysis to *L. intestinalis* can be challenging due to difficulties in assigning orthology, which is essential for transferring functional information. This challenge is primarily due to the large phylogenetic distance between *Ligula* and other model organisms within the Cestoda, coupled with the rapid

molecular evolution of the parasites (Morand, 2000). Consequently, gene ontology may yield uncertain functional inferences.

Our findings include five genes showing a direct correlation between outlier loci and DEGs, either in terms of gene identity or physical proximity, along with notable functional parallels. Of these five DEGs, ANN02916 and ANN012081, are associated with reverse transcription activities. Aligning with a prior research by Nazarizadeh et al. (2024), ANN02916 was identified as significantly downregulated during the transition from the larval to the adult stage in *L. intestinalis*. The reverse transcription process involves DNA sequences in an organism's genome derived from RNA, which is a key process in retroviruses and retrotransposons (Hughes, 2015). Additionally, our study sheds light on the biological processes associated with DNA integration through the annotation of outlier SNPs. These findings suggest that DNA integration and activities related to reverse transcription are pivotal in determining the host specificity and immune evasion strategies of the parasite. Such mechanisms are crucial not only across different host groups but also between various life stages, including larvae and adults (Nazarizadeh et al., 2024). These mechanisms contribute to genetic variability and local adaptation, which are essential for the survival and evolutionary success of the parasite within their host environments.

The tree other genes, ANN07147, ANN10664, and ANN13534, identified as significant in both SNPs under selection and DEGs play critical roles in biological functions linked to the endoplasmic reticulum (ER) membrane and transmembrane transport. Both processes may be important for the survival and host specificity of parasites, the membrane is key for carrier-mediated transport within the tapeworm tegument, greatly influencing the chemical modification of absorbed substances and offering protection against the host’s digestive enzymes. Notably, this membrane includes a glycocalyx layer made up of oligo- or polysaccharide chains attached to external lipid and protein elements. This layer is fundamental for binding various substances, such as inorganic ions and larger organic molecules including host enzymes. While some of these enzymes remain active on the worm’s surface, facilitating contact digestion, others are bound in a non-active form, possibly serving as a defence mechanism to prevent the parasite from being digested by the host. Additionally, our results highlight the biological function of glycosylation, potentially linked to the glycocalyx layer of the tegument. Studies have shown that glycosylation inhibitors like tunicamycin can substantially reduce the incorporation of galactose into the tapeworm's tegument and carcass

(Hildreth et al., 1997; Izvekova et al., 2021). This reduction highlights the importance of glycosylation in maintaining the surface glycocalyx of the tapeworm (Hildreth et al., 1997). Moreover, the presence of glucose transporter homologues in *Taenia solium* suggests a link between glucose absorption through the tegument and glycosylation processes (Rodríguez-Contreras et al., 1998). The tegument's external limiting membrane is coated with carbohydrate-rich polyelectrolytes, indicating the presence of glycosylation and its potential role in the tapeworm's membrane structure and function (Lumsden, 1975). This glycosylation is not only crucial for the tapeworm's nutrient absorption but also plays a significant role in host-parasite interactions. The functional importance of glycolipids in *Spirometra erinaceieuropaei* is a prime example (Yanagisawa et al., 1999). Additionally, the presence of glycoconjugates in the tegument of *Taenia taeniaeformis* may contribute to the parasite's resistance to host digestive processes (Olson et al., 2000).

In conclusion, using whole-genome genotyping and transcriptome analyses we demonstrated how host-specialisation leads to sympatric genetic differentiation in spite of continued gene flow. We proposed disruptive host-mediated selection on genes involved in immunological or physiological interaction with the host as the driver of ecological speciation in this pathogenic parasite of freshwater fish. This study provides an example of a frequently anticipated, but rarely observed, ecological phenomenon of speciation by host specificity.

Data availability

All ddRAD and transcriptome data used in this study are available in the National Center for Biotechnology Information (NCBI) database under the BioProject number PRJNA1088389. Additionally, all VCF files resulting from genotype calling are accessible on Dryad (link: XXXXXXXXX).

Acknowledgements
We would like to thank Anna Mácová and Roman Hrdlička for their valuable assistance in the field. This work was supported by a grant from the Czech Science Foundation (GA19-04676S). Computational resources were provided by the project 'e-Infrastruktura CZ' (e-INFRA CZ LM2018140), which is supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Author contributions

JŠ initiated the research topic. MNaz participated in field collection and analysed the data under the supervision of JŠ and JV. MNov conducted laboratory analyses. MNaz and JŠ drafted the paper. All authors reviewed and approved the final manuscript.

Supplemental information

ddRAD data assembly and SNP calling

Using the process_radtags program from Stacks v.2.5.6, we removed cut sites, barcodes, and adaptors from the ddRAD results, an essential step to ensure data integrity. Subsequently, we assessed the initial quality of our raw sequencing data using FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016), which provide comprehensive quality control reports. For genotype calling and locus assembly in Stacks, we implemented specific criteria to ensure robust data analysis: we included loci present across all populations (-p option) and in at least 80% of individuals within each population (-r option), set a minor allele frequency threshold of 3% (--min-mac), and imposed a maximum observed heterozygosity limit of 80%. In our subsequent analyses focused on population genetic structure and FST-based evaluations, we excluded SNPs in linkage disequilibrium (LD) using the --write-random-snp option in Stacks' population program. To further refine our dataset, we employed vcftools v0.1.16 (Danecek et al., 2011) during post-processing to remove variants with extremely low (<5x) or high (>800x) coverage depth and loci with more than 20% missingness.

Population Genetic Structure

To investigate the impact of host specificity on the genetic structure of parasite populations, we conducted genetic clustering analyses on two datasets: all samples in Lineage A (Dataset B) and only samples from Czechia (Dataset D), aiming to compare the genetic structure of parasite populations on both large and local scales. Firstly, we performed a discriminant analysis of principal components (DAPC; Jombart et al. 2010) utilising the adegenet package in R v4.0.5. Using the K-means method, we determined the optimal number of genetic clusters based on the lowest Bayesian Information Criterion (BIC) value. Subsequently, individuals were assigned to clusters using DAPC and the optim.a.score method to establish the number of principal components to retain. We then visualised the individual memberships in each cluster through plots and showcased the clusters using Principal Component Analysis (PCA). Secondly, we employed the evolutionary clustering method in ADMIXTURE (Alexander and Lange, 2011), leveraging the parallel processing capabilities of AdmixPiPe (Mussmann et al., 2020). In this step, we conducted 20 replicates for each K value ranging from 1 to 7. The best K values were determined by the lowest cross-validation error (CV) across replicates, as suggested by Alexander & Lange (2011). The clustering was displayed using the CLUMPAK server (http://clumpak.tau.ac.il/). Lastly, we used fineRADstructure and RADpainter v.0.2 (Malinsky et al., 2018) to explore the population genetic structure based on the nearest-neighbour haplotype. We ran the Stacks2fineRAD.py script from the fineRADstructure

package to calculate the distribution of alleles and SNPs per locus, and the amount of missing data per individual, allowing a maximum of 10 SNPs per locus and limiting individual missingness to 25% during the conversion of the haplotype file to the RADpainter format. Considering the sensitivity of ddRAD to batch effects caused by minor differences between libraries at the size selection step, we examined the potential impact of missing data on any library-based structure. We utilised the fineRADstructure pipeline with default settings but increased the burn-in iterations to 200,000, with 1,000,000 iterations sampled at 1000 intervals. We evaluated convergence by assigning individuals to populations across multiple independent runs, reviewing the plots for the MCMC output of parameter values to ensure consistent convergence on Bayesian posterior distributions, and obtaining effective parameter sample sizes (above 250) by lengthening the duration of each chain. To plot the co-ancestry heatmap, we utilised the "FinestructureLibrary.R" function in the fineRADstructure package (Malinsky et al., 2018).

Coalescent analysis of Speciation modelling

We calculated the composite likelihood of our observed data within a particular model using the site frequency spectrum (SFS) and the simulation method provided by fastsimcoal2 (Excoffier & Foll, 2011; Excoffier et al., 2013). For the simulations, we selected 25 samples for each genetic group representing the highest probability of assignment to their specific genetic cluster (i.e., we excluded highly admixed parasite populations between two genetic clusters; see the Results section). In the genotype calling process for our 50 samples, we utilised the Stacks software, adopting a stringent selection criterion to substantially reduce the incidence of missing data. This approach involved retaining only those loci that were consistently present across all individuals in both populations (r=1 and p=2). After applying these rigorous criteria, the genotype calling process yielded 30,442 SNPs. Then, using the easySFS Python script (Coffman et al., 2016; Gutenkunst et al., 2009), we generated a folded joint SFS, considering a single SNP for each locus to reduce the effects of linkage disequilibrium. The average mutation rate per site per generation (2.89×10^{-9}) and the divergence time events were obtained from Nazarizadeh et al., 2023 (see figure S1). The divergence time between two parasite lineages was estimated at 300 Kyr (Nazarizadeh et al., 2023; 95% highest posterior density interval: 160-457 Kyr). Given the lack of a precise divergence time between the two parasite lineages, we conducted the analysis at the lower bound splits. subsequently, The time of change in demographic events was estimated as a model parameter and allowed to range from 1 to 160 Kyr (generations). For each demographic

model, we utilized fastsimcoal to optimize the fit to the observed multidimensional minor allele SFS via the composite-likelihood method. We employed specific options, starting with -N 100,000 to denote the number of coalescent simulations, followed by -C 10, which sets the threshold for observed SFS entry count by pooling all entries with fewer than 10 SNPs. Subsequently, we applied -L 40, maintaining the number of expectation-maximization (EM) cycles, and -M 0.001 to define the minimum relative difference in parameter values for the stopping criterion. We assigned wide search ranges with log-uniform distributions to all model parameters. To ascertain the parameter estimates that yield the highest likelihood, we executed 500 independent runs of fastsimcoal 2 for each model (Excoffier et al. 2013).We used an information-theoretic model selection approach based on the Akaike's information criterion (AIC) to determine the probability of each model given the observed data. After the maximum likelihood was estimated for each model in every replicate, we calculated the AIC scores. AIC values for each model were rescaled (DAIC) by calculating the difference between the AIC value of each model and the minimum AIC obtained among all competing models. Point estimates of the different demographic parameters for the best-supported model were selected from the run with the highest maximum composite likelihood. Finally, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier et al., 2013).

Gene flow among parasite populations

To evaluate gene flow among parasite populations in sympatric hosts, we analysed unlinked single nucleotide polymorphism (SNP) data from Czechia (Dataset D) to reconstruct interactions among these populations. We employed the Treemix v.1.12 tool (Pickrell and Pritchard, 2012) to investigate gene flow among populations within a phylogenetic framework. A Maximum Likelihood (ML) tree, based on allelic frequency data, was constructed to infer historical migration events between populations. We calculated individual migration events (n) separately. The second double-digest restriction-site associated DNA (ddRAD) dataset was categorized into six clusters according to host specificity, utilizing data from five distinct host-derived parasite populations. We assessed a series of migration events (m), ranging from 1 to 6 (1+ the total number of populations). The optimal model was selected based on the covariance associated with each migration event, and the tree's stability was confirmed through bootstrap analysis using 1,000 SNP blocks. Results from Treemix were visualized using the popcorn package in R.

Genomic signatures of host specific selection

Furthermore, we analysed the contemporary gene flow among parasite groups from different hosts using Bayesian inference, facilitated by the BA3SNP software (Mussmann et al., 2019), which is specifically designed for this type of genetic analysis. The Bayesian framework allowed us to probabilistically estimate gene flow parameters, incorporating uncertainty in our models. Our analysis involved 10 million iterations, with the initial 1 million steps discarded as burn-in to ensure convergence and stability of the results. Data were sampled at every 1,000th step. To enhance the reliability of our findings, we performed cross-validation on the SNP datasets, which helps in assessing the model's predictive accuracy. We adjusted the -a (allele frequencies) and -f (inbreeding coefficients) parameters within the software to achieve optimal acceptance rates of 20 to 60%, respectively (Mussmann et al. 2019).

The first outlier detection method, utilised via the R package pcadapt (v4.3.3; Luu et al., 2017), conducts an individual-based genome scan grounded in PCA. This method does not rely on prior assumptions about population groupings, thus avoiding the need to force admixed individuals into predefined populations. The initial pcadapt run was performed using 20 PCs (K=20). The ideal number of PCs to retain for later tests was determined following Cattell's rule, as described by Luu et al. (2017). The second outlier method used the Outflank r package to determine the distribution of FST for neutral loci, which was then used to assign q-values to each locus to detect outliers that may be due to spatially heterogeneous selection. In this analysis, we set the 'number_of_samples' parameter to 5 (equal to the number of populations sampled), the 'LeftTrimFraction' to 0.08, the 'RightTrimFraction' to 0.30, and maintained the default setting for the Hmin parameter (0.1). The initial threshold for calculating q-values was 0.05, as set by default (Whitlock and Lotterhos, 2015)

Additionally, we utilised the BayeScan v2.192 to pinpoint loci undergoing divergent selection, a process grounded in the variations in allele frequencies across distinct populations. SNP loci with a false discovery rate (FDR) below 0.05 were chosen as outlier SNPs. The analysis included an initial 20 pilot runs, each with 5000 iterations, succeeded by a main phase of 500,000 iterations with a burn-in period of 250,000 steps to guarantee convergence. The default prior odds value of 10 was retained throughout. Based on their alpha values, the loci were categorized: loci significantly above zero were regarded as under directional selection, while those below zero were considered to be under balancing selection (Moore et al., 2014), with the rest being categorized as neutral. The resulting data, which included FST values, were

imported into R software using the BayeScan package for further analysis following Geweke's diagnostic method, known for its efficient convergence diagnostics and outlier detection. This yielded a comprehensive list and tally of outliers, aiding in the visualization and interpretation of selection patterns across the examined populations.

Furthermore, the HapFLK software was employed for a haplotype-based analysis on regions potentially experiencing selection, with the local haplotype cluster (K) set to 20 and the number of iterations increased to 20, balancing accuracy and computational time. Essentially, HapFLK expands on FST-based analysis, pinpointing genomic areas with notable haplotype divergence between individuals from selected populations, while acknowledging the population structure. This method is efficient in identifying recent selective sweeps differentiating the populations under study. Moreover, regions were identified as potential selection areas if at least two successive SNPs showed a nominal p-value less than or equal to 0.05. To ascertain the random appearance of regions fulfilling this criterion, a resampling test was conducted, where SNP positions were shuffled 1000 times to determine the probability of observing sequences with a specific number of consecutive SNPs meeting the defined criteria. Regions with a resampling p-value under 0.05 were retained for additional analysis. Loci identified as outliers by the four methods were deemed "potential outliers" and were presented through Venn diagrams utilizing the "VennDiagram" package in R (Chen and Boutros, 2011).

Supplemental Table

Table S2. Comparison of different divergence scenarios analysed using fastsimcoal2, for two parasite populations associated with *R. rutilus*, *S. erythrophthalmus*, and *A. alburnus* (RSA), and *A. brama* and *B. bjoerkna* (AB) hosts.

Model Parameters	**Search range (log uniform distribution)**	**Allopatry**	**primary contact**	**Secondary contact**	**Isolation with continuous gene flow**
RSA	100-100000	11,561	7,070	7,861	**5,211**
AB	100-100000	48,323	5,541	4,935	**5,332**
MRCA (LineageA)	resize 0.1-10	11,549	3,723	3,029	**3,627**
Time	1- 160,000	NA	72000	9,100	**58,000**
early mig RSA->AB	$1x10^{-6}$-0.01	NA	NA	$3.3x10^{-4}$/1.64	**$4.2x10^{-6}$/0.02**
early mig AB->RSA	$1x10^{-6}$-0.01	NA	NA	$6.8x10^{-4}$/4.00	**$3.3x10^{-4}$/1.96**
recent mig RSA->AB	$1x10^{-6}$-0.01	NA	$4.4x10^{-4}$/2.42	NA	**$9.4x10^{-4}$/5.00**
recent mig AB->RSA	$1x10^{-6}$-0.01	NA	$9.4x10^{-4}$/6.66	NA	**$7.1x10^{-4}$/4.22**
Log10		-2144.5	-1706.8	-1783.9	**-1623.1**
DAIC		96.4	25.4	28.1	**0.00**
AIC		4367.2	3821.6	3867.5	**3150.2**
AIC_w		0.00	0.19	0.14	**0.81**

Supplemental Figures

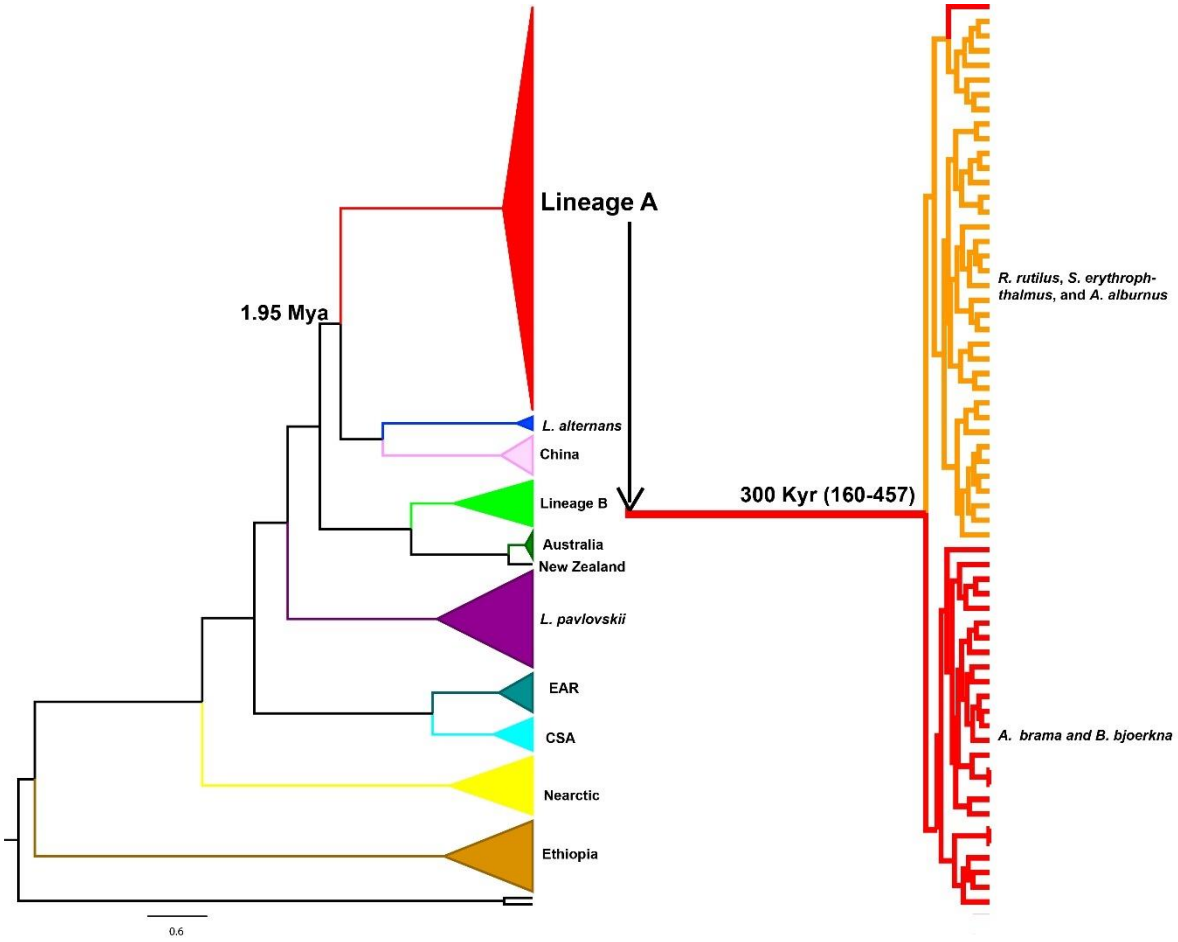


Figure S1. Dated phylogenetic tree derived from ddRAD sequence data, adapted from Nazarizadeh et al. 2023. The tree illustrates the evolutionary relationships among parasite populations. Lineage A, depicted in red, indicates two distinct sub-lineages: one found in *Abramis brama* and *Blicca bjoerkna*, and the other in *Rutilus rutilus*, *Scardinius erythrophthalmus*, and *Alburnus alburnus*. These sub-lineages diverged approximately 300 thousand years ago (Kyr), with the base of Lineage A tracing back to 1.95 million years ago (Mya)

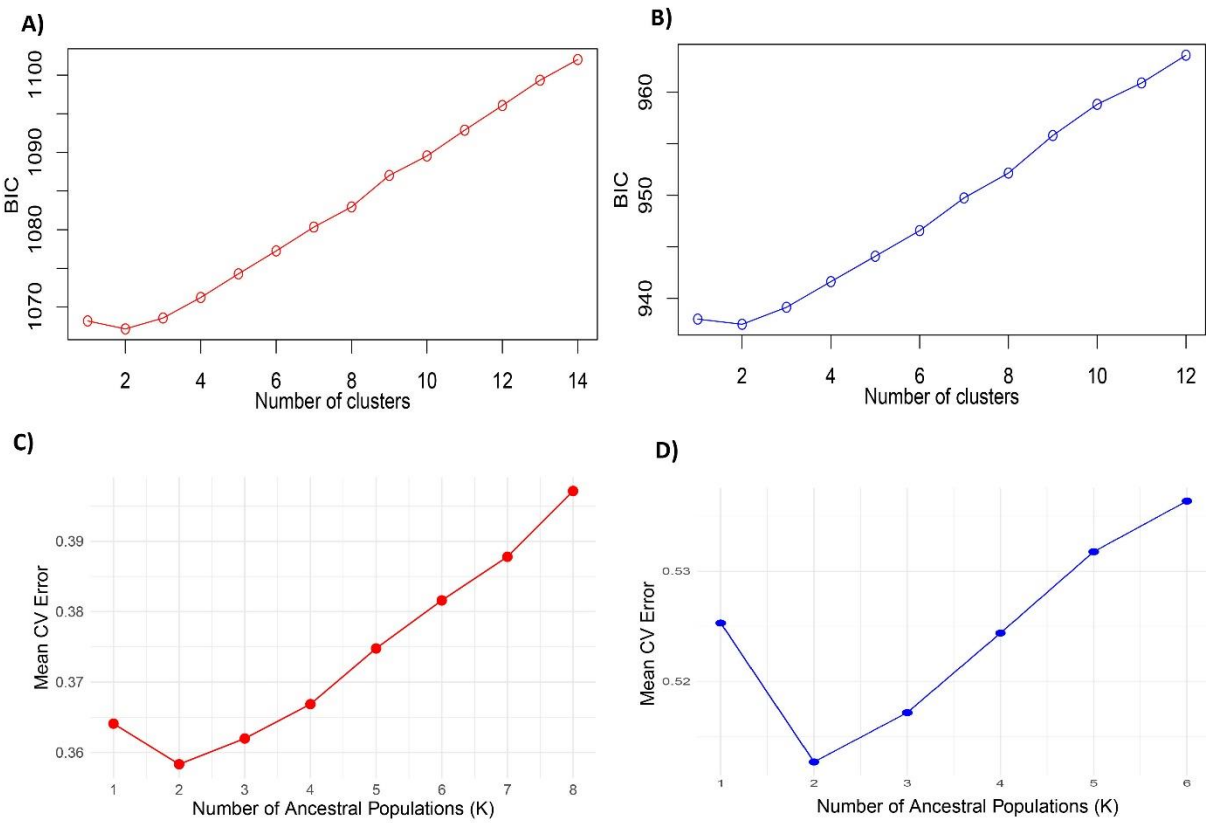


Figure S2. Comparison of genetic clustering Analysis using DAPC and Admixture for Datasets B and D. A) Bayesian Information Criterion (BIC) scores from DAPC for Dataset B indicate an elbow shape for two genetic cluster . B) BIC scores similarly from DAPC for Dataset D also support the existence of two clusters . C) Mean Cross-validation (CV) error from Admixture analysis for Dataset B minimizes at two ancestral populations, indicating the most probable number of genetic clusters in Lineage A across all geographic locations. D) Mean CV error from Admixture analysis for Dataset D shows the lowest point at two ancestral populations, suggesting the optimal number of genetic clusters specific to the Czechia.

A)

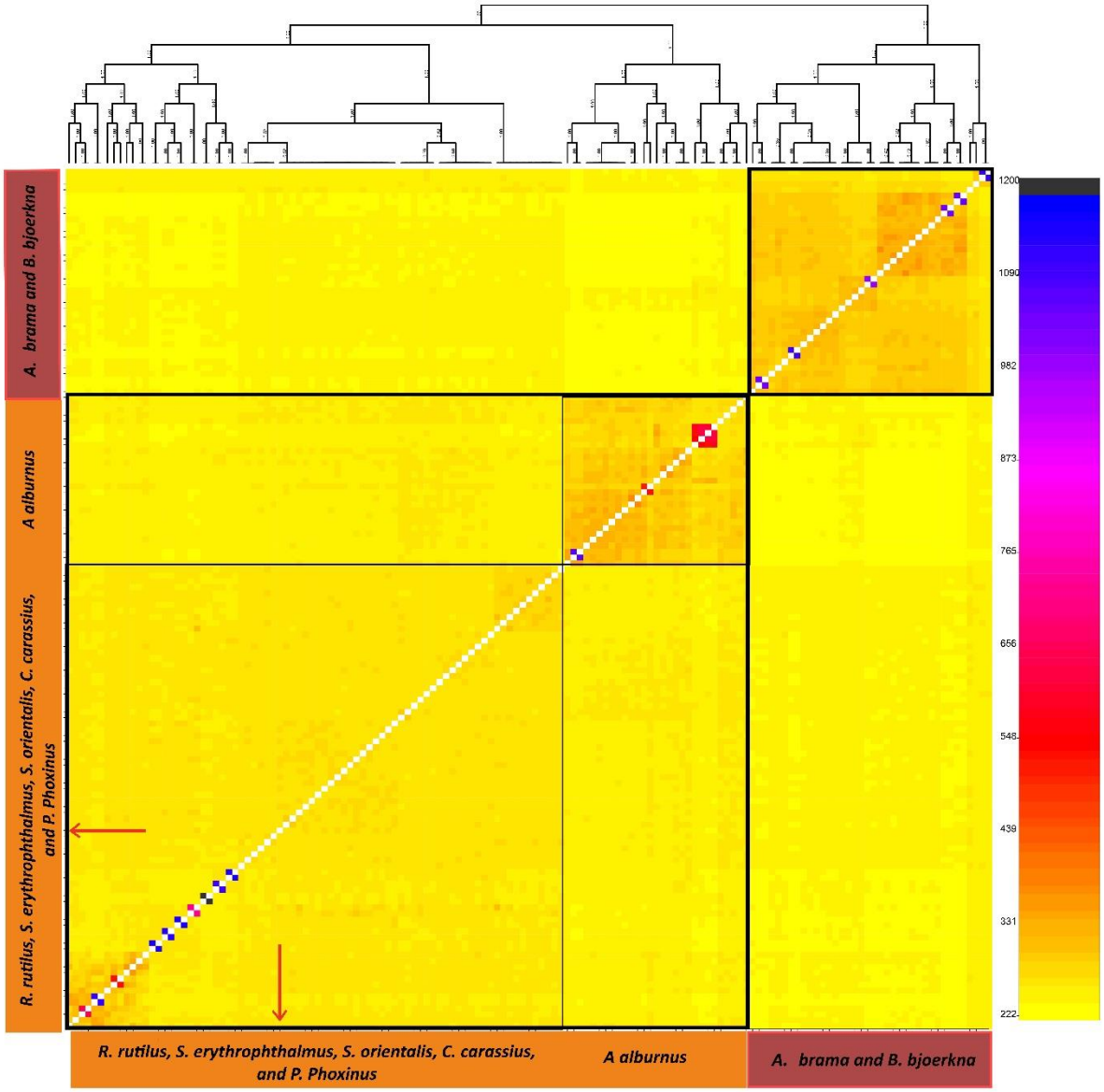


Figure S3 The heatmap generated from FineRADstructure analysis illustrates the genetic differentiation within lineage A of the parasite L. intestinalis, revealing two prominent genetic clusters correlating with distinct host species. The first genetic cluster is specific to hosts *A. brama* and *B. bjoerkna*. In contrast, the second genetic cluster encompasses a broader range of hosts, including *R. rutilus*, *A. alburnus*, *S. erythrophthalmus* , *S. cephalus, C. carassius*, and *P. phoxinus*. Notably, the heatmap highlights a red arrow pointing to a sample from the A. brama and B. bjoerkna cluster, signifying a genetic link with the second cluster.

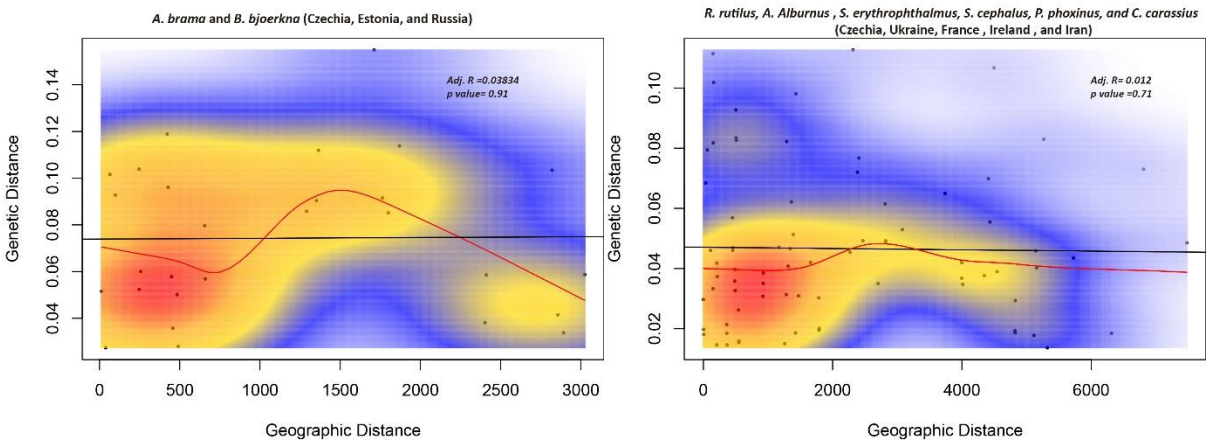


Figure S4. Comparative Mantel tests illustrating the lack of significant correlation between genetic distance and geographic distance in two parasite genetic clusters: one within Abramis brama and Blicca bjoerkna populations across Czechia, Estonia, and Russia (left panel), and another within *Rutilus rutilus, Scardinius erythrophthalmus, Alburnus alburnus, Phoxinus phoxinus*, and *Carassius carassius* populations across Czechia, Ukraine, France, Ireland, and Iran (right panel). Both analyses reveal low and nonsignificant adjusted R-values, indicating that geographic separation does not predict genetic differentiation within these clusters.

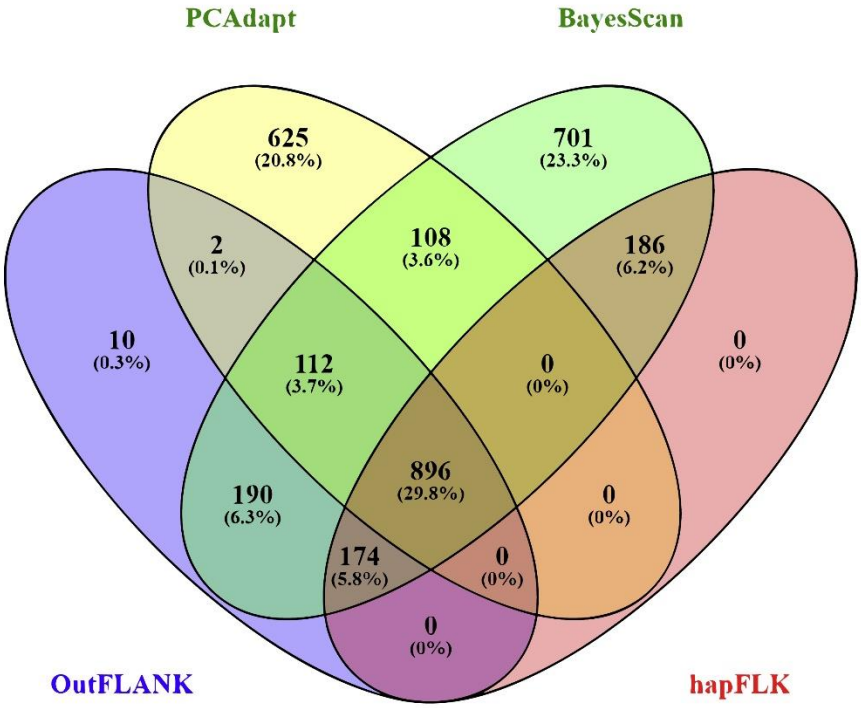


Figure S5. Venn diagram depicting the overlap of genome-wide SNP analyses using four different selection analysis methods: PCAadapt, BayeScan, OutFLANK, and hapFLK. The intersection at the center represents 896 SNPs (29.8% of the total) identified as under selection by all methods, highlighting a consensus in the detection of selective pressures across the genome.

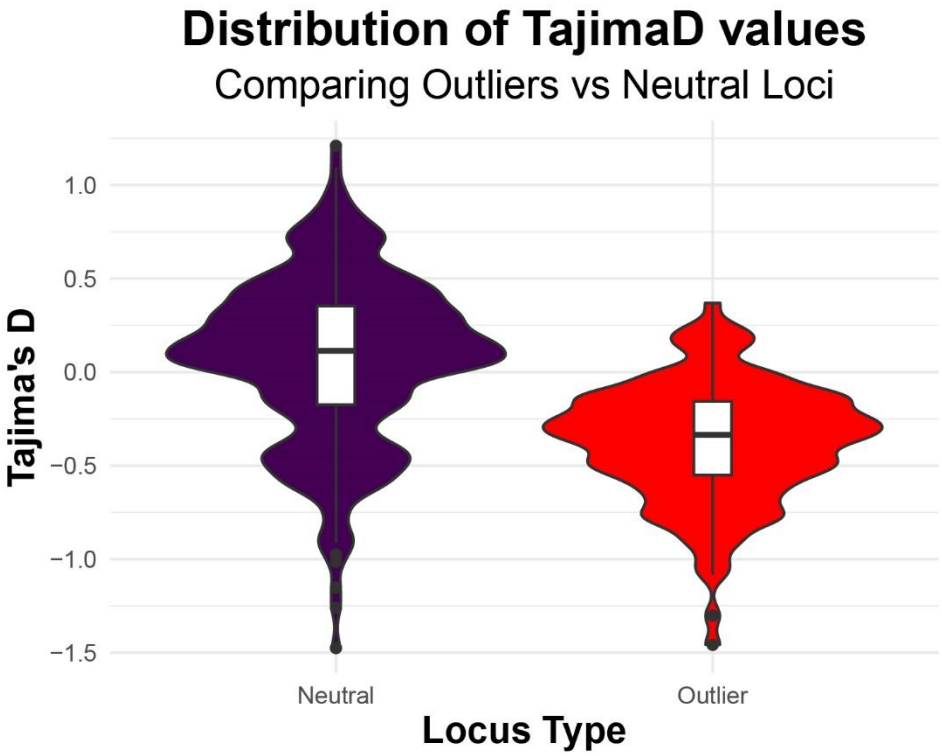


Figure S6 Violin plot contrasting the distribution of Tajima's D values between neutral loci (purple) and outlier loci (red). The data show that outlier loci have a significantly more negative distribution of Tajima's D values, suggesting a deviation from neutrality, potentially due to selective sweeps.

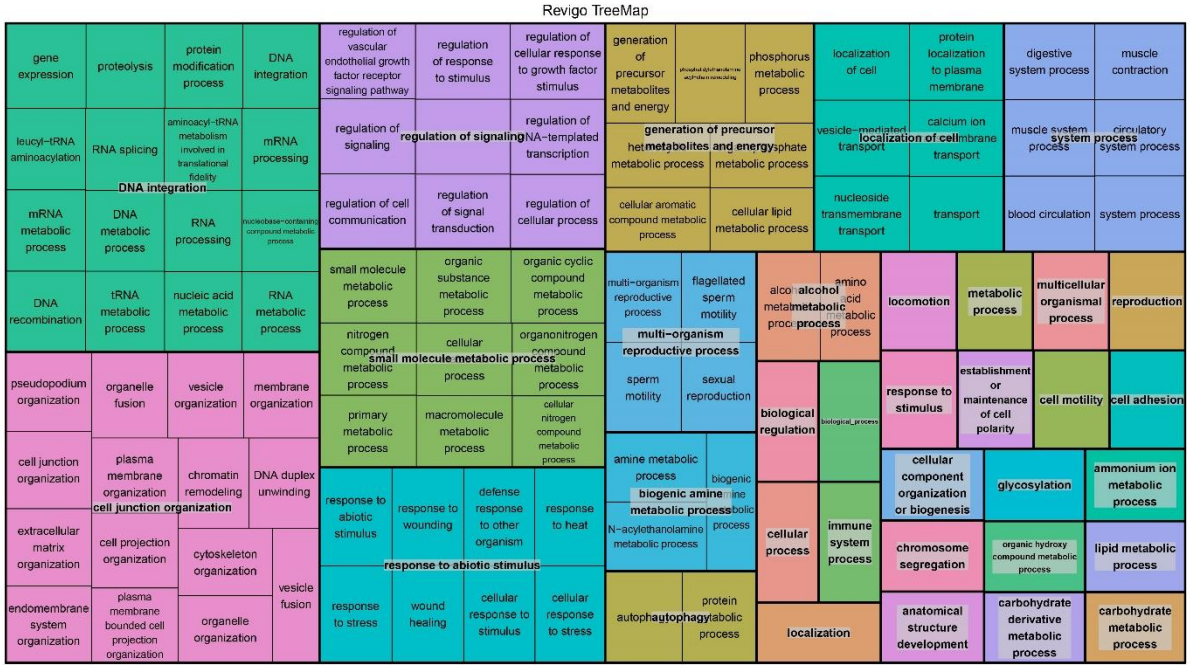


Figure S7. Revigo TreeMap visualization of the functional annotation of SNPs under selection, categorized by biological processes. Each colored block represents a unique cluster of related biological functions, with size indicating the frequency of the process in the dataset. This map highlights the diverse biological processes potentially influenced by selective pressures.

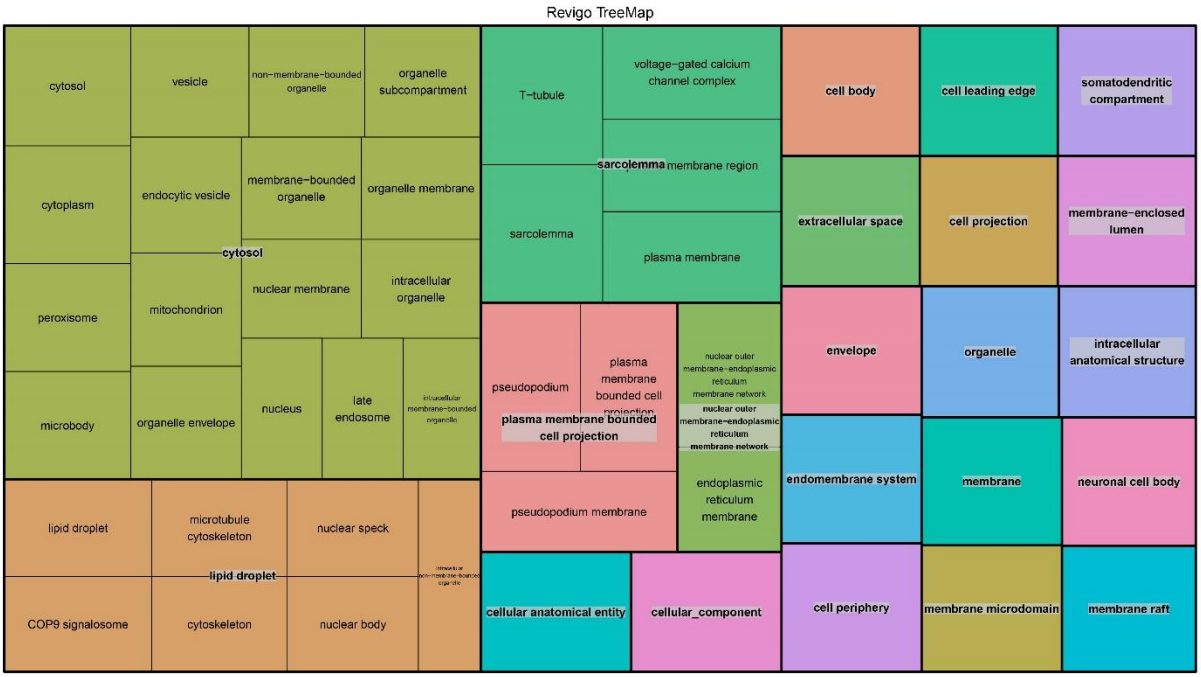


Figure S8 Revigo TreeMap visualization representing the cellular component categorization of SNPs under selection. Each block signifies a distinct cellular component, with size corresponding to the prevalence of SNPs associated with that component in the data. This distribution showcases the complexity of cellular architecture impacted by selection, ranging from broader components like the cytosol and plasma membrane to specific structures such as lipid droplets and organelle membranes

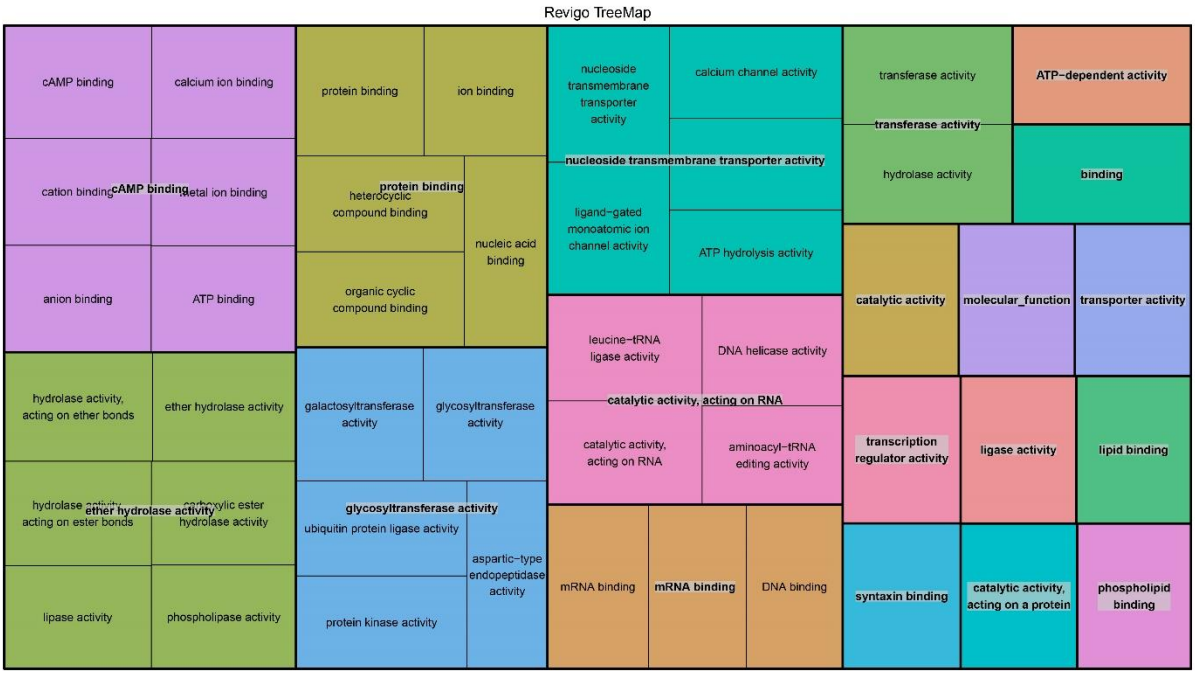


Figure S9. Revigo TreeMap visualization mapping the molecular function categorization of SNPs under selection. Each block represents a distinct molecular function, with size proportional to the occurrence of SNPs associated with that function in the dataset. The map illustrates the intricate network of molecular activities, from binding and catalytic functions to transport and enzyme activities, influenced by selective pressures at the molecular level.

References:

Alexander, D.H., Lange, K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12, 1–6.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Auld, S.K.J.R., Tinsley, M.C., 2015. The evolutionary ecology of complex lifecycle parasites: linking phenomena with mechanisms. Heredity (Edinb). 114, 125–132. https://doi.org/10.1038/hdy.2014.84

Bellis, E.S., von Münchow, C.S., Kronberger, A., Odero, C.O., Kelly, E.A., Xia, T., Huang, X., Wicke, S., Runo, S.M., dePamphilis, C.W., 2022. Genomic signatures of host-specific selection in a parasitic plant. bioRxiv 2002–2022.

Berlocher, S.H., Feder, J.L., 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annu. Rev. Entomol. 47, 773–815.

BioBam, 2019. OmicsBox-Bioinformatics made easy. March 3, 2019.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouzid, W, Lek, S., Mace, M., Ben Hassine, O., Etienne, R., Legal, L., Loot, G., 2008. Genetic diversity of Ligula intestinalis (Cestoda: Diphyllobothriidea) based on analysis of inter-simple sequence repeat markers. J. Zool. Syst. Evol. Res. 46, 289–296.

Bouzid, Wafa, Štefka, J., Hypša, V., Lek, S., Scholz, T., Legal, L., Hassine, O.K. Ben, Loot, G., 2008. Geography and host specificity: Two forces behind the genetic structure of the freshwater fish parasite Ligula intestinalis (Cestoda: Diphyllobothriidae). Int. J. Parasitol. 38, 1465–1479. https://doi.org/10.1016/j.ijpara.2008.03.008

Brunner, F.S., Eizaguirre, C., 2016. Can environmental change affect host/parasite-mediated speciation? Zoology 119, 384–394.

Bush, G.L., 1994. Sympatric speciation in animals: new wine in old bottles. Trends Ecol. Evol. 9, 285–288.

Cháves-González, L.E., Morales-Calvo, F., Mora, J., Solano-Barquero, A., Verocai, G.G., Rojas, A., 2022. What lies behind the curtain: Cryptic diversity in helminth parasites of human and veterinary importance. Curr. Res. Parasitol. Vector-borne Dis. 2, 100094.

Chen, H., Boutros, P.C., 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 1–7.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Clark, N.J., Clegg, S.M., 2017. Integrating phylogenetic and ecological distances reveals new insights into parasite host specificity. Mol. Ecol. 26, 3074–3086.

Coffman, A.J., Hsieh, P.H., Gravel, S., Gutenkunst, R.N., 2016. Computationally efficient composite likelihood statistics for demographic inference. Mol. Biol. Evol. 33, 591–593.

Combes, C., 2001. Parasitism: the ecology and evolution of intimate interactions. University of Chicago Press.

Cruickshank, T.E., Hahn, M.W., 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23, 3133–3157.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10, giab008. https://doi.org/10.1093/gigascience/giab008

de Meeûs, T., Hochberg, M.E., Renaud, F., 1995. Maintenance of two genetic entities by habitat selection. Evol. Ecol. 9, 131–138.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dubinina, M.N., 1980. Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR. Tapeworms (Cestoda, Ligulidae) of the fauna of the USSR.

Duffy, M.A., Brassil, C.E., Hall, S.R., Tessier, A.J., Cáceres, C.E., Conner, J.K., 2008. Parasite-mediated disruptive selection in a natural Daphnia population. BMC Evol. Biol. 8, 80. https://doi.org/10.1186/1471-2148-8-80

Ebert, D., Fields, P.D., 2020. Host–parasite co-evolution and its genomic signature. Nat. Rev. Genet. 21, 754–768.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14, 2611–2620.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C., Foll, M., 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9, e1003905.

Excoffier, L., Foll, M., 2011. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27, 1332–1334.

Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A., Sousa, V.C., 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37, 4882–4885.

Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28, 342–350.

Foll, M., Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180, 977–993.

Funk, D.J., Filchak, K.E., Feder, J.L., 2002. Herbivorous insects: model systems for the comparative study of speciation ecology. Genet. mate choice From Sex. Sel. to Sex. Isol. 251–267.

Giraud, T., Gladieux, P., Gavrilets, S., 2010. Linking the emergence of fungal plant diseases with ecological speciation. Trends Ecol. Evol. 25, 387–395.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695.

Halimi, M., Colagar, A.H., Youssefi, M.R., 2013. Immune response in spirlins (Alburnoides bipunctatus, Bloch 1782) infested by Ligula intestinalis parasite. Vet Ital. 49, 243–246.

Hemstrom, W., Jones, M., 2023. snpR: User friendly population genomics for SNP data sets with categorical metadata. Mol. Ecol. Resour. 23, 962–973.

Henrich, T., Kalbe, M., 2016. The role of prezygotic isolation mechanisms in the divergence of two parasite species. BMC Evol. Biol. 16, 1–10.

Hildreth, M.B., Pappas, P.W., Oaks, J.A., 1997. Effect of tunicamycin on the uptake and incorporation of galactose in Hymenolepis diminuta. J. Parasitol. 555–558.

Hoole, D., Carter, V., Dufour, S., 2010. Ligula intestinalis (Cestoda: Pseudophyllidea): an ideal fish-metazoan parasite model? Parasitology 137, 425–438.

Howard, D.J., Berlocher, S.H., 1998. Endless forms: species and speciation. Oxford University Press on Demand.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314.

Hughes, S.H., 2015. Reverse transcription of retroviruses and LTR retrotransposons. Mob. DNA III 1051–1077.

Hume, J.B., Recknagel, H., Bean, C.W., Adams, C.E., Mable, B.K., 2018. RADseq and mate choice assays reveal unidirectional gene flow among three lamprey ecotypes despite weak assortative mating: Insights into the formation and stability of multiple ecotypes in sympatry. Mol. Ecol. 27, 4572–4590.

Izvekova, G.I., Frolova, T. V, Izvekov, E.I., Kashinskaya, E.N., Solovyev, M.M., 2021. Localization of the proteinase inhibitor activity in the fish cestode Eubothrium rugosum. J. Fish Dis. 44, 1951–1958.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet. 11. https://doi.org/10.1186/1471-2156-11-94

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: Genome-scale protein function classification. Bioinformatics 30, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Le Gac, M., Giraud, T., 2004. What is sympatric speciation in parasites? Trends Parasitol. 20, 207–208.

Le Pennec, G., Retel, C., Kowallik, V., Becks, L., Feulner, P.G.D., 2023. Demographic fluctuations and selection during host-parasite coevolution interactively increase genetic diversity. Mol. Ecol.

Leder, E.H., André, C., Le Moan, A., Töpel, M., Blomberg, A., Havenhand, J.N., Lindström, K., Volckaert, F.A.M., Kvarnemo, C., Johannesson, K., 2021. Post-glacial establishment of locally adapted fish populations over a steep salinity gradient. J. Evol. Biol. 34, 138–156.

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Loot, G., Aulagnier, S., Lek, S., Thomas, F., Guégan, J.-F., 2002. Experimental demonstration of a behavioural modification in a cyprinid fish, Rutilus rutilus (L.), induced by a parasite, Ligula intestinalis (L.). Can. J. Zool. 80, 738–744.

Lumsden, R.D., 1975. Surface ultrastructure and cytochemistry of parasitic helminths. Exp. Parasitol. 37, 267–339.

Luu, K., Bazin, E., Blum, M.G.B., 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. 17, 67–77.

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: population inference from RADseq data. Mol. Biol. Evol. 35, 1284–1290.

Marques, D.A., Meier, J.I., Seehausen, O., 2019. A combinatorial view on speciation and adaptive radiation. Trends Ecol. Evol. 34, 531–544.

Mateus, C.S., Stange, M., Berner, D., Roesti, M., Quintella, B.R., Alves, M.J., Almeida, P.R., Salzburger, W., 2013. Strong genome-wide divergence between sympatric European river and brook lampreys. Curr. Biol. 23, R649–R650.

Mathieu-Bégné, E., Blanchet, S., Mitta, G., Le Potier, C., Loot, G., Rey, O., 2022. Transcriptomic adjustments in a freshwater ectoparasite reveal the role of molecular plasticity for parasite host shift. Genes (Basel). 13, 525.

Matsubayashi, K.W., Kohyama, T.I., Kobayashi, N., Yamasaki, S., Kuwajima, M., Katakura, H., 2017. Genetic divergence with ongoing gene flow is maintained by the use of different hosts in phytophagous ladybird beetles genus Henosepilachna. J. Evol. Biol. 30, 1110–1123.

McCoy, K.D., 2003. Sympatric speciation in parasites–what is sympatry? Trends Parasitol. 19, 400–404.

Mentel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 175–178.

Momigliano, P., Harcourt, R., Robbins, W.D., Jaiteh, V., Mahardika, G.N., Sembiring, A., Stow, A., 2017. Genetic structure and signatures of selection in grey reef sharks (Carcharhinus amblyrhynchos). Heredity (Edinb). 119, 142–153.

Moore, J., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., Chaput, G., Bernatchez, L., 2014. Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. Mol. Ecol. 23, 5680–5697.

Mussmann, S.M., Douglas, M.R., Chafin, T.K., Douglas, M.E., 2019. BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. Methods Ecol. Evol. 10, 1808–1813.

Nazarizadeh, M., Novakova, M., Drabkova, M., Catchen, J.M., Olson, P.D., Stefka, J., 2024. Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of the tapeworm Ligula intestinalis. Proc. R. Soc. B Biol. Sci. 2007–2023.

Nazarizadeh, M., Nováková, M., Loot, G., Gabagambi, N.P., Fatemizadeh, F., Osano, O., Presswell, B., Poulin, R., Vitál, Z., Scholz, T., Halajian, A., Trucchi, E., Kočová, P., Štefka, J., 2023. Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite – The Ligula intestinalis species complex. Mol. Phylogenet. Evol. 180, 107677. https://doi.org/10.1016/j.ympev.2022.107677

Nazarizadeh, M., Peterka, J., Kubečka, J., Vašek, M., Jůza, T., de Moraes, K.R., Čech, M., Holubová, M., Souza, A.T., Blabolil, P., Muška, M., Tsering, L., Bartoň, D., Říha, M., Šmejkal, M., Tušer, M., Vejřík, L., Frouzová, J., Jarić, I., Prchalová, M., Vejříková, I., Štefka, J., 2022. Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of Ligula intestinalis (Cestoda) in Czech water bodies. Folia Parasitol. (Praha). 69, 18. https://doi.org/10.14411/fp.2022.018

Nei, M., 1987. Molecular Evolutionary Genetics Columbia University Press New York 512. Search in.

Nikolakis, Z.L., Carlton, E.J., Pollock, D.D., Castoe, T.A., 2022. A genomic can of worms for schistosome host-specificity. Trends Parasitol.

Nosil, P., 2012. Ecological Speciation, Ecological Speciation. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001

Nosil, P., Harmon, L.J., Seehausen, O., 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24, 145–156.

Olson, E.J., Oaks, J.A., Osmundson, G.D., Hildreth, M.B., 2000. Ultrastructural and lectin-histochemical differences between the scolex/strobila and bladder teguments of the Taenia taeniaeformis strobilocercus. J. Parasitol. 86, 18–24.

Olson, P.D., Littlewood, D.T.J., Griffiths, D., Kennedy, C.R., Arme, C., 2002. Evidence for the co-existence of separate strains or species of Ligula in Lough Neagh, Northern Ireland. J. Helminthol. 76, 171.

Orr, M.R., Smith, T.B., 1998. Ecology and speciation. Trends Ecol. Evol. 13, 502–506.

Papadopoulou, A., Knowles, L.L., 2015. Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a M editerranean P leistocene A ggregate I sland C omplex. Mol. Ecol. 24, 4252–4268.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S.E., Lercher, M.J., 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol. Biol. Evol. 31, 1929–1936.

Pickrell, J., Pritchard, J., 2012. Inference of population splits and mixtures from genome-wide allele frequency data. Nat. Preced. 1.

Poulin, R., 2011. Evolutionary ecology of parasites, Evolutionary ecology of parasites. Princeton university press.

Privé, F., Luu, K., Vilhjálmsson, B.J., Blum, M.G.B., 2020. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. Mol. Biol. Evol. 37, 2153–2154.

Pruitt, K.D., Tatusova, T., Maglott, D.R., 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, D61–D65. https://doi.org/10.1093/nar/gkl842

Rahmatallah, Y., Zybailov, B., Emmert-Streib, F., Glazko, G., 2017. GSAR: Bioconductor package for Gene Set analysis in R. BMC Bioinformatics 18, 1–12.

Räsänen, K., Hendry, A.P., 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. Ecol. Lett. 11, 624–636.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., Ripley, M.B., 2013. Package 'mass.' Cran r 538, 113–120.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28, 4737–4754. https://doi.org/10.1111/mec.15253

Rodriguez-Contreras, D., Skelly, P.J., Landa, A., Shoemaker, C.B., Laclette, J.P., 1998. Molecular and functional characterization and tissue localization of 2 glucose transporter homologues (TGTP1 and TGTP2) from the tapeworm Taenia solium. Parasitology 117, 579–588.

Romero, I.G., Ruvinsky, I., Gilad, Y., 2012. Comparative studies of gene expression and the evolution of gene regulation. Nat. Rev. Genet. 13, 505–516.

Rundle, H.D., Nosil, P., 2005. Ecological speciation. Ecol. Lett. 8, 336–352.

Schluter, D., 1996. Ecological causes of adaptive radiation. Am. Nat. 148, S40–S64.

Simmonds, S.E., Fritts-Penniman, A.L., Cheng, S.H., Mahardika, G.N., Barber, P.H., 2020. Genomic signatures of host-associated divergence and adaptation in a coral-eating snail, Coralliophila violacea (Kiener, 1836). Ecol. Evol. 10, 1817–1837.

Skoracka, A., Kuczyński, L., Szydło, W., Rector, B., 2013. The wheat curl mite Aceria tosichella (Acari: Eriophyoidea) is a complex of cryptic lineages with divergent host ranges: evidence from molecular and plant bioassay data. Biol. J. Linn. Soc. 109, 165–180.

Štefka, J., Hypša, V., Scholz, T., 2009. Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: Microsatellite study of Ligula intestinalis (Cestoda). Mol. Ecol. 18, 1187–1206. https://doi.org/10.1111/j.1365-294X.2008.04074.x

Stelinski, L.L., Liburd, O.E., 2005. Behavioral evidence for host fidelity among populations of the parasitic wasp, Diachasma alloeum (Muesebeck). Naturwissenschaften 92, 65–68.

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLoS One 6, e21800. https://doi.org/10.1371/journal.pone.0021800

Teske, P.R., Sandoval-Castillo, J., Golla, T.R., Emami-Khoyi, A., Tine, M., Von Der Heyden, S., Beheregaray, L.B., 2019. Thermal selection as a driver of marine ecological speciation. Proc. R. Soc. B 286, 20182023.

Thaenkham, U., Chaisiri, K., Hui En Chan, A., 2022. Challenges of Species Identification for Parasitic Helminths, in: Molecular Systematics of Parasitic Helminths. Springer, pp. 131–159.

Thomé, M.T.C., Carstens, B.C., 2016. Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. Proc. Natl. Acad. Sci. 113, 8010–8017.

Tsumura, Y., Uchiyama, K., Moriguchi, Y., Ueno, S., Ihara-Ujino, T., 2012. Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, Cryptomeria japonica. Heredity (Edinb). 109, 349–360.

Villacis-Perez, E., Snoeck, S., Kurlovs, A.H., Clark, R.M., Breeuwer, J.A.J., Van Leeuwen, T., 2021. Adaptive divergence and post-zygotic barriers to gene flow between sympatric populations of a herbivorous mite. Commun. Biol. 4, 853.

Walker, M., Head, M.J., Lowe, J., Berkelhammer, M., BjÖrck, S., Cheng, H., Cwynar, L.C., Fisher, D., Gkinis, V., Long, A., Newnham, R., Rasmussen, S.O., Weiss, H., 2019. Subdividing the Holocene Series/Epoch: formalization of stages/ages and subseries/subepochs, and designation of GSSPs and auxiliary stratotypes. J. Quat. Sci. 34, 173–186.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution (N. Y). 1358–1370.

Whitlock, M.C., Lotterhos, K.E., 2015. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. Am. Nat. 186, S24–S36.

Chapter V

Early evidence of establishment of the tropical bedbug (*Cimex hemipterus*) in Central Europe

Medical and Veterinary Entomology
(DOI: 10.1111/mve.12522)

Medical and Veterinary Entomology

Medical and Veterinary Entomology (2021) **35**, 462–467 doi: 10.1111/mve.12522

Early evidence of establishment of the tropical bedbug (*Cimex hemipterus*) in Central Europe

O. BALVÍN[1], M. SASÍNKOVÁ[1], J. MARTINŮ[2,3], M. NAZARIZADEH[2,3], T. BUBOVÁ[1,4], W. BOOTH[5], E. L. VARGO[6] and J. ŠTEFKA[2,3]

[1]Department of Ecology, Faculty of Environmental Sciences, Czech University of Life Sciences in Prague, Prague, Czech Republic, [2]Institute of Parasitology, Biology Centre CAS, České Budějovice, Czech Republic, [3]Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, [4]National Reference Laboratory for Vector Control, Centre for Epidemiology and Microbiology, National Institute of Public Health, Prague, Czech Republic, [5]Department of Biological Science, The University of Tulsa, Tulsa, OK, U.S.A. and [6]Department of Entomology, Texas A&M University, College Station, TX, U.S.A.

Abstract. In recent decades, the world has witnessed a remarkable resurgence of bedbugs (Hemiptera: Cimicidae). Although populations of the common bedbug, *Cimex lectularius* L., expanded in temperate regions of its original distribution, the tropical bedbug, *C. hemipterus* (F.), increased its abundance in warmer regions, where it also had been historically distributed. However, *C. hemipterus* has recently been observed to be expanding to other areas, e.g. North Australia, Middle East, the United States and Russia. In other parts of Europe, few sporadic and ephemeral introductions of *C. hemipterus* were recorded until recently. We conducted an extensive sampling of European bedbug populations starting in 2002 and found that *C. hemipterus* has recently become locally established. Among 566 examined infestations, nearly all of which involved *C. lectularius*, *C. hemipterus* occurred in six infestations collected since 2019. In at least three cases, the social background of inhabitants of the infested properties indicated that tropical bedbugs likely spread within local communities. Using cytochrome oxidase subunit I, we linked five of the infestations to the most common haplotype found globally, and one to an African haplotype. In all infestations, we observed two *kdr*-associated mutations in the sodium channel gene, which are also commonly found across the world.

Key words. Cytochrome oxidase subunit I, human ectoparasite, insecticide resistance, invasion, *kdr*, mitochondrial network, pest control.

Introduction

After several decades of being almost forgotten, bedbugs have recently experienced a remarkable resurgence around the world. Until recently, pest management companies (PMCs) across Europe have had to deal with only a single human associated species—the common bedbug, *Cimex lectularius*. However, the last few years have witnessed an increase in reports of a second species, the tropical bedbug *C. hemipterus*, which may create new challenges in pest detection and control.

Bedbugs have evolved a very specific strategy of host exploitation. Unlike most other blood-feeding ectoparasites, they do not live on the host body, but rather reside in the shelters of their hosts, visiting the host body only for blood meals (Reinhardt & Siva-Jothy, 2007). Although widely known as human specialists, the common and tropical bedbugs are two species ancestrally associated with bats (Usinger, 1966). In human dwellings, bedbugs are cryptic, remaining hidden and often going unnoticed, until severe infestations develop. Due to this behaviour, monitoring their presence is difficult. Due to the occurrence of

Correspondence: Ondřej Balvín, Department of Ecology, Faculty of Environmental Sciences, Czech University of Life Sciences in Prague, Kamýcká 129, 165 21, Prague 6, Czech Republic. Tel.: +42 073 740 0238; E-mail: o.balvin@centrum.cz

multiple insecticide resistance mechanisms (Dang *et al.*, 2017; Romero, 2018) combined with a limited range of effective insecticides, control becomes challenging, especially in countries with strict regulations on insecticidal compounds (Kilpinen *et al.*, 2011).

Although bedbug spread has been facilitated by urbanization, distribution of the two species appears to be largely mutually exclusive (Zorrilla-Vaca *et al.*, 2015). Historically, the tropical bedbug was distributed in warm regions within the 30° north and south latitudes. It has been the primary bedbug species found in SE Asia or inland India, whereas in other tropical regions both species may occur (Usinger, 1966). The well-documented recent resurgence of bedbugs has comprised both species. Population expansions of *C. hemipterus* first occurred in regions where it prevailed in the past (How & Lee, 2010; Lee, 2013). However, soon after, it became established in other regions, including the Middle East (Ramachandran, 2012; Balfour, 2013; Balvín *et al.*, 2013), North Australia (Doggett *et al.*, 2003; Doggett & Russell, 2008), Florida and Hawaii (Campbell *et al.*, 2016; Lewis *et al.*, 2020).

Across Europe, where *C. lectularius* is widespread, most of the available records of *C. hemipterus* are related to sporadic, ephemeral infestations, largely connected with international travel. An early case in Brno, the Czech Republic, had a clear origin in Nepal (V. Kubáň, 1991, personal communication). Two cases in Northern Italy were reported to originate from travels abroad (Masseti & Bruschi, 2007). Another case from Italy was connected with a visit to Brazil (Masini *et al.*, 2020) and one record was from a plane in Switzerland (M. Schmidt, 2018, personal communication). One of the two cases in Southern France was reported from an apartment occupied by students (Bérenger & Pluot-Sigwalt, 2017); the background of the other case is unknown; similarly, for a case from the United Kingdom (Burgess, 2003). Five cases were reported from Sweden, one from 1981, the rest from 2014 to 2016 (Vinnersten, 2017), with no details on the social background or the travel history of the infested inhabitants given. Nevertheless, particularly frequent travels of the Swedes to Thailand are mentioned, and, at the same time, the author speculates that these cases may indicate a permanent establishment of *C. hemipterus* in Sweden. *Cimex hemipterus*, however, appears to have established in Russia, and likely replaced *C. lectularius* in both Moscow and St. Petersburg (Gapon, 2016). Recent data suggest it has spread further in the European part of Russia (Golub *et al.*, 2020).

The control of bedbugs is particularly problematic not only due to their cryptic behaviour but also due to the evolution of multiple mechanisms that confer resistance to insecticides. In *C. hemipterus*, resistance has been reported multiple times, starting in the 1950s (reviewed in Doggett *et al.*, 2018). The mechanisms of resistance have been examined only for mutations in the sodium channel gene, responsible for the knock-down type of resistance (*kdr*) to pyrethrin, pyrethroids, and some organochlorines. Amino acid changes in the sodium channel gene, both in *C. hemipterus* and *C. lectularius*, are considered a major cause of the bedbug resurgence. In *C. lectularius*, despite some considerable world-wide sampling (Zhu *et al.*, 2010, 2013; Booth *et al.*, 2015; Palenchar *et al.*, 2015; Dang *et al.*, 2015a; Balvín & Booth, 2018; Holleman *et al.*, 2019), only three *kdr*-associated mutations are known, showing unequal distributions at the continent level (Balvín & Booth, 2018; Holleman *et al.*, 2019). In contrast, a total of nine *kdr*-associated mutations were recorded in *C. hemipterus*. Four (L899V, M918I, D953G and L1014F) were found to be unequally distributed across populations in Australia (Dang *et al.*, 2015b) and the remaining five (Y/L995H, V1010L, I1011F, V1016E and L1017F/S) were reported from Sri Lanka (Punchihewa *et al.*, 2019). In the sample representing the invasion of *C. hemipterus* to Hawaii, bedbugs collected in 2009 exhibited three of these mutations, whereas bedbugs found in 2019 possessed four (Lewis *et al.*, 2020).

Specimens of bedbugs are rarely identified to the species level by PMCs in their practice. The specialists in fact have had no reason to do so as only *C. lectularius* has been traditionally known from Europe. It is possible that the identification of *C. hemipterus* in the reported cases was carried out because the tropical bedbug was suspected in connection with trans-continental travel. Also because the two species are difficult to discriminate (Fig. 1), established populations of the tropical bedbug may have been overlooked since the start of the bedbug resurgence or even longer.

Since the resistance levels of *C. hemipterus* compared to *C. lectularius* have not been assessed, the relevance of species identification for the choice of control measures is not currently understood. However, the occurrence of *C. hemipterus* is highly relevant for the consideration of bedbug detection methods. For example, *C. hemipterus* has been shown to have better climbing abilities when compared to *C. lectularius*, due to the presence of a greater number of tenant hairs on its tibial pads (Kim *et al.*, 2017). As such, the higher vertical friction force that these impart enables adults to easily escape the monitors traditionally used for bedbug detection. Knowledge of the species identity of local populations is therefore important.

This study reports on the detection of multiple, likely established populations of *C. hemipterus* in Europe, focusing primarily on Central Europe. Species identification was made through morphological and genetic methods, including both mtDNA and *kdr*-associated mutation presence. Examining the sequences of the cytochrome oxidase subunit I gene of the current samples and using those available in GenBank, we compared the European populations to those sampled in the original areas of distribution. The presence of *kdr*-associated mutations is also reported.

Materials and methods

Records of *C. hemipterus* were made during thorough monitoring of European bedbug populations starting in 2002, with the vast majority of collections derived from two time periods: 2009–2013 and 2019–2020 (Table S1, Fig. S1). The material was gathered with the assistance of multiple PMCs who also provided a brief record of the social background of the affected inhabitants.

For species identity, we examined between 1 and 10 individuals per infestation. Specimens from 2002 to 2018 were identified based on morphological characters (Usinger, 1966).

Fig. 1. Adult females of *Cimex lectularius* (left) and *C. hemipterus* (right). *Cimex lectularius* has a wider body, and broader pronotum relative to head width. The lateral lobes of the pronotum are broad in *C. lectularius*, broadening forwards and rounded in the front, whereas they are narrow, parallel and frontally rather pointed in *C. hemipterus*.

Table 1. Details of *Cimex hemipterus* records reported in the study, collected from Central European homes in 2019–2020.

Country	Locality and social context	Date	No. of individuals	COI GenBank accession no.
Czech Republic	Prague; hostel for foreign workers	02 Dec 2019	2	MW449538
Slovakia	Bratislava; apartment for socially disadvantaged	19 Jul 2019	2	MW449539
Slovakia	Bratislava; private apartment, seniors	09 Jul 2019	2	MW449540
Slovakia	Bratislava; hostel for socially disadvantaged	28 Nov 2019	1	MW449542
Slovakia	Bratislava; private apartment, young family	03 Oct2019	3	MW449541
Switzerland	Genève; hotel	05 Jan2020	3	MW449543

Additionally, for 169 infestations out of 267, the species was confirmed through sequencing of a fragment of the cytochrome oxidase subunit I gene (COI), following the methods outlined in Balvín *et al.* (2012). Samples collected between 2019 and 2020 (299 locations) were identified based on morphology, and those believed to be *C. hemipterus* were confirmed by sequencing the COI. Sequences of *C. hemipterus* were compared to those publicly available in GenBank with the length of 530 bp and more (Table S2). Sequences were shortened for sites with undefined states to 513 bp and an intraspecific mitochondrial network was constructed using the TCS Statistical Parsimony method (Clement *et al.*, 2000) implemented in PopART v.1.7.2 (Bandelt *et al.*, 1999, http://popart.otago.ac.nz).

Specimens identified as *C. hemipterus* were then sequenced for two fragments of the para-type sodium channel gene, in an effort to identify mutations previously connected with *kdr* in *C. hemipterus* (Dang *et al.*, 2015b; Punchihewa *et al.*, 2019). Primer combinations used were BBParaF1/BBParaR1 and BBParaF3/BBParaR3, following Zhu *et al.* (2010). Chromatograms were visualized using CodonCode Aligner version 9.0.1 (CodonCode Corporation, Dedham, MA, USA) and manually inspected.

Results

A total of 566 infestations were analysed, from the years 2002–2020, collected from 15 European countries (Table S1). Of these, 284 samples came from the Czech Republic and Slovakia. Between the years 2002 and 2018, 267 infestations from 15 countries were analysed. Among these, only *C. lectularius* was identified. Since 2019, specimens from 299 sample locations spanning seven countries were analysed. This latter sampling yielded *C. hemipterus* in six infestations from three countries (Table 1). Four cases were found in Bratislava, Slovakia, one each in an apartment and a hostel both rented to socially disadvantaged people, one in an apartment inhabited by seniors, and the final in an apartment occupied by a young family. In Prague, Czech Republic, one case was recorded in a hostel for workers, mostly of Ukrainian origin. In Geneva, Switzerland, *C. hemipterus* was found in a hotel.

Thirteen specimens morphologically identified as *C. hemipterus* were confirmed also by COI sequencing. A dataset of 79 sequences (66 downloaded from GenBank) was collapsed into a network consisting of 19 unique haplotypes (Fig. 2). The resulting network contained one central haplotype

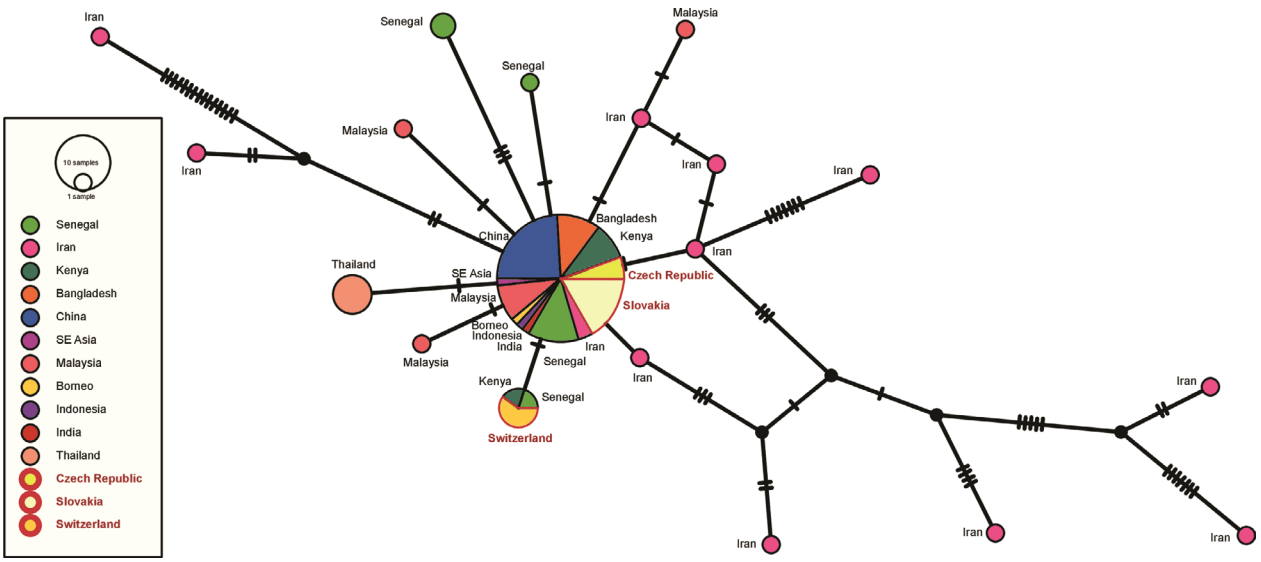


Fig. 2. Mitochondrial network of *Cimex hemipterus* based on cytochrome oxidase subunit I sequences (513 bp) reconstructed using TCS Statistical Parsimony method (Clement *et al.*, 2000). Circle size represents the number of specimens. Crossed lines correspond to the number of nucleotide substitutions. Country names and circles in red represent samples reported in this study.

comprising samples from geographically distant localities, including all of those sampled from the Czech Republic and Slovakia. The samples from Geneva differed from the central haplotype by 2 bp, sharing a haplotype with sequences previously reported from Senegal and Kenya. Senegalese, Malai and Thai haplotypes were separated from the main haplotype by two or more bp mutations. By contrast, Iranian samples created a complex network with more distant haplotypes.

Two *kdr*-associated mutations (M918I and L1014F) were detected in the sodium channel gene, consistently in a homozygous state across all populations examined.

Discussion

The results presented here reveal convincing evidence of the recent establishment of local *C. hemipterus* populations in Central Europe. Among our records dating back to 2002, only in the collections made since 2019 were *C. hemipterus* detected. This suggests that the invasion was not synchronous with the re-emergence of *C. lectularius* across Europe. It thus started only recently and had not been largely overlooked by PMCs, as hypothesized. The establishment of local populations is supported by the information on the social background of the inhabitants of the infested properties, which indicates that the *C. hemipterus* infestations were likely transferred within cities or communities for at least three cases. This is in contrast to the information known for most previous cases recorded from Europe, which, except for findings in Russia and possibly Sweden, indicated rather sporadic and ephemeral introductions, connected with international travel.

Among the cases reported here, three findings from Bratislava were made among inhabitants who were very unlikely to travel to areas with frequent occurrence of *C. hemipterus*. The case from Prague was recorded among foreign workers. However, their nationality was European. One case is likely connected with a travel to areas with a regular occurrence of *C. hemipterus* (Geneva, hotel), and the origin of one case is inconclusive, with an introduction resulting from foreign travel or within-city spread equally plausible (Bratislava, young family).

Our reconstruction of the relationship among European samples with those from the rest of the world showed limited mtDNA variability. Excluding samples from Iran, the network showed a starburst-like pattern, likely reflecting a recent population expansion. However, the number of recorded mutations suggests that the expansion occurred deep in the past and is not connected to the recent resurgence. One of the likely scenarios behind this pattern may be the expansion of host species from bats to humans. The distance of the Iranian samples (Samiei *et al.*, 2020) from the rest may suggest a deeper history of association of *C. hemipterus* with humans in the region. However, such diversity is highly unlikely to stay restricted to one country through the history of humans.

The mitochondrial data, in concert with the single *kdr*-associated variant found, provide only limited information on the origin of the European population, when compared with data available in GenBank. The COI haplotype found in infestations sampled in the Czech Republic and Slovakia was previously reported across many countries and appears to be the most common COI variant of *C. hemipterus*, at least in Asia. Similarly, the sodium channel variant, containing M918I and L1014F mutations, has been reported multiple times, from India and Australia (Dang *et al.*, 2015b), and from China (Zhao *et al.*, 2020), but not in Hawaii (Lewis *et al.*, 2020) or Sri Lanka (Punchihewa *et al.*, 2019). Although the country of origin cannot be pinpointed accurately at this time, the genetic uniformity of the Czech and Slovakian infestations may suggest a single introduction followed by spread. However, this is speculative,

as they represent the most common and widespread variants for the examined loci in the original areas of distribution. The Swiss finding is linked by its COI haplotype to two African countries, suggesting possible import from this continent.

The presence of M918I and L1014F mutations suggests that the new European populations of *C. hemipterus* are resistant to pyrethroids, however, the level of resistance has yet to be determined. Nevertheless, the cooperating PMCs reported no problems controlling the sampled infestations using treatments commonly implemented for *C. lectularius* control. Although the authors are not aware of current insecticidal treatments failing to control *C. hemipterus* that would otherwise be effective for the control of *C. lectularius*, insecticide resistance mechanisms are diverse and there is a potential for new, species-specific forms to evolve.

In conclusion, the authors document a relatively recent and, for now, low level invasion of the *C. hemipterus* into Central Europe. Within Europe, the first occupied areas were Russia and possibly Sweden. It is possible that the climate had been originally a reason for different areas of distribution of the two bedbug species. The current situation indicates that climate no longer constrains bedbug distribution and that indoor conditions, including central heating and/or the use of indoor humidifiers, mostly used in bedrooms, may facilitate the spread of both bedbug species, including the presumably tropical *C. hemipterus*. At this moment, no evidence suggests that the European population of *C. hemipterus* is harder to control than *C. lectularius*. However, in an area with limited spectrum of insecticides available on the market, the occurrence of two species may represent a broader genetic pool, upon which evolution may yield further resistance mechanisms. Currently, it is clear that the successful establishment of invasions is in its early stages; however, understanding the occurrence and spread of *C. hemipterus* is important with respect to bedbug detection methods. With regard to that, bedbug-detection dogs trained for *C. lectularius* have been found to be capable of locating other cimicid species, specifically *C. hemipterus* and *C. pipistrelli* (A. Hamker, J. Hofferová, 2021, personal communication). However, bedbug monitoring using conventional climb-up type traps may fail to identify early-stage infestations of *C. hemipterus* due to their greater climbing ability. The reliability of detection is important in evaluating the efficiency of bedbug control measures, but also in the monitoring of eventual further spread of the tropical bedbug. Therefore, we recommend that PMCs make efforts to identify bedbug species at least occasionally, particularly in cases where traditional treatments have failed.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Map of the infestations examined in this study.

Table S1. List of bedbug material examined.

Table S2. List of GenBank accession codes of COI sequences of *Cimex hemipterus* compared with our samples and used in PopArt genealogy reconstruction.

Acknowledgements

We are deeply thankful to several dozen pest control specialists, colleagues, and friends who helped us collect bedbug material for the population studies. Those that contributed to the material examined in this study are listed as collectors in Table S1. The study was supported by a grant from the Ministry of Education, Youth and Sports of the Czech Republic InterAction no. LTAUSA18032 (Czech Republic based authors), a grant from the National Science Foundation (DEB-1754394) (Warren Booth), a grant from Bayer Crop Sciences and the Urban Entomology Endowment at Texas A&M University (Ed Vargo). All authors declared that they have no conflicts of interests regarding the presented data.

Author's contributions

M. Sasínková, T. Bubová and O. Balvín organized the sample collection. O. Balvín identified the specimens morphologically. O. Balvín, J. Martinů and M. Sasínková carried out the molecular investigations. All authors took part in interpretation of the molecular data and writing the manuscript. All authors gave final approval for publication.

Data availability statement

All data used in study are enclosed as supplementary files, with nucleotide sequences deposited at GenBank with reference codes in Supplementary Table 2.

References

Balfour, J. (2013) *Arthropod Public Health Pests in the Emirates. How to Recognize them, How to Control Them*. Gulf Printing Press, Dubai.

Balvín, O. & Booth, W. (2018) Distribution and frequency of pyrethroid resistance-associated mutations in host lineages of the bed bug (Hemiptera: Cimicidae) across Europe. *Journal of Medical Entomology*, **55**, 923–928.

Balvín, O., Kment, P. & Linnavuori, R.E. (2013) Order Hemiptera, suborder Heteroptera, infraorder Cimicomorpha, family Cimicidae *Arthropod Fauna of the UAE* (ed. by A. Van Harten), pp. 36–68. Dar Al Ummah Printing, Abu Dhabi.

Balvín, O., Munclinger, P., Kratochvíl, L. & Vilímova, J. (2012) Mitochondrial DNA and morphology show independent evolutionary histories of bedbug *Cimex lectularius* (Heteroptera: Cimicidae) on bats and humans. *Parasitology Research*, **111**, 457–469.

Bandelt, H.J., Forster, P. & Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bérenger, J.M. & Pluot-Sigwalt, D. (2017) Présence en France de la Punaise de lit tropicale, *Cimex hemipterus* (Fabricius, 1803) (Hemiptera, Heteroptera, Cimicidae). *Bulletin de la Société entomologique de France*, **122**, 423–427.

Booth, W., Balvín, O., Vargo, E.L., Vilímová, J. & Schal, C. (2015) Host association drives genetic divergence in the bed bug, *Cimex lectularius*. *Molecular Ecology*, **24**, 980–992.

Burgess, I. (2003) Bugs of the past—or are they on the up? *Professional Pest Control, Spring*, **2003**, 16–17.

Campbell, B.E., Hoehler, P.G., Buss, L.J. & Baldwin, R.W. (2016) Recent documentation of the tropical bed bug (Hemiptera: Cimicidae) in Florida since the common bed bug resurgence. *Florida Entomologist*, **99**, 549–551.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Dang, K., Doggett, S.L., Veera Singham, G. & Lee, C.-Y. (2017) Insecticide resistance and resitance mechanisms in bed bugs, *Cimex* spp. (Hemiptera: Cimicidae). *Parasites and Vectors*, **10**, 318.

Dang, K., Toi, C.S., Lilly, D.G., Bu, W. & Doggett, S.L. (2015a) Detection of knockdown resistance mutations in the common bed bug, *Cimex lectularius* (Hemiptera: Cimicidae), in Australia. *Pest Management Science*, **71**, 914–922.

Dang, K., Toi, C.S., Lilly, D.G., Lee, C.-Y., Naylor, R. & Doggett, S.L. (2015b) Identification of putative *kdr* mutations in the tropical bed bug, *Cimex hemipterus* (Hemiptera: Cimicidae). *Pest Management Science*, **71**, 1015–1020.

Doggett, S.L., Geary, M.J., Crowe, W.J., Wilson, P. & Russell, R.C. (2003) Has the tropical bed bug, *Cimex hemipterus* (Hemiptera: Cimicidae), invaded Australia? *Environmental Health Journal*, **3**, 80–82.

Doggett, S.L., Miller, D.M. & Lee, C.Y. (2018) *Advances in the Biology and Management of Modern bed Bugs*. Wiley-Blackwell, Hoboken.

Doggett, S.L. & Russell, R.C. (2008) The resurgence of bed bugs, *Cimex* spp. (Hemiptera: Cimicidae) in Australia. *Proceedings of the Sixth International Conference on Urban Pests*. OOK-Press, Budapest.

Gapon, D. (2016) First records of the tropical bed bug *Cimex hemipterus* (Heteroptera: Cimicidae) from Russia. *Zoosystematica Rossica*, **25**, 239–242.

Golub, V.B., Aksenenko, E.V., Soboleva, V.A. & Kornev, I.I. (2020) New data on the distribution of the tropical bed bug *Cimex hemipterus* and the western conifer seed bug *Leptoglossus occidentalis* (Heteroptera: Cimicidae, Coreidae) in the European part of Russia. *Russian Journal of Biological Invasions*, **11**, 97–100.

Holleman, J.G., Robison, G.A., Bellovich, I.J. & Booth, W. (2019) Knockdown resistance-associated mutations dominate populations of the common bed bug (Hemiptera: Cimicidae) across the south central United States. *Journal of Medical Entomology*, **56**, 1678–1683.

How, Y. & Lee, C. (2010) Survey of bed bugs in infested premises in Malaysia and Singapore. *Journal of Vector Ecology*, **35**, 89–94.

Kilpinen, O., Kristensen, M. & Vagn Jensen, K.M. (2011) Resistance differences between chlorpyrifos and synthetic pyrethroids in *Cimex lectularius* population from Denmark. *Parasitology Research*, **109**, 1461–1464.

Kim, D.Y., Billen, J., Doggett, S.L. & Lee, C.Y. (2017) Differences in climbing ability of *Cimex lectularius* and *Cimex hemipterus* (Hemiptera: Cimicidae). *Journal of Economic Entomology*, **110**, 1179–1186.

Lee, C. (2013) *Bed Bugs in Asia – Perspective from Southeast Asia*. Global Bed Bug Summit, Denver, CO.

Lewis, C.D., Levine, B.A., Vargo, E.L., Schal, C. & Booth, W. (2020) Recent detection of multiple populations of the tropical bed bug (Hemiptera: Cimicidae) exhibiting *kdr*-associated mutations in Hawaii. *Journal of Medical Entomology*, **57**, 1077–1081.

Masini, P., Zampetti, S., Miñón Llera, G., Biancolini, F., Moretta, I. & Sringeni, L. (2020) Infestation by the tropical bedbug *Cimex hemipterus* (Hemiptera: Cimicidae): first report in Italy. *Journal of the European Academy of Dermatology and Venereology*, **34**, e28–e30.

Masseti, M. & Bruschi, F. (2007) Bedbug infestations recorded in Central Italy. *Parasitology International*, **51**, 81–83.

Palenchar, D.J., Gellatly, K.J., Yoon, K.S., Mumcuoglu, K.Y., Shalom, U. & Clark, J.M. (2015) Quantitative sequencing for the determination of *kdr*-type resistance allele (V419L, L925I, I936F) frequencies in common bed bug (Hemiptera: Cimicidae) populations collected from Israel. *Journal of Medical Entomology*, **52**, 1018–1027.

Punchihewa, R., de Silva, W.A.P.P., Weeraratne, T.C. & Karunaratne, S.H.P.P. (2019) Insecticide resistance mechanisms with novel '*kdr*' type gene mutations in the tropical bed bug *Cimex hemipterus*. *Parasites & Vectors*, **12**, 310.

Ramachandran, D. (2012) Regional challenges associated with bed bug control. *NPMA Pest World East Conference*, Dubai, UAE.

Reinhardt, K. & Siva-Jothy, M.T. (2007) Biology of the bed bugs (Cimicidae). *Annual Review of Entomology*, **52**, 351–374.

Romero, A. (2018) Insecticide resistance *Advances in the Biology and Management of Modern Bed Bugs* (ed. by S.L. Doggett, D.M. Miller & C.-Y. Lee), pp. 273–280. John Wiley & Sons Ltd, Oxford.

Samiei, A., Tavassoli, M. & Mardani, K. (2020) The phylogenetic analysis of *Cimex hemipterus* (Hemiptera: Cimicidae) isolated from different regions of Iran using cytochrome oxidase subunit I gene. *Journal of Arthropod-Borne Diseases*, **14**, 239–249.

Usinger, R.L. (1966) *Monograph of Cimicidae*. Entomological Society of America, Washington, DC.

Vinnersten, P. (2017) *Cimex hemipterus* (Fabricius, 1803) enetablerad vägglusarti Sverige (Heteroptera, Cimicidae). *Entomologisk Tidskrift*, **138**, 67–70.

Zhao, Y., Feng, X., Li, M. & Qiu, X. (2020) The double-mutation (M918I + L1014F) *kdr* allele is fixed in *Cimex hemipterus* populations in Guangxi, China. *Bulletin of Entomological Research*, **110**, 506–511.

Zhu, F., Gujar, H., Gordon, J.R., Haynes, K.F., Potter, M.F. & Palli, S.R. (2013) Bed bugs evolved unique adaptive strategy to resist pyrethroid insecticides. *Scientific Reports*, **3**, 1456.

Zhu, F., Wigginton, J., Romero, A., Moore, A., Ferguson, K. & Palli, S.R. (2010) Widespread distribution of knockdown resistance mutations in the bed bug, *Cimex lectularius* (Hemiptera: Cimicidae), populations in the United States. *Archives of Insect Biochemistry and Physiology*, **73**, 245–257.

Zorrilla-Vaca, A., Silva-Medina, M.M. & Escandón-Vargas, K. (2015) Bedbugs, *Cimex* spp.: their current world resurgence and healthcare impact. *Asian Pacific Journal of Tropical Disease*, **5**, 342–352.

Accepted 16 April 2021
First published online 4 May 2021

Supplemental information

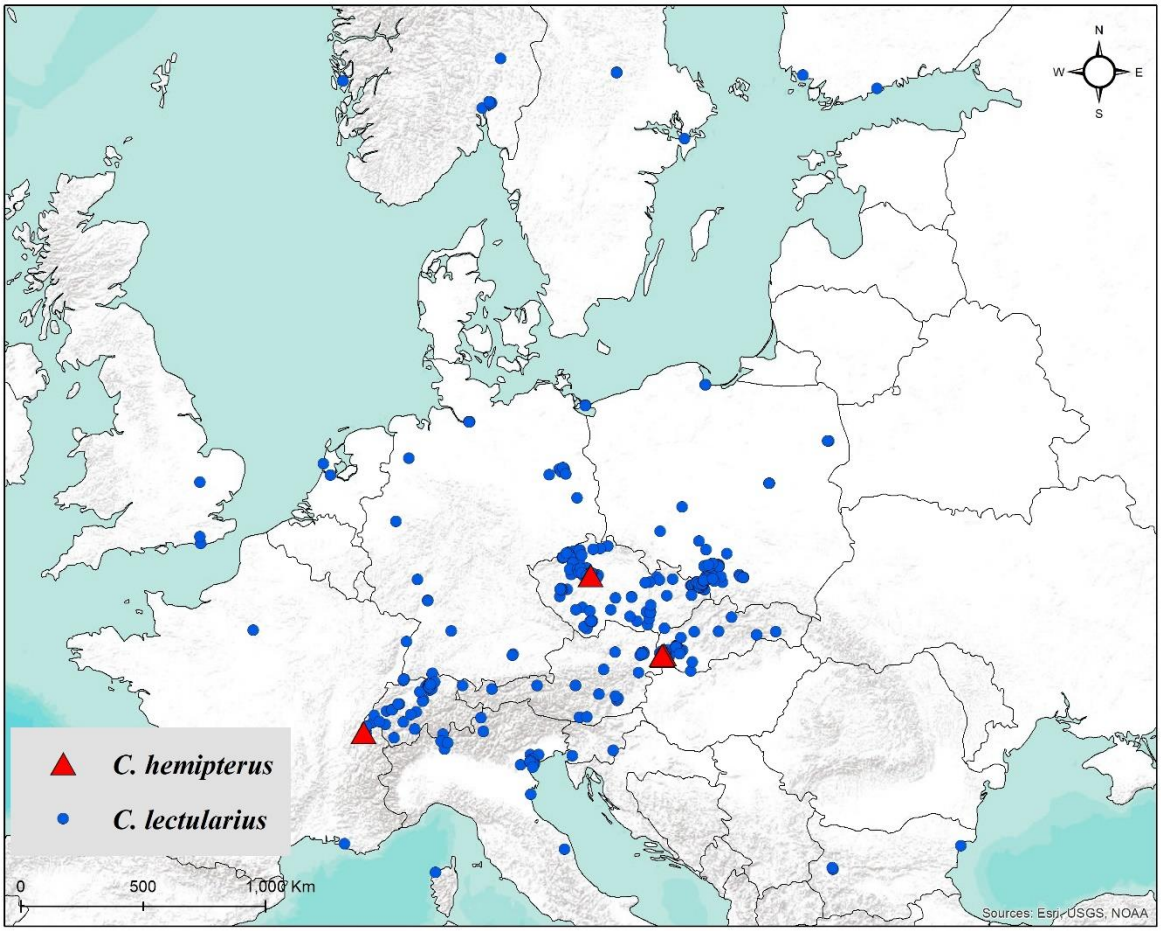


Fig S1 The distribution map of *C. lectularis* and *C. hemipterus* in Europe

Chapter VI

Host Influence on Genetic Diversity and Adaptability in *Cimex lectularius*: Insights from Genome-Wide SNP and Transcriptome Analyses

Unpublished manuscript

Host Influence on Genetic Diversity and Adaptability in *Cimex lectularius*: Insights from Genome-Wide SNP and Transcriptome Analyses

Masoud Nazarizadeh[1,2], Markéta Sasínková[3], Jana Martinů[1], Milena Nováková[2], Jan Štefka[1,2], Jana Křemenová[4], Tomáš Bartonička[4], Ondřej Balvín[3]

[1]Faculty of Science, University of South Bohemia, Ceské Budejovice, Czech Republic

[2] Institute of Parasitology, Biology Centre CAS, Ceské Budejovice, Czech Republic

[3]Department of Ecology, Faculty of Environmental Sciences, Czech University of Life Sciences, Prague, Czech Republic,

[4]Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlarska 2 Brno, Czech Republic

Abstract

In the present study, we investigate genetic and transcriptomic divergence in *Cimex lectularius* populations parasitizing humans and bats. We explore how host switching through ecological fitting allows parasites to exploit different host species using their existing traits, potentially leading to speciation. Using genome-wide SNP and transcriptome analyses, we revealed significant genetic differences between human- and bat-associated bedbugs. Bat-associated bedbugs (BL), presumably representing the original host association, showed higher genetic diversity and lower inbreeding compared to human-associated bedbugs (HL), attributed to the ancient genetic background of the host lineage and the stability of local bed bug populations. Conversely, human-associated bedbugs faced genetic bottlenecks due to stable environments and pest control measures, resulting in lower diversity and higher inbreeding rates. Transcriptomic analysis identified distinct gene expression profiles related to blood digestion enzymes. BL bedbugs exhibited significant and numerous gene expression changes when switching from bat to human blood, while HL bedbugs showed minimal changes, indicating versatile physiological mechanisms. Additionally, HL bedbugs had enhanced odorant binding capabilities, suggesting an evolutionary advantage in variable host environments These findings highlight the impact of host association on bedbug genetic structure and adaptability, emphasizing the need for pest management strategies that consider genetic diversity and

evolutionary dynamics. Our study provides insights into the molecular basis of host specialization and parasite adaptation.

Keywords: *Cimex lectularius*, Bat-associated bedbugs, human-associated bedbugs, host switching

Introduction

Host switching is a dynamic evolutionary process in which a parasite shifts to a new host species. This process drives microevolutionary adaptations at the genetic level in the short term and can potentially lead to macroevolutionary outcomes, such as the emergence of new species, over longer timescales (D'Bastiani et al., 2023; Favret and Voegtlin, 2004; Nazarizadeh et al., 2023). During this process parasites adapt to different hosts in a sloppy fitness space, utilizing pre-existing traits to access new resources and occasionally modifying these traits to more effectively exploit the new host environment (Araujo et al., 2015; Brooks et al., 2006). Factors such as the genetic plasticity of the parasite, ecological availability and susceptibility of potential new hosts, and the presence of vectors play key roles in the success of host switching (Agosta and Brooks, 2020; Araujo et al., 2015; Brooks et al., 2019; De Fine Licht, 2018). Moreover, host switching explains the incongruence in co-speciation between hosts and parasites, where the evolutionary paths of the host and the parasite tangle uprather than mirror each other (Agosta and Brooks, 2020; Brooks et al., 2019). This incongruence often occurs because parasites can switch hosts more frequently than hosts can evolve new defence mechanisms, leading to a mismatch in evolutionary rates (Doña et al., 2017; Huyse et al., 2005; Kvičerová and Hypša, 2013). As a result, pathogen-host shifts are of considerable concern for humans, wildlife, and agriculture, with significant economic and public health impacts that threaten food biosecurity and human health (Brooks et al., 2019). Consequently, understanding the factors influencing the success of host-switching and subsequent speciation events is important for understanding parasite diversification.

The study of host switching is critically enhanced by the use of genomic, transcriptomic, and population genomic approaches (Cruz and Freitas-Castro, 2019; Huo et al., 2021; Nazarizadeh et al., 2023; Wang et al., 2016). These methods provide in-depth insights into how parasites adapt to new hosts through genetic and molecular mechanisms (Nazarizadeh et al., 2024b). For instance, population genomics can offer a broader perspective on how these adaptations might vary across different populations of the same species, highlighting the role of genetic diversity in the success of host switching (Wang et al., 2016). Additionally, transcriptomics further

complements these insights by examining changes in gene expression that occur as the parasite interacts with and adapts to a new host (Nazarizadeh et al., 2024a; Xia et al., 2017). These techniques elucidate the complex genetic and molecular foundations enabling parasites to overcome host-specific barriers, develop new ecological niches, and possibly lead to speciation (Nazarizadeh et al., 2024b). Phenotype plasticity mediated by differentially expressed genes (DEGs) is understood as one of the major mechanisms of population differentiation and, consequently, ecological speciation (Amwoma et al., 2024; Benoit et al., 2016; Birget et al., 2017; Chen et al., 2017; Talbot et al., 2017). Therefore, studying these adaptations not only sheds light on the evolutionary processes driving host switching but also helps predict and manage potential outbreaks by understanding the factors that contribute to the emergence of new parasite-host interactions.

The common bedbug, *Cimex lectularius*, is member of a specialized blood-feeding family of Heteroptera. This species, along with the tropical bedbug (*C. hemipterus*), was originally associated with bats, and has successfully adapted to human environment, a host shift believed to have occurred approximately 245,000 years ago (Balvín et al., 2012). While many bedbug species show strong host specificity, often linked to specific bat or bird families, the common bedbug displays more generalist tendencies (Roth et al., 2019; Talbot et al., 2019). It is capable of utilizing almost any warm-blooded host but has formed permanent associations primarily with humans (human-associated lineage, HL), bats (bat-associated lineage, BL), and likely pigeons and poultry, though the latter associations are less documented due to scarce data (Usinger, 1966).

Recent population genetic studies, analyzing two fragments of mitochondrial genomes, showed a high number of shared haplotypes (Balvín et al., 2012; Booth et al., 2015). However, analyses of microsatellite loci indicated significant genetic differentiation between the bat and human lineages, despite their presence in sympatry (Booth et al., 2015; Sasínková et al., 2023). Similar to the findings from the microsatellite data, notable morphological differences were observed between the two lineages (Balvín et al., 2012). However, Sasínková et al. (2023) found no post-copulation reproductive barriers, as crosses between the two lineages produced viable and fertile offspring. Additionally, there were no significant differences in the number of eggs laid or the fitness of the offspring between crosses within the same lineage and those between lineages (DeVries et al., 2020).

Both the human-associated and bat-associated lineages of bedbugs successfully reproduce on both human and bat hosts. However, HL exhibits slower development rates (Wawrocka and Bartonička, 2013) and consumes fewer blood meals when reared on bats (Sasínková et al., 2023). This suggests that HL has become better adapted to human hosts, particularly in terms of feeding, and shows inadequate adaptation to its ancestral bat hosts. Additionally, HL may have developed longer and thinner legs (Balvín et al., 2012) to facilitate active dispersal and adapt to moving between apartments (Cooper et al., 2015; Wang et al., 2010). In contrast, BL retained short and strong legs (Balvín et al. 2012) in order to cling on bat body for dispersal, as the active dispersal is impossible due to the spatial isolation of individual bat roosts. Given the lengthy period of mutual isolation between these lineages, observing such adaptations is not surprising. Despite their largely sympatric distribution, the reasons behind their enduring separation remain unclear.

Here, we aim to uncover the mechanisms underlying speciation via host switching in *C. lectularius* in both BL and HL populations, where all hosts can coexist in sympatry. The primary objective of our study was to differentiate between genetically determined changes in gene expression, anticipated to arise from the divergence of host lineages and adaptation to new hosts, and phenotypic plasticity, which demonstrates the parasite's capacity for host switching and evolutionary potential. Utilizing a comprehensive approach that includes the analysis of genome-wide SNPs from double digest restriction-site associated DNA (ddRADseq) and transcriptome data, we aim to: ascertain whether host specificity in bedbugs influences the population structure of parasites, and reconstruct transcriptome profiles and identify DEGs between two host-associated populations of *C. lectularius*. Additionally, to further elucidate the complex dynamics of parasite evolution, particularly in scenarios where host switching occurs between humans and bats, we aim to analyse specific genes that contribute to phenotypic plasticity and potential pre-adaptations.

Material and methods

Sampling and in vitro cross feeding experiment

To assess the level of population structure between bedbug lineages, we included 29 samples from bat-associated populations and 48 samples from human-associated populations in the Czech Republic (Hanušovice, CZ, 2016; Hoštějn, CZ, 2016; Raškov, CZ, 2016). The samples were preserved in 96% ethanol and stored in a refrigerator prior to DNA extraction. To investigate the effects of blood meals on gene expression in bedbugs associated with bats and humans, we isolated RNA from the guts of bedbugs for transcriptome sequencing. This

approach is based on the fact that the bedbug's gut is essential for digesting blood meals (Rost-Roszkowska et al., 2017). We analysed 32 biological replicates across four treatments: eight samples from the human lineage of bedbugs fed on their original diet (human blood), eight samples from the human lineage after six generations of feeding on bat blood, eight samples from the bat lineage of *C. lectularius* sourced from original bat blood, and eight samples from bat lineage bedbugs fed with human blood for six generations (Table S1). In order to standardize for sex, life stage and feeding status, only adult, virgin females were used, 48 hours after they were fed as adults for the first time, 2-3 weeks after the last larval feeding. For populations with diet altered from the original host association, we sampled individuals from the 6th generation after the blood switch. All samples were euthanized and immediately stored at -80 degrees Celsius to preserve RNA integrity.

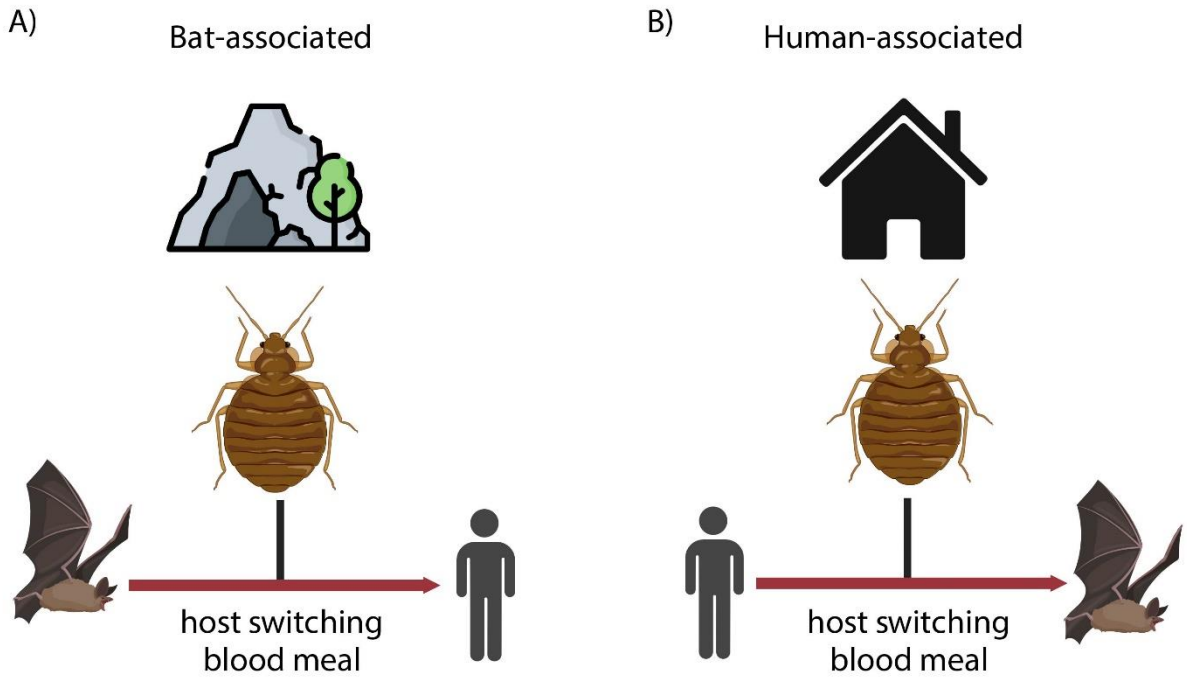


Figure 1: Experimental design for the transcriptome study on host switching in *C. lectularius*. Panel A illustrates the transition of bat-associated bedbugs to a human blood meal, Panel B shows human-associated bedbugs switching to bat blood. This in vitro experiment investigates the transcriptomic responses in two host-associated lineages when exposed to the opposite host's blood.

DNA extraction and ddRADseq library preparation

Genomic DNA was extracted from the specimens using the DNeasy Blood and Tissue Kit (Qiagen). The extracted DNA's quality and quantity were assessed using a 0.8% agarose gel and a Qubit 2.0 Fluorometer. Then, two ddRAD libraries were prepared according to the protocols modified from Peterson et al. (2012) and Schweyen et al. (2014). The genomic DNA was digested at 37°C for two hours with EcoRI (Thermo Scientific) and NlaIII (NEB) restriction enzymes, and the completeness of digestion was confirmed on a 1.5% agarose gel. Adapters that included random spacers (as per Schweyen et al., 2014) were attached to the DNA in a 30 µl reaction mix that included specific adapter concentrations, T4 ligase buffer, T4 DNA ligase, ddH2O, and 75 - 300 ng of digested DNA per sample. This mixture was incubated at 22°C for 60 minutes, with a subsequent heat inactivation step at 65°C for 10 minutes. Afterward, 48 to 72 uniquely barcoded samples were pooled into a single library, vortexed, and equally distributed into strip tubes. The library preparation included two stages of magnetic bead purification. Initially, all samples were diluted with 10 µl of ddH2O and combined into two tubes for incubation on a magnetic plate. DNA was then eluted in 16 µl of ddH2O, and the eluates from both tubes were combined for a final incubation on the magnetic plate. DNA quantity was again verified using the Qubit 2.0 Fluorometer. Size selection was performed on DNA fragments ranging from 383-449 bp using the Pippin Prep system (Sage Science). PCR amplification involved 6.25 ng of size-selected DNA, P5 forward and P7 reverse primers, 5X-HF buffer, dNTPs, ddH2O, and Phusion polymerase, undergoing 12 cycles. The quality and concentration of the final libraries were evaluated using agarose gel electrophoresis and the Qubit 2.0 Fluorometer. The libraries were sequenced on an Illumina NovaSeq platform with 150 PE reads(Novogene, UK).

Transcriptome sequencing

To analyze the transcriptome profiles of BL and HL bedbug samples, RNA was extracted from all 32 specimens. using the acid guanidinium thiocyanate-phenol-chloroform extraction method, as described by Chomczynski and Sacchi (1987) with reagents supplied by Invitrogen (Carlsbad, CA, USA). The RNA yield was quantified using the Qubit RNA Broad Range Assay Kit from Thermo Fisher, and RNA integrity was verified using an Agilent Bioanalyzer 2100 (Agilent Technologies, USA). After quality assessment, the RNA samples were processed commercially to create cDNA libraries, which were then sequenced on the Illumina NovaSeq 150 PE read platform (Novogene, UK).

ddRAD data processing and reference assembly

Demultiplexing, adapter sequence trimming, low quality read filtering, and PCR duplicate removal were conducted using the "preprocess_ddradtags.pl" script (Schweyen et al., 2014). We then evaluated the initial quality of our sequencing data using FastQC (Andrews, 2010) and MultiQC (Ewels et al., 2016). All paired-end reads were aligned to the bed bug reference genome using Bowtie v2.5.0 (Langmead and Salzberg, 2012) and alignments were processed to the BAM format with Samtools v1.18 (Danecek et al., 2021). Using the Stacks software (Rochette et al., 2019), we performed reference assembly for genotype calling and locus assembly, applying stringent criteria to ensure accurate data analysis. This included the use of loci present across all populations and in at least 80% of individuals within each population, a minor allele frequency threshold of 3%, and a maximum heterozygosity limit of 80%. We excluded SNPs in linkage disequilibrium using Stacks' population program and refined our dataset with vcftools v0.1.16 (Danecek et al., 2011) to remove variants with extreme coverage depths and significant missingness. We produced two SNP matrices in VCF format: dataset A, encompassing 49,908 SNPs across 78 samples from both BL and HL bedbug populations with an average locus coverage of 21x; and dataset B, containing 10,288 SNPs post-LD filtering with an average coverage of 18x.

Genome-wide diversity

We used the R package SambaR (de Jong et al., 2021) to compute several metrics of genetic diversity - including nucleotide diversity (Pi), expected heterozygosity (ExeHe), observed heterozygosity (ObsHe), Multi-locus heterozygosity (MHL), standardised Multi locus heterozygosity (sMLH), Watterson's estimator theta (θ), Tajima ‘s D, and private alleles (K), relatedness coefficient (r), and inbreeding coefficient (F) utilising the dataset A.

Population Genetic Structure

To explore how host specificity affects the genetic structure of bedbugs, we implemented three genetic clustering techniques on an LD-filtered dataset B. First, we conducted a principal component analysis (PCA) and a discriminant analysis of principal components (DAPC) using Plink V 1.90 (Purcell et al., 2007) and the adegenet package (Jombart and Collins, 2015) in R v4.0.5, respectively. We identified the optimal cluster numbers using the K-means method,

selecting configurations with the lowest Bayesian Information Criterion (BIC). After assigning individuals to clusters using DAPC and the optim.a.score method, we visualized the clusters with PCA plots. For our second method, we used ADMIXTURE, enhanced with AdmixPiPe's parallel processing capabilities (Mussmann et al., 2020), to perform evolutionary clustering. We analyzed 20 replicates for each K value from 1 to 4, choosing the best K based on the lowest cross-validation error (Figure S1), and displayed these clusters on the CLUMPAK server (Kopelman et al., 2015). Lastly, we examined the population structure further using fineRADstructure and RADpainter v.0.2 (Malinsky et al., 2018), adjusting for allele and SNP distribution and capping individual missing data at 25% to prepare data for RADpainter format. Analysis was run using a burn-in period of 200,000 iterations and a total of 1,000,000 iterations sampled at every 1000..

Gene expression analysis in bedbugs across different blood meals

A total of 36 RNA-seq samples from BL and HL *C. lectularius* were analyzed to compare transcriptomic differences related to varying blood meals. Low-quality reads and Illumina remaining adapters were removed from all samples using Trimmomatic v0.33 (Bolger et al., 2014), employing a quality threshold of 15, a sliding window of 4:20. Subsequent to trimming, paired-end reads were aligned to the reference genome using Hisat2 (Kim et al., 2019). Reads with a mapping quality below 20 or those that mapped to multiple loci were discarded using Samtools v1.10. The resulting SAM files were sorted, converted to binary format (BAM) , and genomic features were quantified using featureCounts v1.06 from the Subread package (Liao et al., 2014). Table S2 details the mapped reads per sample replicate and gene model. To discover gene plasticity related to host switching blood meals, to identify gene plasticity, We used DESeq2 (Love et al., 2014) implemented in iDEP V 2.01 (Ge et al., 2018) to analyze differentially expressed genes (DEGs) between bed bug populations on bat and human lineages, as well as those that exposed to their non-original blood. Initially, raw counts were normalized using a log2 transformation of count per million (CPM+C), ensuring that only genes with at least 0.5 counts per million were included. DEGs were stringently identified based on a p-value threshold of less than 0.05 and an absolute log-fold change exceeding 1.

Pathway analysis

Gene Set Enrichment Analysis (GSEA) (Shi and Walker, 2007) implemented in iDEP, applying a significant false discovery rate (FDR) of 0.05, was used to identify overrepresented and underrepresented gene sets associated with DEGs between bat and human blood meals, as well as those resulting from switched blood meals and their interactions.

Results

Genome wide Diversity

Table 1 shows the results of ddRAD analysis on genetic diversity in *C. lectularius* across BL and HL. Nucleotide diversity (Pi) is higher in the bat lineage (0.185) compared to the human lineage (0.122). Measures of Multi-Locus Heterozygosity (MLH) and its standardized form (sMLH) also indicate slightly greater genetic diversity in bats, with values of 0.104 and 1.366, respectively, versus 0.099 and 0.790 in humans. Waterson's Theta (θ), which reflects effective population size, is nearly double in bats at 0.213, compared to 0.112 in humans. Tajima's D, used to assess the neutrality of mutations, shows values close to zero in both lineages, -0.025 for bats and 0.01 for humans, suggesting a neutral evolutionary pattern with no strong deviations from neutrality due to selection or demographic changes (Table 1). The Nei genetic distance and Fst calculations indicated a 3% genetic divergence and a value of 0.16 between BL and HL, respectively.

Table 1: Genetic diversity in bedbug populations associated with bat (BL) and human hosts (HL). Abbreviations: number of samples (N), nucleotide diversity (Pi), expected heterozygosity under Hardy-Weinberg equilibrium (ExeHe), observed heterozygosity (ObsHe), multilocus heterozygosity (MLH), standardized MLH (sMLH), Watterson's estimator theta (θ), Tajima's D value for each population, private allelels (K), relatedness coefficient (r) and inbreeding coefficient (F).

Parasite population	N	Pi	ExeHe	ObsHe	MLH	sMLH	θ	Tajima D	K	r	F
BL	29	0.185	0.195	0.094	0.104	1.366	0.213	-0.02	21,1	0.41	0.45
HL	78	0.122	0.131	0.054	0.099	0.790	0.112	0.01	4,57	0.44	0.55

Population genetic structure

The results of non-model-based genetic structure analyses (PCA and DAPC) distinctly identified two genetic clusters between HL and BL *C. lectularius*. PCA separated these two parasite populations along its PC1 axis, explaining 12.02% of the variance (Figure 2A and B).

Consistently, model-based genetic clustering confirmed two distinct genetic structures, similar to those revealed by PCA and DAPC, with some level of admixture observed among some individuals (Figure 2C).

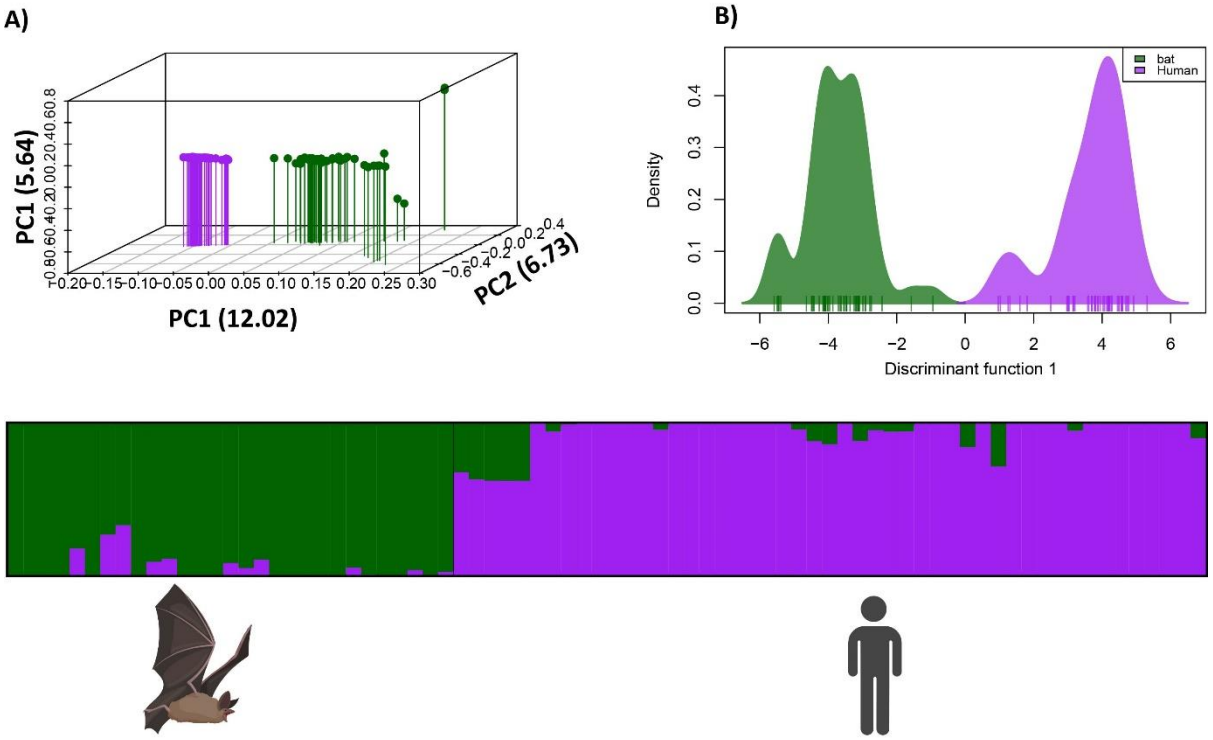


Figure 2: Analysis of genetic structure in bat and human-associated bedbug populations. A) Principal Component Analysis (PCA) showing the genetic variation along the first three principal components (PC1, PC2, and PC3) with percentages of variance explained in parentheses. Samples from bat hosts are shown in purple, and those from human hosts are in green. B) Discriminant Analysis of Principal Components (DAPC) illustrating the density distribution of the genetic data along discriminant function 1 for bat (green) and human (purple) host-associated populations. C) Admixture analysis displaying the proportion of genome admixture in populations associated with bats and humans, depicted in corresponding colors.

The results from fineRADstructure aligned with those obtained from PCA and admixture analyses, confirming the presence of two genetic clusters in *C. lectularius*, associated with human and bat populations. The topology derived from this analysis indicates these clusters with strong posterior probabilities. Each cluster shows a high level of shared ancestry within

its respective population, while exhibiting relatively less shared ancestry between the two lineages (Figure 3).

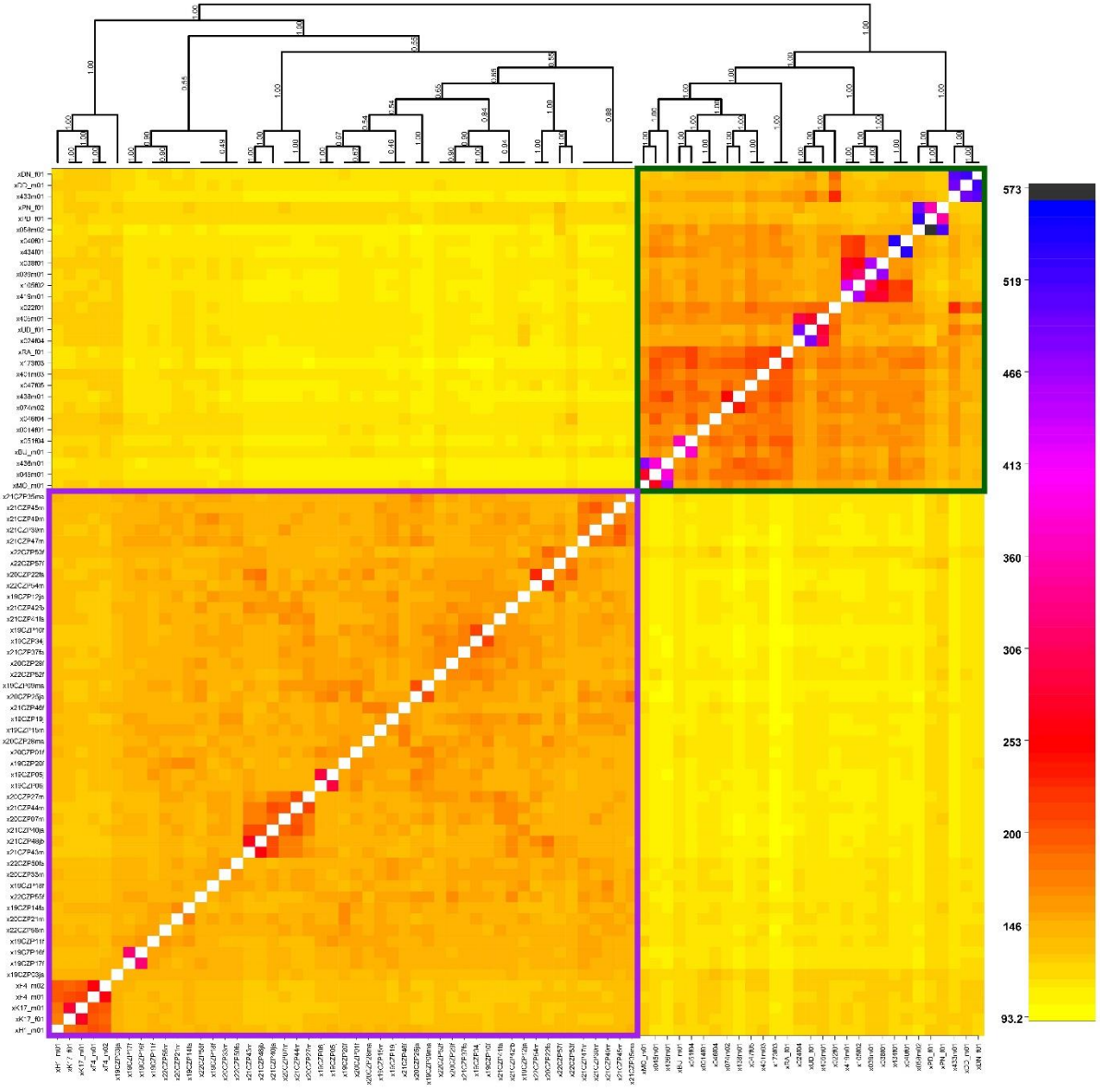


Figure 3: Coancestry heatmap from FineRADstructure analysis of bedbug populations. The heatmap represents a coancestry matrix, illustrating pairwise relatedness between individual bedbugs ranging from light yellow (low coancestry) to deep red (high coancestry).The dendrogram above the matrix categorizes individuals based on genetic similarity, highlighting potential substructures within the population.

Gene expression analysis of BL and HL bedbug populations based on the original blood meal

The DEG analysis identified two distinct transcriptome profiles between BL and HL bedbugs, revealing 146 DEGs: 67 genes were upregulated and 79 were downregulated (Figure 4A and B). Moreover, the results of pathway analysis revealed a distinction in gene expression between bedbugs fed on bat versus human blood across three key categories: molecular functions, biological processes, and cellular components. In molecular functions, genes involved in molecular adaptor activity, oxidoreduction-driven active transmembrane transporter activity, and NADH dehydrogenase complex assembly were significantly underrepresented. Conversely, enhancements were observed in activities of structural molecules, various enzymes including aspartic-type and serine-type endopeptidases, as well as in binding functions related to odorants and Hsp70 proteins. In terms of biological processes, pathways associated with the assembly of mitochondrial respiratory chain complex I and protein complexes within the endoplasmic reticulum were notably underrepresented. Cellular components such as the extracellular region saw an increase in gene expression. These results categorize into three main groups: molecular function, biological process, and cellular component, providing a detailed view of the transcriptomic changes between the two feeding groups (Figure 4).

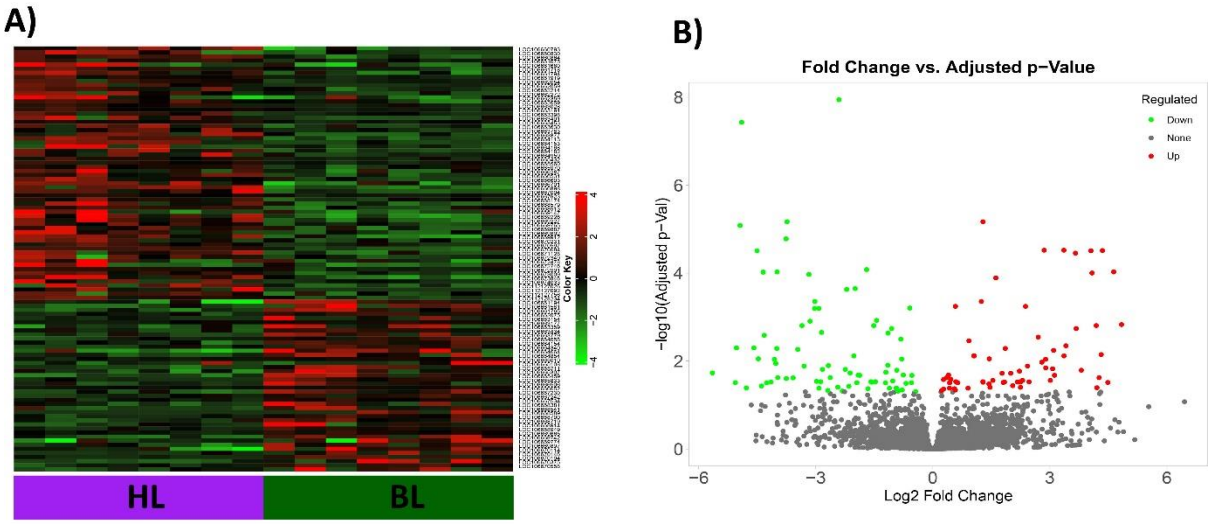


Figure 4: Differential Gene Expression analysis of bedbug lineages from original blood meals. (A) Heatmap of DEGs analysis between BL and HL bedbug lineages. A total of 146 DEGs are displayed, with gene expression levels color-coded from high (red) to low (green). (B) Volcano plot illustrating the log2 fold change versus the adjusted p-value for the DEGs identified. In total, 67 genes were found to be upregulated (red) and 79 downregulated (green) in the comparison between the two lineages.

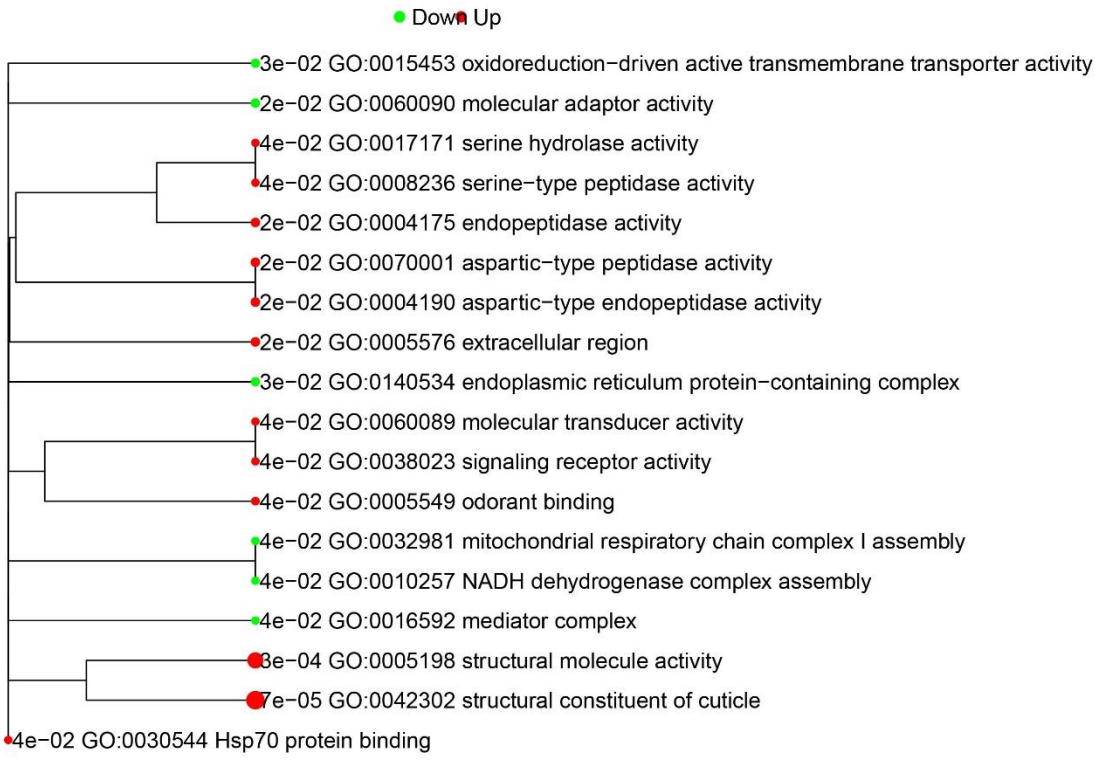


Figure 4 The pathway analysis tree of DEGs associated with the BL and HL bedbug lineages raised on their original blood meals.

Gene expression variability in bedbugs during host switching blood meals

We identified 61 DEGs from 16 samples of HL bedbugs, originally feeding on human blood, which were switched to bat blood (Figure 1A). The DEGs comprised eight upregulated and 53 downregulated genes. Subsequent enrichment analysis highlighted significant differences in various biological pathways, categorized into cellular components, molecular functions, and biological processes. In terms of cellular components, There is a noted downregulation from the original blood type to non-original blood types in recycling endosome and glyoxylate and dicarboxylate metabolism, illustrating a reduction in these functional activities within the human lineage. Conversely, significant upregulation is observed in several categories, particularly in ligand-gated ion channel activities such as gated channel activity, extracellular ligand-gated ion channel activity, and ligand-gated ion channel activity, indicating enhanced ion transport and receptor functions. This upregulation extends to structural enhancements in the cuticle, emphasizing physical structural differences. Additionally, various receptor activities are significantly enhanced, including neurotransmitter receptor activity, neuropeptide Y receptor activity, glutamate receptor activities, and odorant binding, supporting a divergence in sensory and signalling pathways. Further upregulation is seen in synaptic signalling and the

neuropeptide signalling pathway, which highlight differences in synaptic and neuropeptide-mediated communication between the blood meals (Figure 5).

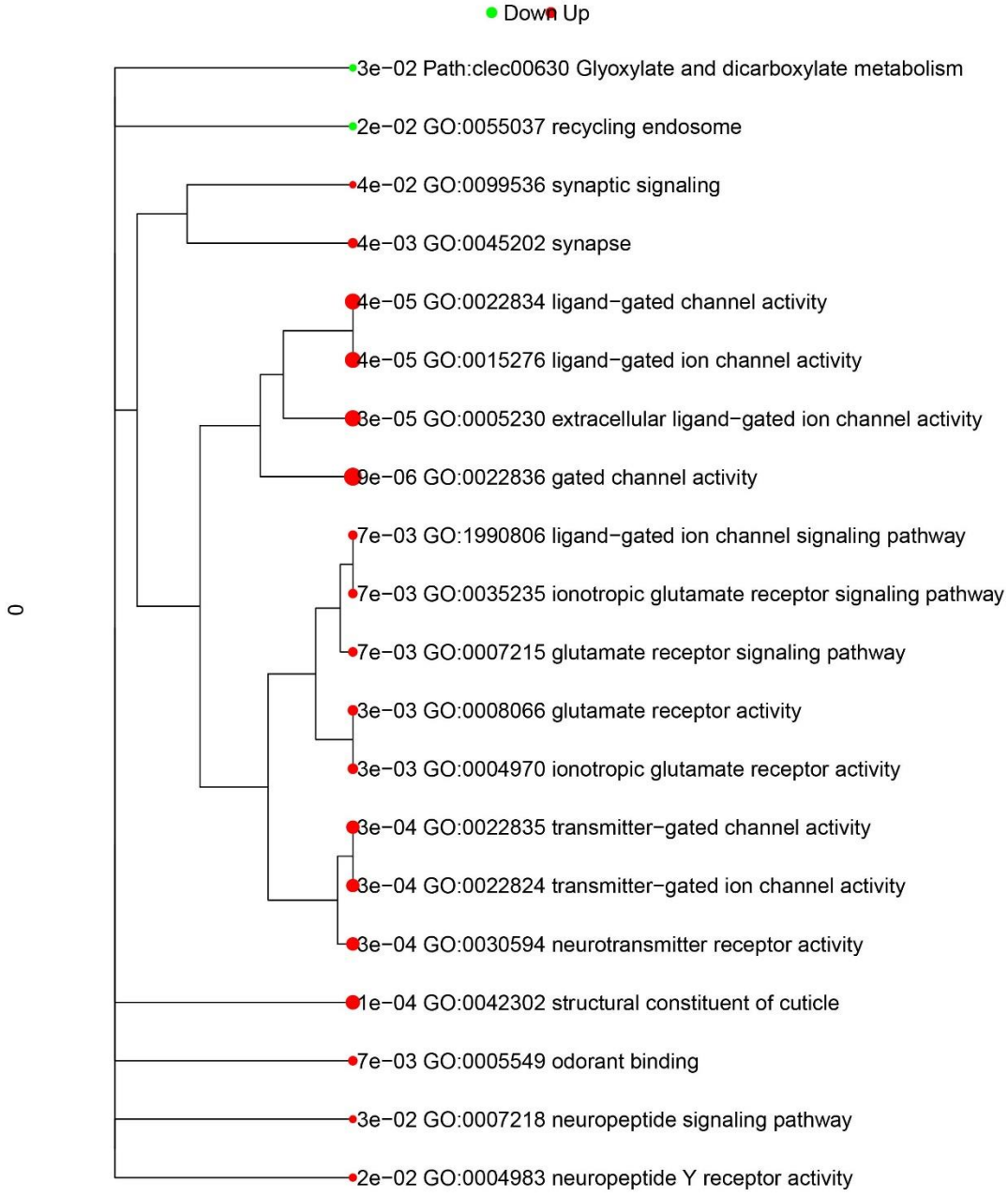


Figure 5: hierarchical clustering tree illustrating the correlation among significant pathways in HL bedbugs after diet switch from human to bat blood.

The result of gene expression analysis for cross feeding experiment uncovered 1663 DEGs from 16 samples of BL lineages, originally feeding on bat blood and diet switched to human blood (Figure 1B). These include 860 upregulated and 803 downregulated genes. Moreover,

the pathway analysis revealed a significant downregulation in the proton-transporting ATP synthase complex, integral and intrinsic components of the endoplasmic reticulum membrane, and DNA-templated transcription elongation. Additionally, the pathway for oxidative phosphorylation and mitochondrial electron transport from NADH to ubiquinone also showed significant downregulation. Conversely, there was a significant upregulation in pathways related to DNA and protein complexes, such as protein-DNA complex and DNA packaging complex. Enhancements were also noted in phospholipid metabolism processes, including phospholipid dephosphorylation and phosphatidylinositol dephosphorylation, as well as increased activity in odorant binding and various receptor activities including transmembrane signalling receptor activity. These results suggest intricate shifts in cellular function and structural organization between bat and human blood-fed bedbugs, reflecting potentially adaptive changes in response to different host environments (Figure 6).

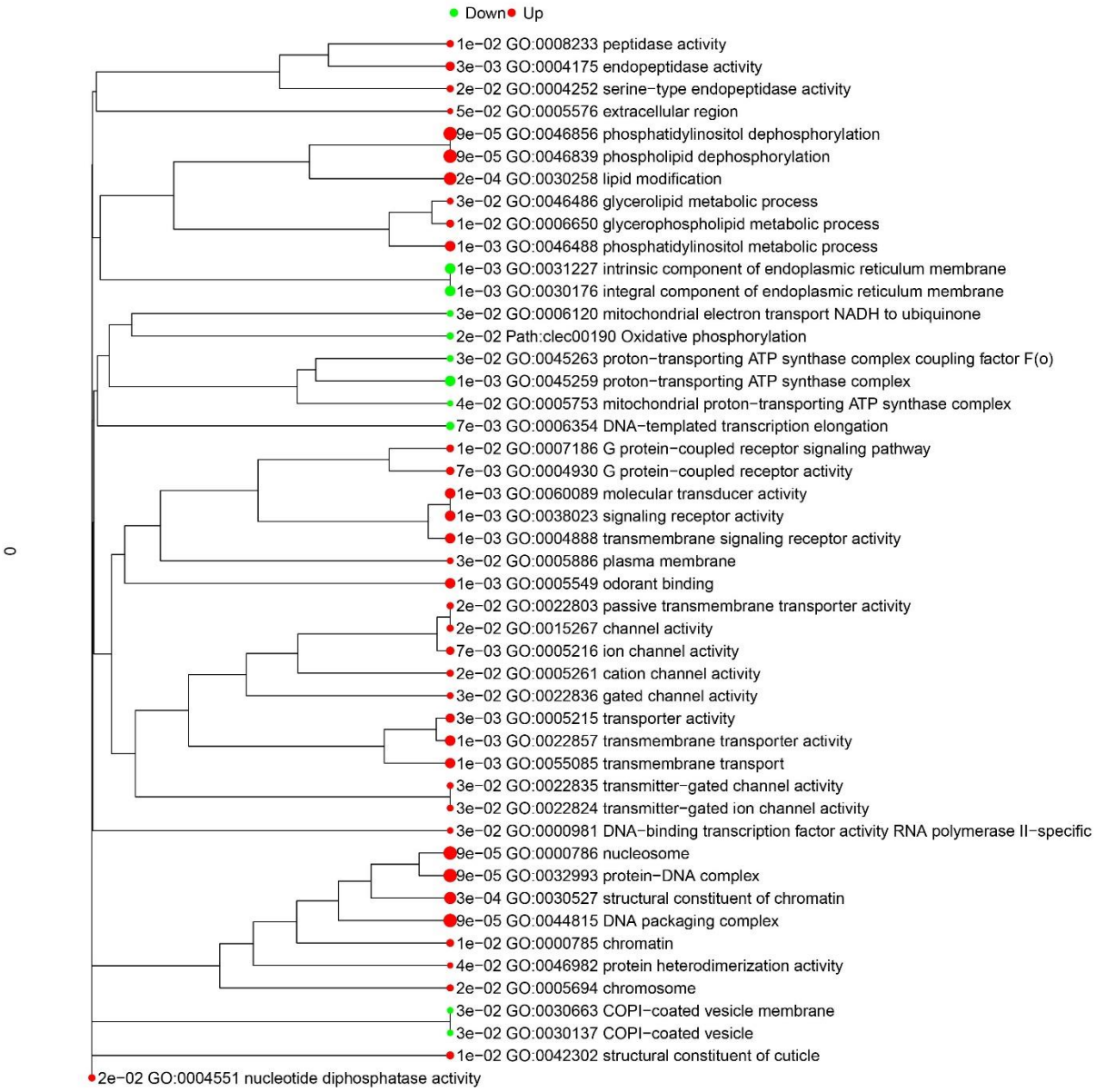


Figure 6: Significantly regulated pathways in BL *C. lectularius* exposed to human blood.

To identify DEGs demonstrating gene plasticity in response to different blood meals, we analysed the intersection of DEGs from human-associated bedbugs fed with bat blood and bat-associated bedbugs fed with human blood. The analysis revealed 1632 unique DEGs in the BL group when switching from bat to human blood, and 29 unique DEGs in the HL group when the switch was from human to bat blood. Additionally, 32 DEGs (all down regulated genes from original blood to opposite blood meal) were found to be common between both groups. The result of gene ontology revealed several key biological processes and molecular functions,

crucial to the adaptation and response mechanisms of bedbugs to varying blood sources. The biological processes uncovered genes related to "protein phosphorylation" and "fatty acid biosynthetic process", highlighting shifts in signal transduction and lipid metabolism (Figure S2). In terms of cellular components, changes were evident in the "plasma membrane" and "membrane" related genes, reflecting possible alterations in cell structure and transport mechanisms to accommodate the metabolic demands imposed by the different blood meals (Figure S3). On the molecular function front, activities such as "protein kinase activity", "heme binding", "phosphopantetheine binding", and "ubiquitin-protein transferase activity" were prevalent, indicating modifications in protein regulation and degradation pathways. Furthermore, "ATP binding" and "calcium ion binding" suggest changes in energy usage and signalling pathways that could be essential for survival and adaptation (Figure S4).

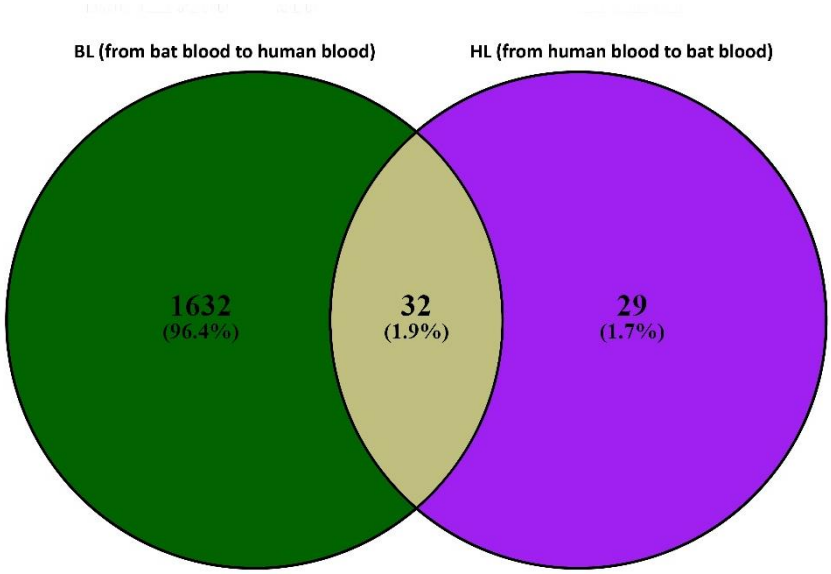


Figure 7: Venn diagram demonstrating the common DEGs identified between BL and HL bedbug after switching from their native blood meals.

Discussion

Host switching through ecological fitting allows parasites to exploit new host species using their existing traits, without significant evolutionary changes. This adaptability not only demonstrates their ability to tap into new ecological opportunities but could also lead to speciation as they utilize diverse resources (Matsubayashi and Yamaguchi, 2022). In this study, we used *C. lectularius* as a model organism, examining two lineages, one parasitizing humans and one parasitising bats. Our research, integrating genome-wide SNPs and transcriptome data, robustly confirms the genetic divergence between the two host-associated lineages of *C. lectularius*. These findings are consistent with previous microsatellite and morphological studies, whereas mitochondrial studies were not informative enough due to incomplete lineage sorting (shared haplotypes) between the lineages (Balvín et al., 2012; Booth et al., 2015; Sasínková et al., 2023). Finally, we employed transcriptome data to contrast the different blood meals between BL and HL, investigating whether any biological functions could potentially lead to parasite adaptation between these lineages. Below, we discuss possible explanations for the observed genetic and transcriptome patterns in relation to parasite adaptation to different hosts.

Host influence on genetic diversity of bedbugs

The results of the genome-wide SNP analysis revealed greater diversity in BL bedbugs compared to the HL, in agreement with findings from an earlier microsatellite study (Booth et al., 2015). This diversity can be attributed to ecological and evolutionary factors between the lineages. For instance, bats, being highly mobile hosts, enable broader gene flow among bedbug populations through their extensive flight range and utilization of multiple roost sites (Willis and Brigham, 2004). Therefore, such connectivity can promote frequent interbreeding across different populations, thereby enhancing genetic diversity. Additionally, the diverse and changing natural habitats of bats expose bedbugs to varied environmental conditions and selective pressures, fostering the genetic variation necessary for adaptation. In contrast, human environments are generally more stable and isolated, potentially leading to uniform selection pressures and genetic bottlenecks (Robison et al., 2015; Talbot et al., 2017). Moreover, the evolutionary relationship between bedbugs and bats is likely more ancient, providing more time for coevolution and diversification, whereas the relatively recent host-switch to humans reduced diversity in HL. Therefore, this complex interaction of host behaviour, environmental

variability, and evolutionary history might underpin the observed differences in genetic diversity between bedbug lineages associated with bats and humans.

The results of our study further confirm that bat-associated bedbug populations not only exhibit higher genetic diversity but also lower levels of inbreeding and more private alleles, suggesting less genetic drift and more stable population dynamics (Table 1). Conversely, human-associated populations, often subjected to intense pest control measures, display signs of recent bottlenecks and founder effects, which contribute to their reduced genetic diversity (Booth et al., 2015). Additionally, our results uncovered relatively low relatedness in bat-associated bedbug populations compared to those associated with humans. A low relatedness coefficient signifies greater genetic diversity, which is typical of larger populations with more genetic mixing (Rousset, 2002). This diversity can confer advantages such as greater resilience to diseases and environmental changes due to the presence of a wider variety of genetic traits (Reda et al., 2023). These findings likely highlight the significant impact of ecological niches and human intervention on genetic structure and adaptability in bedbug populations, emphasizing the need for integrated pest management strategies that consider genetic factors to effectively control these pests.

Transcriptomic analysis of host adaptation in BL and HL bedbugs

The results of the DEGs analysis clearly distinguish the transcriptome profiles of BL and HL bedbugs from each other. Moreover, the pathway analysis uncovered several hydrolase activities (hydrolyzing O-glycosyl compounds, acting on glycosyl bonds, aspartic-type endopeptidase activity, serine-type peptidase activity) associated with four DEGs: LOC106664154, LOC106665010, LOC106667230, and LOC106672237. All these transcripts were overrepresented in the HL bedbugs compared to the BL bedbugs. Blood meals from different hosts contain various proteins or lipids that need to be digested efficiently (Lam et al., 2022; Li et al., 2021). Moreover, the lipid composition was shown to differ, with the proportion of saturated fatty acids being higher in bat blood (53 %) than in human blood (42%) (Kremenova PhD thesis). Therefore, hydrolase activities, particularly those involving the hydrolysis of peptide bonds (peptidase activities), are critical for bedbugs as they feed on blood from different hosts. Aspartic-type endopeptidase and aspartic-type peptidase activities are essential because they break down large blood proteins like hemoglobin into smaller peptides and amino acids, which are easier for hematophagous parasites to absorb and utilize (Vorel et al., 2021). Moreover, endopeptidase activity, which cleaves peptide bonds within proteins, is

crucial for the initial breakdown of large blood proteins into smaller fragments, facilitating further digestion and nutrient acquisition (Santiago et al., 2017). Serine-type peptidase and serine hydrolase activities play a significant role in the final stages of protein digestion, breaking down peptides into amino acids that bedbugs can readily absorb (Borges-Veloso et al., 2012; Francischetti et al., 2010). These enzymatic activities enable bedbugs to adapt to the varying protein compositions in bat and human blood, ensuring efficient processing and nutrient absorption, which is vital for their survival, host-switching ability, and reproductive success.

Wawrocka and Bartonička (2013) revealed through cross-feeding experiments that bat-associated bedbugs exhibit higher survival and moulting rates when fed on bat blood compared to human blood. Conversely, when both lineages were fed on human blood, no significant disadvantages were observed for BL bedbugs compared to HL, which are typically associated with human hosts. This suggests that BL bedbugs do not experience significant disadvantages when fed on human blood, despite their usual association with bats. In our study, a notable increase of 1663 DEGs was observed in BL bedbugs when they switched from bat to human blood. Conversely, only 61 DEGs were identified in HL bedbugs when they switched from human to bat blood. This stark contrast may suggest that HL bedbugs could be genetically constrained, possibly due to lower heterozygosity, leading to reduced genetic plasticity. Meanwhile, the high number of DEGs in BL bedbugs could indicate their capacity for more extensive regulatory responses when their diet is altered, hinting at a substantial genetic adaptation and physiological stress when transitioning to human blood. Thus, these findings highlight significant differences in genetic responsiveness and adaptability between the bedbug lineages, offering important insights into host specialization and evolutionary dynamics in parasitic species

The results of pathway analysis indicated that odorant binding is significantly overrepresented in HL bedbugs compared to BL bedbugs. Additionally, this molecular function was also substantial in our cross-feeding experiments, where the BL lineage switched from bat to human blood and the HL lineage switched from human to bat blood. Human scents are known to attract bedbugs and can direct their movements in olfactometers (Hentley et al., 2017). Moreover, different human odorants caused varying neuronal responses with different firing frequencies and temporal dynamics, indicating a complex interaction between bedbugs' sensory perception and the chemical makeup of human odours (Liu et al., 2017) . Liu and Liu (2015) identified

two odorant receptors that were functionally characterized, showing specific responses to aldehyde human odorants. These findings suggest that HL bedbugs may possess a more adaptable or extensive array of odorant receptors that enable them to effectively process and respond to a wider variety of host-derived chemical cues. This capability could confer an evolutionary advantage in environments where host variability is high. In contrast, BL bedbugs might exhibit a more restricted odorant receptor profile, potentially limiting their ability to adapt to different hosts as efficiently as HL bedbugs. Therefore, this differential expression of odorant binding capabilities demonstrates the potential for divergent evolutionary paths in bedbug populations depending on their typical host exposure.

The results of the present study identified 32 common DEGs in our cross-feeding experiments, which compared the transcriptome profiles of the BL lineage switching from bat to human blood and the HL lineage switching from human to bat blood. These DEGs might contribute to phenotypic plasticity in BL and HL bedbugs in response to changes in the blood type. Among them, several DEGs are involved in critical processes such as protein phosphorylation, or various hydrolase activities, including the hydrolysis of glycosyl bonds and peptidase activities. These processes are essential for adapting to different blood sources by modulating digestive enzyme activity and nutrient absorption efficiency (Wang et al., 2021). Additionally, DEGs associated with oxidoreductase activity and response to oxidative stress play a role in managing the oxidative challenges posed by different blood types (Pawłowska et al., 2023). Furthermore, transporters and ion channels identified among the DEGs facilitate the efficient uptake of nutrients and maintenance of cellular homeostasis (Di Maggio et al., 2016). DEGs related to heme binding are crucial for detoxifying free heme released during blood digestion, thereby preventing oxidative damage and supporting the safe utilization of heme as a nutrient (Toh et al., 2010). Therefore, these DEGs enable bedbugs to effectively digest and utilize nutrients from both bat and human blood, highlighting the molecular basis for their host-switching capabilities and adaptability.

In conclusion, the results of our study further confirmed that bat-associated bedbug populations not only exhibit higher genetic diversity but also lower levels of inbreeding and more private alleles, indicating less genetic drift and more stable population dynamics. Conversely, human-associated populations, often subject to intense pest control measures, display signs of recent bottlenecks and founder effects, contributing to their reduced genetic diversity. Moreover, the analysis of DEGs highlighted significant differences in gene expression linked to host adaptation. Bat-associated bedbugs showed extensive gene regulation when switching from bat to human blood, reflecting their adaptability and potential evolutionary pressures from their primary hosts. Conversely, the minimal changes in DEGs in human-associated bedbugs suggest a reduced genetic plasticity, possibly due to more consistent environmental pressures. These findings emphasize the profound impact of host association on the genetic diversity and adaptability of parasites, underscoring the complex interplay between genetic makeup and environmental challenges

Supplemental information

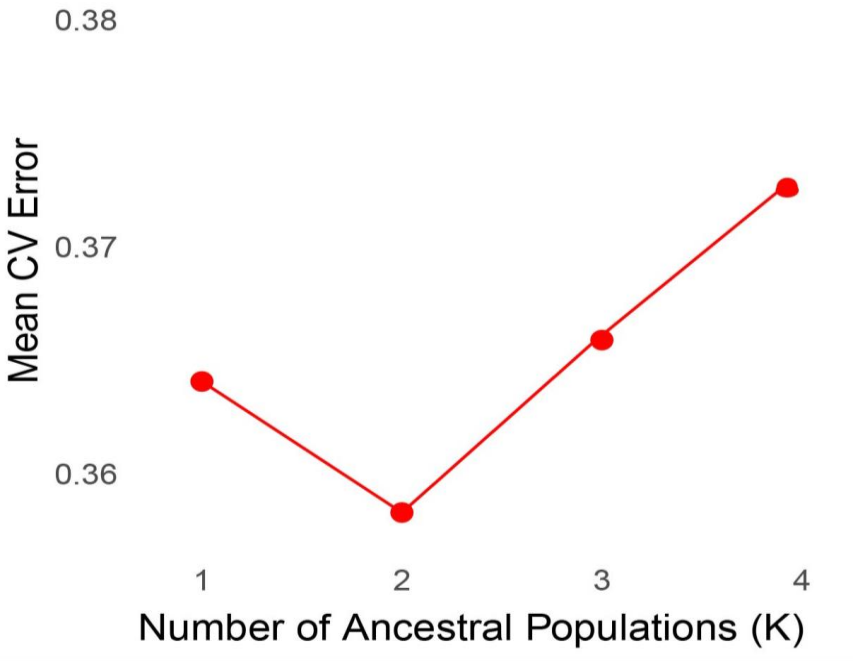


Figure S1. Cross-validation error from DAPC analysis revealing two genetic clusters in *Cimex lectularius* associated with bats and humans.

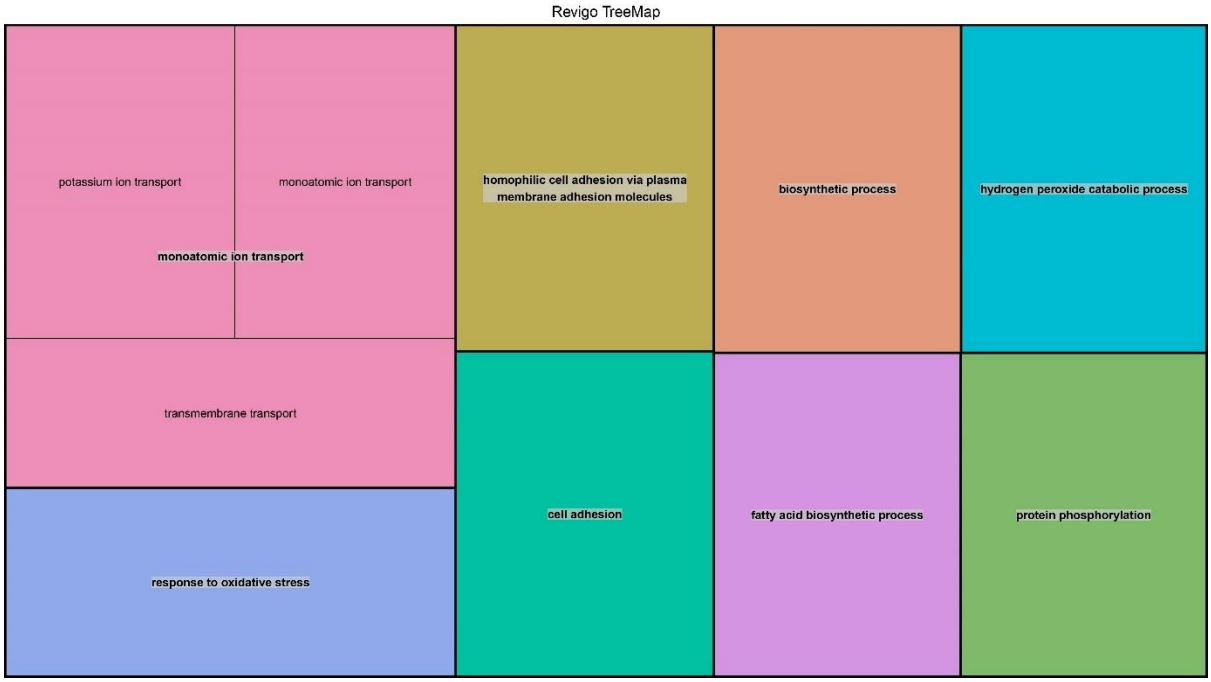


Figure S2. Treemap visualization of enriched biological processes among 32 DEGs in *C. lectularius*. These genes represent common DEGs between bat-associated *C. lectularius* switching from their original bat blood meal to a non-original human blood meal, and human-associated *C. lectularius* switching from their original human blood meal to a non-original bat blood meal. The biological processes are categorized based on their Gene Ontology (GO) terms, with terms grouped by their representative functions, including catalase activity, protein binding, microtubule binding, symporter activity, oxidoreductase activity, transferase activity, and signalling receptor activity.

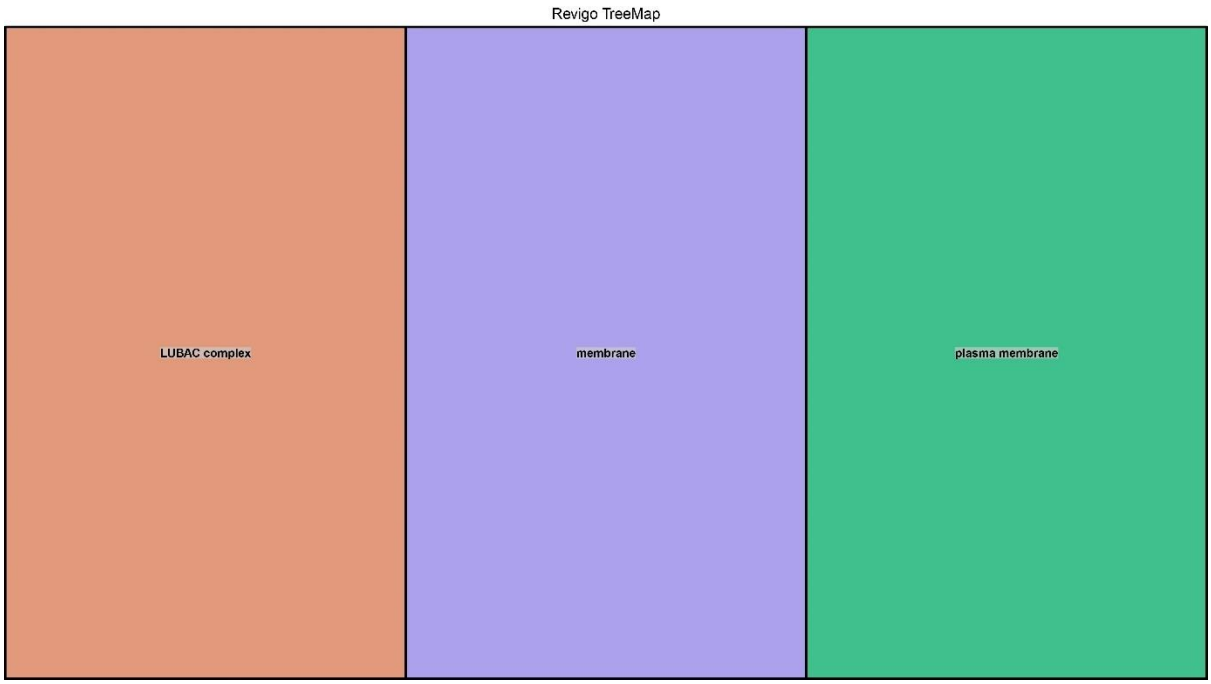


Figure S3. Treemap visualization depicting enriched cellular components among DEGs in *C. lectularius*. These DEGs are associated with the switch from their native blood meal source to an alternative source in both bat-associated and human-associated *C. lectularius*. The cellular components are categorized by their Gene GO terms, emphasizing key elements such as the plasma membrane, membrane, and LUBAC complex. The terms are grouped by their representative functions, offering an in-depth view of the cellular components implicated in the dietary transition.

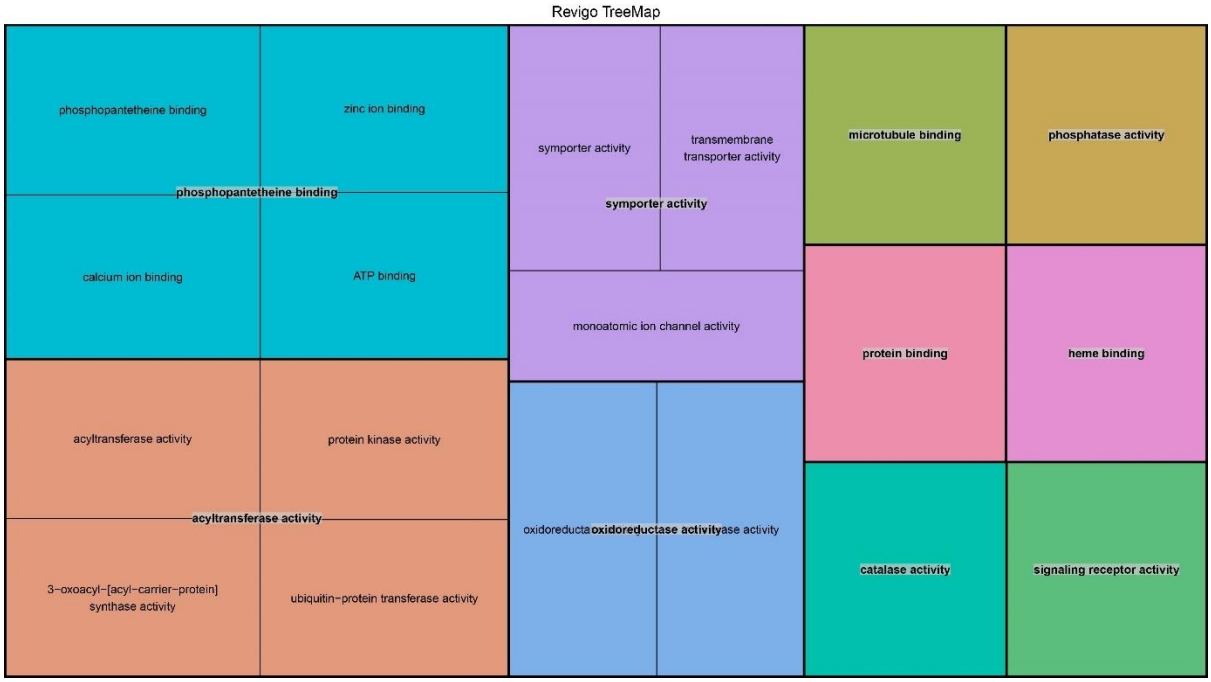


Figure S4. Treemap visualization demonstrating enriched molecular functions among DEGs in *C. lectularius*. These DEGs are involved in the transition from their original blood meal source to an alternative source in both bat-associated and human-associated *C. lectularius*. Molecular functions are categorized by their GO terms, with key functions including catalase activity, protein binding, microtubule binding, symporter activity, monoatomic ion channel activity, transmembrane transporter activity, oxidoreductase activity, transferase activity, acyltransferase activity, protein kinase activity, ubiquitin-protein transferase activity, phosphatase activity, heme binding, phosphopantetheine binding, calcium ion binding, ATP binding, zinc ion binding, and signaling receptor activity. The terms are grouped by their representative functions, providing an in-depth look at the molecular activities associated with the dietary switch.

References

Agosta, S.J., Brooks, D.R., 2020. The Stockholm Paradigm, in: The Major Metaphors of Evolution: Darwinism Then and Now. Springer, pp. 219–242. https://doi.org/10.1007/978-3-030-52086-1_10

Amwoma, J.G., Kituyi, S., Wakoli, D.M., Ochora, D.O., Chemwor, G., Maisiba, R., Okore, W., Opot, B., Juma, D., Muok, E.M.O., Garges, E.C., Egbo, T.E., Nyabuga, F.N., Andagalu, B., Akala, H.M., 2024. Comparative analysis of peripheral whole blood transcriptome from asymptomatic carriers reveals upregulation of subsets of surface proteins implicated in Plasmodium falciparum phenotypic plasticity. Biochem. Biophys. Reports 37, 101596. https://doi.org/10.1016/j.bbrep.2023.101596

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Araujo, S.B.L., Braga, M.P., Brooks, D.R., Agosta, S.J., Hoberg, E.P., von Hartenthal, F.W., Boeger, W.A., 2015. Understanding Host-Switching by Ecological Fitting. PLoS One 10, e0139225. https://doi.org/10.1371/journal.pone.0139225

Balvín, O., Munclinger, P., Kratochvíl, L., Vilímová, J., 2012. Mitochondrial DNA and morphology show independent evolutionary histories of bedbug Cimex lectularius (Heteroptera: Cimicidae) on bats and humans. Parasitol. Res. 111, 457–469. https://doi.org/10.1007/s00436-012-2862-5

Benoit, J.B., Adelman, Z.N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E.C., Szuter, E.M., Hagan, R.W., Gujar, H., Shukla, J.N., Zhu, F., Mohan, M., Nelson, D.R., Rosendale, A.J., Derst, C., Resnik, V., Wernig, S., Menegazzi, P., Wegener, C., Peschel, N., Hendershot, J.M., Blenau, W., Predel, R., Johnston, P.R., Ioannidis, P., Waterhouse, R.M., Nauen, R., Schorn, C., Ott, M.C., Maiwald, F., Johnston, J.S., Gondhalekar, A.D., Scharf, M.E., Peterson, B.F., Raje, K.R., Hottel, B.A., Armisén, D., Crumière, A.J.J., Refki, P.N., Santos, M.E., Sghaier, E., Viala, S., Khila, A., Ahn, S.J., Childers, C., Lee, C.Y., Lin, H., Hughes, D.S.T., Duncan, E.J., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Wheeler, D., Panfilio, K.A., Vargas Jentzsch, I.M., Vargo, E.L., Booth, W., Friedrich, M., Weirauch, M.T., Anderson, M.A.E., Jones, J.W., Mittapalli, O., Zhao, C., Zhou, J.J., Evans, J.D., Attardo, G.M., Robertson, H.M., Zdobnov, E.M., Ribeiro, J.M.C., Gibbs, R.A., Werren, J.H., Palli, S.R., Schal, C., Richards, S., 2016. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat. Commun. 7, 1–10. https://doi.org/10.1038/ncomms10165

Birget, P.L.G., Repton, C., O'Donnell, A.J., Schneider, P., Reece, S.E., 2017. Phenotypic plasticity in reproductive effort: malaria parasites respond to resource availability. Proc. R. Soc. B Biol. Sci. 284, 20171229. https://doi.org/10.1098/rspb.2017.1229

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Booth, W., Balvín, O., Vargo, E.L., Vilímová, J., Schal, C., 2015. Host association drives genetic divergence in the bed bug, Cimex lectularius. Mol. Ecol. 24, 980–992. https://doi.org/10.1111/mec.13086

Borges-Veloso, A., Saboia-Vahia, L., Cuervo, P., Pires, R.C., Britto, C., Fernandes, N., D'Avila-Levy, C.M., De Jesus, J.B., 2012. Proteolytic profiling and comparative analyses

of active trypsin-like serine peptidases in preimaginal stages of Culex quinquefasciatus. Parasit. Vectors 5, 123. https://doi.org/10.1186/1756-3305-5-123

Brooks, D.R., Hoberg, E.P., Boeger, W.A., 2019. The Stockholm Paradigm Climate Change and Emerging Disease. University of Chicago Press. https://doi.org/10.7208/chicago/9780226632582.001.0001

Brooks, D.R., León-Règagnon, V., McLennan, D.A., Zelmer, D., 2006. Ecological fitting as a determinant of the community structure of platyhelminth parasites of anurans. Ecology 87, S76–S85.

Chen, E.-H., Hou, Q.-L., Wei, D.-D., Jiang, H.-B., Wang, J.-J., 2017. Phenotypic plasticity, trade-offs and gene expression changes accompanying dietary restriction and switches in Bactrocera dorsalis (Hendel) (Diptera: Tephritidae). Sci. Rep. 7, 1988. https://doi.org/10.1038/s41598-017-02106-3

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156–159.

Cooper, R., Wang, C., Singh, N., 2015. Mark-Release-Recapture Reveals Extensive Movement of Bed Bugs (Cimex lectularius L.) within and between Apartments. PLoS One 10, e0136462. https://doi.org/10.1371/journal.pone.0136462

Cruz, A.K., Freitas-Castro, F., 2019. Genome and transcriptome analyses of Leishmania spp.: opening Pandora's box. Curr. Opin. Microbiol. 52, 64–69. https://doi.org/10.1016/j.mib.2019.05.004

D'Bastiani, E., Princepe, D., Marquitti, F.M.D., Boeger, W.A., Campião, K.M., Araujo, S.B.L., 2023. Effect of Host-Switching on the Ecological and Evolutionary Patterns of Parasites. Syst. Biol. 72, 912–924. https://doi.org/10.1093/sysbio/syad022

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. Gigascience 10, giab008. https://doi.org/10.1093/gigascience/giab008

De Fine Licht, H.H., 2018. Does pathogen plasticity facilitate host shifts? PLOS Pathog. 14, e1006961. https://doi.org/10.1371/journal.ppat.1006961

de Jong, M.J., de Jong, J.F., Hoelzel, A.R., Janke, A., 2021. SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. Mol. Ecol. Resour. 21, 1369–1379. https://doi.org/https://doi.org/10.1111/1755-0998.13339

DeVries, Z.C., Santangelo, R.G., Booth, W., Lawrence, C.G., Balvín, O., Bartonička, T., Schal, C., 2020. Reproductive compatibility among populations and host-associated lineages of the common bed bug (Cimex lectularius L.). Ecol. Evol. 10, 11090–11099. https://doi.org/10.1002/ece3.6738

Di Maggio, L.S., Tirloni, L., Pinto, A.F.M., Diedrich, J.K., Yates, J.R., Benavides, U., Carmona, C., Da Silva Vaz, I., Berasain, P., 2016. Across intra-mammalian stages of the liver f luke Fasciola hepatica: A proteomic study. Sci. Rep. 6, 1–14. https://doi.org/10.1038/srep32796

Doña, J., Sweet, A.D., Johnson, K.P., Serrano, D., Mironov, S., Jovani, R., 2017. Cophylogenetic analyses reveal extensive host-shift speciation in a highly specialized and host-specific symbiont system. Mol. Phylogenet. Evol. 115, 190–196. https://doi.org/10.1016/j.ympev.2017.08.005

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Favret, C., Voegtlin, D.J., 2004. Speciation by host-switching in pinyon Cinara (Insecta: Hemiptera: Aphididae). Mol. Phylogenet. Evol. 32, 139–151. https://doi.org/10.1016/j.ympev.2003.12.005

Francischetti, I.M.B., Calvo, E., Andersen, J.F., Pham, V.M., Favreau, A.J., Barbian, K.D., Romero, A., Valenzuela, J.G., Ribeiro, J.M.C., 2010. Insight into the Sialome of the Bed Bug, Cimex lectularius. J. Proteome Res. 9, 3820–3831. https://doi.org/10.1021/pr1000169

Ge, S.X., Son, E.W., Yao, R., 2018. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinformatics 19, 534. https://doi.org/10.1186/s12859-018-2486-6

Hentley, W.T., Webster, B., Evison, S.E.F., Siva-Jothy, M.T., 2017. Bed bug aggregation on dirty laundry: A mechanism for passive dispersal. Sci. Rep. 7, 1–5. https://doi.org/10.1038/s41598-017-11850-5

Huo, S.-M., Yan, Z.-C., Zhang, F., Chen, L., Sun, J.-T., Hoffmann, A.A., Hong, X.-Y., 2021. Comparative genome and transcriptome analyses reveal innate differences in response to host plants by two color forms of the two-spotted spider mite Tetranychus urticae. BMC Genomics 22, 569. https://doi.org/10.1186/s12864-021-07894-7

Huyse, T., Poulin, R., Théron, A., Theron, A., 2005. Speciation in parasites: a population genetics approach. Trends Parasitol. 21, 469–475.

Jombart, T., Collins, C., 2015. Analysing genome-wide SNP data using adegenet 2.0. 0.

Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. <scp>Clumpak</scp> : a program for identifying clustering modes and packaging population structure inferences across *K*. Mol. Ecol. Resour. 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kvičerová, J., Hypša, V., 2013. Host-parasite incongruences in rodent Eimeria suggest significant role of adaptation rather than cophylogeny in maintenance of host specificity. PLoS One 8, e63601.

Lam, L.K.M., Dobkin, J., Eckart, K.A., Gereg, I., DiSalvo, A., Nolder, A., Anis, E., Ellis, J.C., Turner, G., Mangalmurti, N.S., 2022. Bat Red Blood Cells Express Nucleic Acid–Sensing Receptors and Bind RNA and DNA. ImmunoHorizons 6, 299–306. https://doi.org/10.4049/immunohorizons.2200013

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Li, D., Wang, Mengxuan, Mao, T., Wang, Mingwen, Zhang, Q., Wang, H., Pang, L., Sun, X., Duan, Z., 2021. The Functional Characterization of Bat and Human P[3] Rotavirus VP8*s. Virol. Sin. 36, 1187–1196. https://doi.org/10.1007/s12250-021-00400-z

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu, F., Chen, Z., Liu, N., 2017. Molecular Basis of Olfactory Chemoreception in the Common Bed Bug, Cimex lectularius. Sci. Rep. 7, 45531. https://doi.org/10.1038/srep45531

Liu, F., Liu, N., 2015. Human Odorant Reception in the Common Bed Bug, Cimex lectularius. Sci. Rep. 5, 15558. https://doi.org/10.1038/srep15558

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: population inference from RADseq data. Mol. Biol. Evol. 35, 1284–1290.

Matsubayashi, K.W., Yamaguchi, R., 2022. The speciation view: Disentangling multiple causes of adaptive and non-adaptive radiation in terms of speciation. Popul. Ecol. 64, 95–107. https://doi.org/10.1002/1438-390X.12103

Mussmann, S.M., Douglas, M.R., Chafin, T.K., Douglas, M.E., 2020. AdmixPipe: population analyses in Admixture for non-model organisms. BMC Bioinformatics 21, 337. https://doi.org/10.1186/s12859-020-03701-4

Nazarizadeh, M., Nováková, M., Drábková, M., Catchen, J., Olson, P.D., Štefka, J., 2024a. Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm Ligula intestinalis. Proc. R. Soc. B Biol. Sci. 291, 2007–2023. https://doi.org/10.1098/rspb.2023.2563

Nazarizadeh, M., Nováková, M., Loot, G., Gabagambi, N.P., Fatemizadeh, F., Osano, O., Presswell, B., Poulin, R., Vitál, Z., Scholz, T., Halajian, A., Trucchi, E., Kočová, P., Štefka, J., 2023. Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite – The Ligula intestinalis species complex. Mol. Phylogenet. Evol. 180, 107677. https://doi.org/10.1016/j.ympev.2022.107677

Nazarizadeh, M., Nováková, M., Vlček, J., Štefka, J., 2024b. Host-associated genetic differentiation in the face of ongoing gene flow: ecological speciation in a pathogenic parasite of freshwater fish. bioRxiv 2001–2024. https://doi.org/10.1101/2024.01.27.577373

Pawłowska, M., Mila-Kierzenkowska, C., Szczegielniak, J., Woźniak, A., 2023. Oxidative Stress in Parasitic Diseases—Reactive Oxygen Species as Mediators of Interactions between the Host and the Parasites. Antioxidants 13, 38. https://doi.org/10.3390/antiox13010038

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS One 7, e37135. https://doi.org/10.1371/journal.pone.0037135

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J.,

Sklar, P., De Bakker, P.I.W., Daly, M.J., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575.

Reda, A.G., Huwe, T., Koepfli, C., Assefa, A., Tessema, S.K., Messele, A., Golassa, L., Mamo, H., 2023. Amplicon deep sequencing of five highly polymorphic markers of Plasmodium falciparum reveals high parasite genetic diversity and moderate population structure in Ethiopia. Malar. J. 22, 376. https://doi.org/10.1186/s12936-023-04814-w

Robison, G.A., Balvin, O., Schal, C., Vargo, E.L., Booth, W., 2015. Extensive mitochondrial heteroplasmy in natural populations of a resurging human pest, the bed bug (Hemiptera: Cimicidae). J. Med. Entomol. 52, 734–738.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol. Ecol. 28, 4737–4754. https://doi.org/10.1111/mec.15253

Rost-Roszkowska, M.M., Vilimova, J., Włodarczyk, A., Sonakowska, L., Kamińska, K., Kaszuba, F., Marchewka, A., Sadílek, D., 2017. Investigation of the midgut structure and ultrastructure in Cimex lectularius and Cimex pipistrelli (Hemiptera: Cimicidae). Neotrop. Entomol. 46, 45–57. https://doi.org/10.1007/s13744-016-0430-x

Roth, S., Balvín, O., Siva-Jothy, M.T., Di Iorio, O., Benda, P., Calva, O., Faundez, E.I., Khan, F.A.A., McFadzen, M., Lehnert, M.P., 2019. Bedbugs evolved before their bat hosts and did not co-speciate with ancient humans. Curr. Biol. 29, 1847–1853.

Rousset, F., 2002. Inbreeding and relatedness coefficients: what do they measure? Heredity (Edinb). 88, 371–380. https://doi.org/10.1038/sj.hdy.6800065

Santiago, P.B., De Araújo, C.N., Motta, F.N., Praça, Y.R., Charneau, S., Bastos, I.M.D., Santana, J.M., 2017. Proteases of haematophagous arthropod vectors are involved in blood-feeding, yolk formation and immunity - a review. Parasites and Vectors 10, 1–20. https://doi.org/10.1186/s13071-017-2005-z

Sasínková, M., Balvín, O., Vandrovcová, J., Massino, C., Weig, A.R., Reinhardt, K., Otti, O., Bartonička, T., 2023. Despite genetic isolation in sympatry, post-copulatory reproductive barriers have not evolved between bat- and human-associated common bedbugs (Cimex lectularius L.). Front. Zool. 20, 36. https://doi.org/10.1186/s12983-023-00514-y

Schweyen, H., Rozenberg, A., Leese, F., 2014. Detection and Removal of PCR Duplicates in Population Genomic ddRAD Studies by Addition of a Degenerate Base Region (DBR) in Sequencing Adapters. Biol. Bull. 227, 146–160. https://doi.org/10.1086/BBLv227n2p146

Shi, J., Walker, M.G., 2007. Gene set enrichment analysis (GSEA) for interpreting gene expression profiles. Curr. Bioinform. 2, 133–137.

Talbot, B., Balvín, O., Vonhof, M.J., Broders, H.G., Fenton, B., Keyghobadi, N., 2017. Host association and selection on salivary protein genes in bed bugs and related blood-feeding ectoparasites. R. Soc. Open Sci. 4, 170446. https://doi.org/10.1098/rsos.170446

Talbot, B., Keyghobadi, N., Fenton, B., 2019. Bed bugs: The move to humans as hosts. FACETS 4, 105–110. https://doi.org/10.1139/facets-2018-0038

Toh, S.Q., Glanfield, A., Gobert, G.N., Jones, M.K., 2010. Heme and blood-feeding parasites: friends or foes? Parasit. Vectors 3, 108. https://doi.org/10.1186/1756-3305-3-108

Usinger, R.L., 1966. Monograph of Cimicidae (Hemiptera, Heteroptera).

Vorel, J., Cwiklinski, K., Roudnický, P., Ilgová, J., Jedličková, L., Dalton, J.P., Mikeš, L., Gelnar, M., Kašný, M., 2021. Eudiplozoon nipponicum (Monogenea, Diplozoidae) and its adaptation to haematophagy as revealed by transcriptome and secretome profiling. BMC Genomics 22, 1–17. https://doi.org/10.1186/s12864-021-07589-z

Wang, C., Gao, W., Yan, S., Zhu, X.Q., Suo, X., Liu, X., Gupta, N., Hu, M., 2021. N-glycome and N-glycoproteome of a hematophagous parasitic nematode Haemonchus. Comput. Struct. Biotechnol. J. 19, 2486–2496. https://doi.org/10.1016/j.csbj.2021.04.038

Wang, C., Saltzmann, K., Chin, E., Bennett, G.W., Gibb, T., 2010. Characteristics of <I>Cimex lectularius</I> (Hemiptera: Cimicidae), Infestation and Dispersal in a High-Rise Apartment Building. J. Econ. Entomol. 103, 172–177. https://doi.org/10.1603/EC09230

Wang, Shuai, Wang, Sen, Luo, Y., Xiao, L., Luo, X., Gao, S., Dou, Y., Zhang, H., Guo, A., Meng, Q., 2016. Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. Nat. Commun. 7, 1–12. https://doi.org/10.1038/ncomms12845

Wawrocka, K., Bartonička, T., 2013. Two different lineages of bedbug (Cimex lectularius) reflected in host specificity. Parasitol. Res. 112, 3897–3904.

Willis, C.K.R., Brigham, R.M., 2004. Roost switching, roost sharing and social cohesion: forest-dwelling big brown bats, Eptesicus fuscus, conform to the fission–fusion model. Anim. Behav. 68, 495–505. https://doi.org/10.1016/j.anbehav.2003.08.028

Xia, Y., Fei, B., He, J., Zhou, M., Zhang, D., Pan, L., Li, S., Liang, Y., Wang, L., Zhu, J., Li, P., Zheng, A., 2017. Transcriptome analysis reveals the host selection fitness mechanisms of the Rhizoctonia solani AG1IA pathogen. Sci. Rep. 7, 10120. https://doi.org/10.1038/s41598-017-10804-1

Summary

Summary of the results

The interaction between hosts and parasites is a complex and dynamic relationship that profoundly influences their evolutionary trajectories and diversification. While the evolutionary history of host-parasite relationships often reveals patterns of co-speciation, where host and parasite lineages diverge in tandem, there are also instances of host switching, where parasites jump to new host species, leading to rapid adaptation and diversification. This intricate interaction is shaped by multiple factors, including genetic variability, environmental influences such as climate change, divergent selection pressures in host-parasite systems, and the host's immune defences and the parasite's evasion strategies. Using two model organisms, *Ligula* tapeworms and *Cimex* bedbugs, I demonstrate how diversification occurs in parasites with distinct life cycles and identify the key factors influencing parasite interaction with their hosts, leading to speciation. By analysing the genomic diversity and gene expression profiles across different populations of hosts and parasites, I could identify the genetic markers linked to resistance and susceptibility, decipher the mechanisms of host immune responses, and expose the adaptive strategies employed by parasites. Finally, the analyses highlighted key adaptations at various life cycle stages, emphasizing the role of genetic diversity and ecological differentiation in shaping parasite populations.

In Chapter I, my colleagues and I investigated the prevalence and population genetic structure of plerocercoids of the tapeworm *Ligula intestinalis* across various cyprinoid fish species in six water bodies in the Czech Republic. We found that the highest frequency of parasitism was in lake Medard, with roach exhibiting the highest infection rates among the fish species studied. Genetic analyses using mitochondrial genes (Cytb and COI) placed all populations in the common European Clade A, but also revealed significant genetic differentiation between parasite population in freshwater bream and those in roach, bleak, and rudd. This host specificity was supported by haplotype network and PCoA analyses, which showed distinct parasite haplotypes in freshwater bream compared to the other species. Our findings suggested that both ecological factors and genetic diversity may influence the infection rates and distribution patterns of *L. intestinalis*. The results provided indication that host specificity could be the mechanisms driving differentiation in *L. intestinalis*.,

In Chapter II, we revealed the evolutionary history and diversification of the globally distributed *L. intestinalis* species complex, using mitochondrial DNA and genomic SNP data from 139 specimens across 21 countries. We uncovered 10 evolutionary lineages, with the deepest divergence occurring around 4.99-5.05 million years ago. Our historical biogeography analyses revealed that the parasite's ancestor diversified through multiple vicariance events across the Palearctic, Afrotropical, and Nearctic regions. Cyprinoids were the ancestral hosts, and new lineages emerged through biogeographic dispersal and host-switching events from the late Pliocene to Pleistocene. We found that *Ligula* lineages exhibited high host specificity, with almost no overlap in host genera, even in sympatric regions. Species delimitation analyses indicated extensive cryptic diversity, identifying at least 10 putative species. Moreover, historical gene flow and introgression events were detected, particularly between the Nearctic lineage and *L. alternans*, and from Lineage B into China and *L. pavlovskii*. Demographic analyses revealed population expansions and contractions, with significant ancestral bottlenecks and range expansions in various lineages. These findings provide a comprehensive understanding of the evolutionary pressures and mechanisms driving the diversification and host specificity of the *L. intestinalis* species complex.

In Chapter III, we carried out a comprehensive genomic and transcriptomic analysis of *L. intestinalis* to reveal significant insights into its complex life cycle and host interactions. We achieved the first highly resolved genome assembly for *L. intestinalis*. The genome is of substantial size (775.3 Mbp), inflated with repetitive sequences comprising over 60% of the genome length. Comparative RNA analyses identified 3922 differentially expressed genes between the plerocercoid larvae and adult stages, uncovering genes linked to reproductive development in adults and immune evasion strategies in larvae. In adults, we found significant enrichment of genes associated with cytoskeleton organization, cilium formation, and microtubule-based movement, which are crucial for sperm production and overall reproductive development. Specific molecular functions important for vitellogenesis and the formation of spermatozoa were also identified, including genes involved in axoneme assembly, sperm motility, and microtubule-based processes. Additionally, the study noted a high frequency of unique orthologous clusters, particularly related to DNA integration and retrovirus-related proteins, suggesting potential functional diversification and host specialization. Synteny analysis demonstrated significant genome conservation with other tapeworms, while gene

family evolution indicated notable expansions and contractions, contributing to the parasite's adaptation and speciation. These findings provide valuable resources for further research into developmental biology, host-parasite coevolution, and the molecular mechanisms of parasitism, advancing our understanding and potential management of parasitic infections.

In Chapter IV, building on the results of the previous studies aimed at identifying host-specific lineages (Chapter I) and obtaining the genome assembly (Chapter III), we examined the mechanisms underlying ecological speciation in the tapeworm *L. intestinalis*, with a particular focus on host-associated genetic differentiation amidst ongoing gene flow. Utilizing genome-wide SNPs and transcriptome data, we analysed the population genetic structure and degree of gene flow among parasites from different cyprinid fish hosts occurring in sympatry. Our genetic clustering analysis identified significant divergence between parasites infecting bream and those infecting roaches, bleaks, and rudds. Historical demography modelling suggested that this divergence likely resulted from isolation with continuous gene flow. Selection analyses identified 896 SNPs under selection, exhibiting higher nucleotide diversity and genetic divergence compared to neutral loci. Transcriptome profiling corroborated these results, revealing distinct gene expression profiles for the two parasite populations, with specific genes involved in immune evasion driving ecological speciation.

In Chapter V, we focused on the human-associated spread of non-native *C. hemipterus* across Central Europe, into the area of a historically prevalent *C. lectularius*, demonstrating how modern travel and urbanization have facilitated the migration of bedbug species beyond their historical geographical limits. By integrating genetic analysis with field data from infestations recorded since 2002, we have provided evidence that *C. hemipterus* has not only arrived but has established new populations in non-tropical environments. Our analysis of 566 infestations revealed that *C. hemipterus* occurred in six infestations since 2019, suggesting its recent establishment. Genetic sequencing identified two kdr-associated mutations, M918I and L1014F, indicating insecticide resistance, a factor complicating eradication efforts. Concurrently, we compared these findings with the historically prevalent *Cimex lectularius* to assess differences in spread patterns and resistance profiles, which highlighted the ongoing challenges in managing both species. This chapter details our methodology, from sampling to genetic sequencing, and discusses the implications of our findings for pest management and public health policies in Europe. We highlight the critical role of local spread within communities, contrasting with earlier assumptions that infestations were predominantly linked

to international travel. This insight emphasizes the need for revised monitoring strategies that consider both species of bedbugs due to their distinct behavioural and physiological traits.

In Chapter VI, we investigated the genetic and transcriptomic divergence in *C. lectularius* populations parasitizing humans and bats, focusing on how host switching influences genetic diversity and adaptability. Our genome-wide SNP and transcriptome analyses revealed significant genetic differences between human-associated and bat-associated bedbugs. Bat-associated bedbugs exhibited higher genetic diversity and lower inbreeding compared to their human-associated counterparts, attributed to bats' mobility and diverse habitats, promoting gene flow. Transcriptomic profiling showed distinct gene expression profiles related to blood digestion enzymes, with bat-associated bedbugs exhibiting significant genetic adaptation when switching from bat to human blood. Conversely, human-associated bedbugs exhibited minimal changes, indicating versatile physiological mechanisms and enhanced odorant-binding capabilities. These findings highlight the complex interplay between host specificity, genetic diversity, and adaptability in bedbug populations. Together, all these six chapters provided a comprehensive picture of the genetic and ecological dynamics shaping the *Ligula and Cimex* and its interactions with various hosts.

Future perspectives

Building on the findings from these six chapters, future research can explore parallel evolution in the *L. intestinalis* species complex to understand how different lineages independently adapt to similar ecological conditions. For *C. lectularius*, a detailed set of samples has been collected from their re-emerging populations across Europe. Processing these samples with ddRAD data will facilitate a comprehensive genetic analysis, enabling comparisons of genetic structures within European populations. Additionally, through an already established international collaboration (team of Dr. Booth at Virginia Tech, USA), it will also be possible to compare the genetic structures of parasite populations from Europe with those in Canada and the USA. This comparison will help track gene flow across continents and reveal adaptive strategies these populations have evolved in response to urban environments and pest control measures. Understanding these patterns is crucial for predicting future spreads and developing targeted control strategies. Furthermore, examining the adaptive responses within the *L. intestinalis* species complex alongside *C. lectularius* offers a unique comparative model to study convergent evolutionary strategies. This approach provides insights into the resilience and adaptability of these organisms under similar pressures. Below, I discuss in more detail the possible routes of such research, with some steps already initiated.

By using detailed sampling of ecological replicates and extensive genetic resources, studies can discern between different genomic mechanisms driving parallel evolution. Key questions include whether parallel evolution operates by selecting on standing genetic variation or by promoting the evolution of novel variants and gene functions (Roesti et al., 2015; Wos et al., 2022). The research should aim to quantify the extent of genetic reuse across different lineages of *L. intestinalis* and identify specific loci under selection. *Ligula* lineages in Europe (Lineage A and B) and Africa (Ethiopia) represent ideal candidates, because each of these lineages uses multiple fish hosts and intra-lineage host driven differentiation was found in at least some of them (In Chapter II; Štefka et al., 2009). Whole genome resequencing and transcriptomic analysis of several representatives from each lineage will not only elucidate the genetic basis of parallel adaptation but also provide insights into the predictability of parasite evolution, which is crucial for managing emerging parasitic threats.

Another critical future direction is to improve genomic resources for parasites such as *Ligula* intestinalis and *Cimex* species. Although genomic resources exist for both taxa, obtained here for Ligula (Chapter III) and by Benoit et al. (2016) for *Cimex*, achieving chromosome-level genome assemblies through advanced sequencing technologies like PacBio HiFi and integrating these with transcriptomic data will significantly enhance our understanding of genome structure and function. Future research should aim to refine these genomic assemblies and utilize advanced bioinformatics tools to analyse gene functions, regulatory elements, and structural variations. These efforts will enable precise functional genomics studies, revealing the molecular mechanisms of host-parasite interactions and adaptations. Moreover, detailed annotations of parasite genomes can uncover gene families involved in key biological processes such as host invasion, immune evasion, and reproduction. Enhanced genomic resources will facilitate the development of targeted interventions, such as novel therapeutics and vaccines, to combat parasitic diseases effectively. Additionally, a well-annotated genomic database can serve as a valuable resource for the broader scientific community, enabling collaborative research and accelerating discoveries in parasitology and related fields.

Future studies should investigate also the changing ecological and environmental conditions influencing host-parasite dynamics. This includes examining how climate change, habitat fragmentation, and other environmental pressures affect the prevalence and distribution of parasitic infections. Research should focus on understanding the interactions of parasites with different host species across various ecosystems, revealing patterns of host switching (for example due to host introductions as in Bouzid et al., 2013), local adaptation, and speciation. Long-term ecological studies and field experiments can provide insights into how environmental changes drive the evolution of host-parasite relationships.

Future studies should build upon the foundational findings regarding the genetic and developmental processes of *Ligula intestinalis*, particularly leveraging insights gained from in vitro experiments. In Chapter III of my thesis, we successfully cultured the tapeworm from plerocercoid to adult stage and identified several genes that play critical roles in the lifecycle of the tapeworm. However, we were unable to complete its entire life cycle, particularly lacking in understanding the genes employed at the procercoid stage. Therefore, future research should aim to culture a complete lifecycle of the tapeworm and utilize advanced genomic and transcriptomic techniques to investigate gene expression profiles across various life stages. This should focus especially on the transitions between the coracidium and procercoid stages

within the primary copepod hosts. Additionally, while Chapter II of my thesis employed sophisticated bioinformatic analysis to estimate the mutation rate per generation for the tapeworm, in vitro experiments sequencing trios (parents against progeny) allows for a more precise estimate. Such data will be invaluable for future population genomics studies aimed at revealing precise demographic histories not only of this tapeworm but also of other tapeworms within the order Diphyllobothriidea, where the majority of the species share similar genome sizes (see chapter III).

RADseq protocol used in Chapter VI to differentiate between bat and human-associated bedbugs provides enough information to study population structure and routes of re-emergence of bed bug populations in developed countries. Indeed, our current sample set involves not only a majority of the countries in Europe, but in some cases also dense sampling within cities. For example, we obtained 220 specimens from Oslo, which allows for a detailed analysis of genetic diversity and potential migration patterns within these urban settings. The comprehensive data set enables an in-depth understanding of the ecological dynamics and adaptive strategies of bedbugs as they navigate urban complexities. Furthermore, by estimating the contemporary gene flow among bedbug populations from different buildings across various locales, we can trace lineage connections and dispersal mechanisms, shedding light on the critical factors driving the resurgence of bedbugs at both local and regional scales. This extensive sampling and detailed genetic profiling set the stage for developing more effective control strategies tailored to the specific characteristics of bedbug populations in various European cities

References

Benoit, J.B., Adelman, Z.N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E.C., Szuter, E.M., Hagan, R.W., Gujar, H., Shukla, J.N., Zhu, F., Mohan, M., Nelson, D.R., Rosendale, A.J., Derst, C., Resnik, V., Wernig, S., Menegazzi, P., Wegener, C., Peschel, N., Hendershot, J.M., Blenau, W., Predel, R., Johnston, P.R., Ioannidis, P., Waterhouse, R.M., Nauen, R., Schorn, C., Ott, M.C., Maiwald, F., Johnston, J.S., Gondhalekar, A.D., Scharf, M.E., Peterson, B.F., Raje, K.R., Hottel, B.A., Armisén, D., Crumière, A.J.J., Refki, P.N., Santos, M.E., Sghaier, E., Viala, S., Khila, A., Ahn, S.J., Childers, C., Lee, C.Y., Lin, H., Hughes, D.S.T., Duncan, E.J., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Wheeler, D., Panfilio, K.A., Vargas Jentzsch, I.M., Vargo, E.L., Booth, W., Friedrich, M., Weirauch, M.T., Anderson, M.A.E., Jones, J.W., Mittapalli, O., Zhao, C., Zhou, J.J., Evans, J.D., Attardo, G.M., Robertson, H.M., Zdobnov, E.M., Ribeiro, J.M.C., Gibbs, R.A., Werren, J.H., Palli, S.R., Schal, C., Richards, S., 2016. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat. Commun. 7, 1–10. https://doi.org/10.1038/ncomms10165

Bouzid, W., Štefka, J., Bahri-Sfar, L., Beerli, P., Loot, G., Lek, S., Haddaoui, N., Hypša, V., Scholz, T., Dkhil-Abbes, T., 2013. Pathways of cryptic invasion in a fish parasite traced using coalescent analysis and epidemiological survey. Biol. Invasions 15, 1907–1923.

Roesti, M., Kueng, B., Moser, D., Berner, D., 2015. The genomics of ecological vicariance in threespine stickleback fish. Nat. Commun. 6, 1–14. https://doi.org/10.1038/ncomms9767

Štefka, J., Hypša, V., Scholz, T., 2009. Interplay of host specificity and biogeography in the population structure of a cosmopolitan endoparasite: Microsatellite study of Ligula intestinalis (Cestoda). Mol. Ecol. 18, 1187–1206. https://doi.org/10.1111/j.1365-294X.2008.04074.x

Wos, G., Arc, E., Huelber, K., Konečná, V., Knotek, A., Požárová, D., Bertel, C., Kaplenig, D., Mandakova, T., Neuner, G., 2022. Parallel local adaptation to an alpine environment in Arabidopsis arenosa. J. Ecol.

Author’s Curriculum vitae

Author's Curriculum vitae

Name: Msc. Masoud Nazarizadeh

Date of Birth: 12/04/1990

Nationality: Iranian

Phone: +420 605831106

Email: Nazarizadeh@yahoo.com

ORCID: 0000-0002-1891-9870

- **Education:**

2020-now PhD candidate in Parasitology, Faculty of Science, University of South Bohemia

2014-2016 Master's degree in environmental science, Faculty of Science, University of Tehran

2010-2014 Bachelor's degree in environmental science, faculty of Science, University of Tehran

- **Employment History**

2020- now Researcher, Laboratory of Molecular Ecology and Evolution, Institute of parasitology, Biology Centre CAS

2016-2019- Research assistant, Faculty of Science, University of Tehran.

- **Internship**

October-December 2022 Peter D Olson, Department of Life Sciences, The Natural History

Museum, London, United Kingdom

May-July 2021 Emiliano Trucchi, Department of Life and Environmental Sciences, Marche

Polytechnic University, Ancona, Italy

- **Teaching**

2021 – 2023 teaching assistant, Molecular ecology, Faculty of Science, University of South Bohemia.

2021 – 2023 teaching assistant, Population genetics, Faculty of Science, University of South Bohemia.

- **Publications (H-index 8)**

Damnjanovic, D., **Nazarizadeh, M**., Wisniewska, M., Pavel, V., Chutný, B., Johnsen, A., Nováková, M., Štefka, J., (2024) Limited genetic depletion despite extinction risk: genomic diversity of a peripheral population of red-spotted bluethroats in central Europe, *Zoological Journal of the Linnean Society* (In press).

Kamali K, **Nazarizadeh M**, Fatemizadeh F, Salmabadi S, Hung CM, Kaboli M, (2024) Integrating phylogenetic, phylogeographic, and morphometric analyses to reveal cryptic lineages within the genus *Asaccus* (Reptilia: Squamata: Phyllodactylidae) in Iran, *BMC Zoology 9, 12 (2024).*

Nazarizadeh, M., Nováková, M., Drábková, M., Catchen, J., Olson, P.D. and Štefka, J., 2024. Highly resolved genome assembly and comparative transcriptome profiling reveal genes related to developmental stages of tapeworm *Ligula intestinalis*. *Proceedings of the Royal Society B*, *291*(2015), p.20232563.

Nazarizadeh, M., Nováková, M., Vlček, J. and Štefka, J., 2024. Host-associated genetic differentiation in the face of ongoing gene flow: ecological speciation in a pathogenic parasite of freshwater fish. bioRxiv, pp.2024-01.

Nazarizadeh, M., Nováková, M., Loot, G., Gabagambi, N.P., Fatemizadeh, F., Osano, O., Presswell, B., Poulin, R., Vitál, Z., Scholz, T. and Halajian, A., 2023. Historical dispersal and host-switching formed the evolutionary history of a globally distributed multi-host parasite–The Ligula intestinalis species complex. *Molecular Phylogenetics and Evolution*, 180, p.107677.

Nazarizadeh, M., Peterka, J., Kubečka, J., Vašek, M., Jůza, T., de Moraes, K.R., Čech, M., Holubová, M., Souza, A.T., Blabolil, P. and Muška, M., 2022. Different hosts in different lakes: prevalence and population genetic structure of plerocercoids of *Ligula intestinalis* (Cestoda) in Czech water bodies. *Folia Parasitologica*, 69, pp.1-10.

Nazarizadeh, M., Martinů, J., Nováková, M., Stanko, M. and Štefka, J., 2022. Phylogeography of the parasitic mite *Laelaps agilis* in Western Palearctic shows lineages lacking host specificity but possessing different demographic histories. *BMC zoology*, 7(1), p.15.

Balvín, O., Sasínková, M., Martinů, J., **Nazarizadeh, M**., Bubová, T., Booth, W., Vargo, E.L. and Štefka, J., 2021. Early evidence of establishment of the tropical bedbug (Cimex hemipterus) in Central Europe. *Medical and Veterinary Entomology*, 35(3), pp.462-467.

Ahmadi, M., Hemami, M.R., Kaboli, M., **Nazarizadeh, M**., Malekian, M., Behrooz, R., Geniez, P., Alroy, J. and Zimmermann, N.E., 2021. The legacy of Eastern Mediterranean mountain uplifts: rapid disparity of phylogenetic niche conservatism and divergence in mountain vipers. *BMC Ecology and Evolution*, 21(1), p.130.

Kazemi, E., **Nazarizadeh, M**., Fatemizadeh, F., Khani, A. and Kaboli, M., 2021. The phylogeny, phylogeography, and diversification history of the westernmost Asian cobra (Serpentes: Elapidae: *Naja oxiana*) in the Trans-Caspian region. *Ecology and Evolution*, 11(5), pp.2024-2039.

Asadi, A., Montgelard, C., **Nazarizadeh, M**., Moghaddasi, A., Fatemizadeh, F., Simonov, E., Kami, H.G. and Kaboli, M., 2019. Evolutionary history and postglacial colonization of an Asian pit viper (Gloydius halys caucasicus) into Transcaucasia revealed by phylogenetic and phylogeographic analyses. *Scientific Reports*, 9(1), p.1224.

Chen, Y.C., **Nazarizadeh, M**., Lei, F.M., Yang, X.J., Yao, C.T., Dong, F., Dong, L., Zou, F.S., Drovetski, S.V., Liu, Y. and Huang, C.C., 2019. The niches of nuthatches affect their lineage evolution differently across latitude. *Molecular Ecology*, 28(4), pp.803-817.

Ahmadi, M., Naderi, M., Kaboli, M., **Nazarizadeh, M**., Karami, M. and Beitollahi, S.M., 2018. Evolutionary applications of phylogenetically-informed ecological niche modelling (ENM) to explore cryptic diversification over cryptic refugia. *Molecular Phylogenetics and Evolution*, 127, pp.712-722.

Behrooz, R., Kaboli, M., Arnal, V., **Nazarizadeh, M**., Asadi, A., Salmanian, A., Ahmadi, M. and Montgelard, C., 2018. Conservation below the species level: suitable evolutionarily significant units among Mountain Vipers (the Montivipera raddei complex) in Iran. *Journal of Heredity*, *109*(4), pp.416-425.

Yalpanian, S., **Nazarizadeh, M**., Asadi, A., Omidi, M., Rezaei, A. and Kaboli, M., 2018. A study of climate niche modeling of the Eurasian Nuthatch (*Sitta europaea*) in theAlborz and Zagros mountains. *Journal of Natural Environment*, *71*(2), pp.269-285.

Nazarizadeh, M., Kaboli, M., Rezaie, H.R., Harisini, J.I. and Pasquet, E., 2016. Phylogenetic relationships of Eurasian Nuthatches (Sitta europaea Linnaeus, 1758) from the Alborz and Zagros Mountains, Iran. *Zoology in the Middle East*, *62*(3), pp.217-226.

Nazarizadeh, M., Fatemizadeh, F., Kaboli, M., Cheraghi, S., Hashemi, A. and Tohidifar, M., 2015. Home Range of Pleske's Ground Jay Podoces pleskei in a Breeding Season in Touran Biosphere Reserve, North-Central Iran. *Podoceps*.

- **Workshop Attendance**

2020 Workshop on Genomics, Cesky Krumlov, Czech Republic

2020 Workshop on population and speciation genomics, Cesky Krumlov, Czech Republic

- **Conferences**

Nazarizadeh M., Novakova M., Drabkova M., Stefka J. genomic differentiation in populations of a tapeworm parasite associated with specificity to different sympatric hosts: a case of ecological speciation. society for molecular biology and evolution (SMBE23), 23-27, 08, 2023, Ferrara, Italy. (poster)

Nazarizadeh M., Novakova M., Stefka J. Historical biogeography of a multi-host parasite (Cestoda: *Ligula intestinalis*). Congress of the European Society for Evoluationary biology (ESEB), 14-19, 08, 2022, Prague, Czech Republic. (poster)

Novakova M., **Nazarizadeh M**., Stefka J. Ecological speciation via co-evolution? Population genomic. differentiation in a freshwater parasite. Congress of the European Society for Evoluationary biology (ESEB), 14-19,08, 2022, Prague, Czech Republic. (poster)

Cvetković T., **Nazarizadeh M**., Flegrová T., Drábková M., Trucchi E., Štefka J. Differences in population structure between two octopus species, *Eledone moschata* and *E. cirrhosa*, in the Mediterranean Sea. 9th Società Italiana di Biologia Evoluzionistica (SIBE), 4-7,09, 2022, Ancona, Italy.

Sasínková M., Balvín O., Martinů J., **Nazarizadeh M**., Booth W., Vargo E., Štefka J. Population genetic structure of the common bed bug in Europe and the USA. The 26th International Congress of Entomology ICE 2022, 17-22, 2022, Helsinki, Finland

- **Language**

Persian, English (*C1 level*; TOEFL iBT score: 95, with a writing skill of 27 out of 30; *C2 level*)

- **Bioinformatics Skills (professional skilled in R, Python, Bash)**

Genomic Analyses (NGS and Third generation sequencing technologies)

e.g., BWA, BOWTIE2, SAMTOOLS, BCFTOOLS, GATK, BlobTools , Selection Analyses, demographic history analyses, Chromosome synteny

Genome-wide SNPs (e.g., Stacks, ipyrad, pyRAD, dDocent, TreeMix, Admixture, Faststructure, fineRADstructure, DAPC, PCA, demographic history analyses)

Transcriptome analyses (e.g., STAR, Hisat2, Subread, edgeR, DEseq2, DiCoExpress)

Amplicon analyses (e.g., QIIME2, vsearch, DADA2)

Phylogeny Analyses (DNAsp- Mega- MrBayes- RAxML- PHyML- JModel Test- Figtree - Network- Popart- Bioedit- Phyml- Arliquin- Genpop- Microchecker- Structure. Seqscape- PAUP- phyde- IMa- Phase- Geneland- Iqtree. Spilttree-BAPS)

Phylogeography Analysis (e.g., Pop ABC-Beast-GenGIS- SPREAD- BEAST- ANECA- GEODIS- TCS)

Biogeography Analysis (e.g., BioGeoBEARS R Package-Lagrange- RASP- DIVA)

Ecological Niche Analysis (e.g., MaXEnt- ENMTOOLS- ecospat- hypervolume- adehabitat R Packages)

Phylogenetic Niche Conservatism Analysis (e.g., Phyloclim R package)

Home range analysis (e.g.,Kernel Density Estimation (KDE)- Maximum Convex Polygon (MCP)- adehabitatHR r package)

ArcGIS (ESRI)