RESEARCH

Open Access

Adaptive evolution of stress response genes in parasites aligns with host niche diversity

Armando J. Cruz-Laufer^{1,2*}, Maarten P. M. Vanhove¹, Lutz Bachmann³, Maxwell Barson⁴, Hassan Bassirou⁵, Arnold R. Bitja Nyom^{5,6}, Mare Geraerts^{1,7}, Christoph Hahn^{8*}, Tine Huyse⁹, Gyrhaiss Kapepula Kasembele^{1,10}, Samuel Njom⁵, Philipp Resl⁸, Karen Smeets¹ and Nikol Kmentová^{1,11}

Abstract

Background Stress responses are key the survival of parasites and, consequently, also the evolutionary success of these organisms. Despite this importance, our understanding of the evolution of molecular pathways dealing with environmental stressors in parasitic animals remains limited. Here, we tested the link between adaptive evolution of parasite stress response genes and their ecological diversity and species richness. We comparatively investigated antioxidant, heat shock, osmoregulatory, and behaviour-related genes (*foraging*) in two model parasitic flatworm lineages with contrasting ecological diversity, *Cichlidogyrus* and *Kapentagyrus* (Platyhelminthes: Monopisthocotyla), through whole-genome sequencing of 11 species followed by in silico exon bait capture as well as phylogenetic and codon analyses.

Results We assembled the sequences of 48 stress-related genes and report the first *foraging (For)* gene orthologs in flatworms. We found duplications of heat shock (*Hsp*) and oxidative stress genes in *Cichlidogyrus* compared to *Kapentagyrus*. We also observed positive selection patterns in genes related to mitochondrial protein import (*Hsp*) and behaviour (*For*) in species of *Cichlidogyrus* infecting East African cichlids—a host lineage under adaptive radiation. These patterns are consistent with a potential adaptation linked to a co-radiation of these parasites and their hosts. Additionally, the absence of cytochrome P450 and kappa and sigma-class glutathione S-transferases in monogenean flatworms is reported, genes considered essential for metazoan life.

Conclusions This study potentially identifies the first molecular function linked to a flatworm radiation. Furthermore, the observed gene duplications and positive selection indicate the potentially important role of stress responses for the ecological adaptation of parasite species.

Keywords Comparative genomics, Positive selection, Monopisthocotyla, Heat shock proteins, Oxidative stress

Background

Evolutionary theory predicts that a species' ability to maintain homeostasis against environmental stressors fundamentally affects its adaptive potential [1]. This

*Correspondence: Armando J. Cruz-Laufer armando.cruzlaufer@uhasselt.be Christoph Hahn christoph.hahn@uni-graz.at Full list of author information is available at the end of the article paradigm also applies to metazoan parasites. Parasites cause many neglected tropical diseases in humans [2] but remain often overlooked as groups of pathogens, which also applies to research on their stress responses. Stress responses might be of high relevance to parasite evolution due to their role in parasite adaptation. Stress responses can determine a parasite's infectivity and virulence [3, 4]. Effective stress responses can also increase fitness of individuals and populations (microevolution) and permit species to expand host repertoires and geographical ranges, which may give rise to new parasite



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.



species and diseases (macroevolution) [5]. Furthermore, understanding parasite adaptation matters in a world where human activity promotes the rise of emerging infectious diseases as environmental disturbance creates new ecological opportunities for parasite species [6, 7]. Nonetheless, research on parasite stress responses remains largely limited to few well-known humaninfecting species for the purpose of drug development [8, 9]—or studies focus on macroevolutionary adaptations of major parasite clades, e.g. flatworm classes or insect orders [10, 11]. Stress response pathways are rarely comparatively analysed below the level of these major lineages. Here, we aim to address this knowledge gap on how stress response systems evolve in parasite lineages that are closely related and functionally alike.

In parasitology, the ability to use a broad spectrum of resources, i.e. host species, is often considered indicative of an increased adaptive potential, specifically in ectoparasites, which are directly exposed to the environmental stressors experienced by their hosts. Several stress-related proteins have been characterised as determining host usage in parasites, e.g. in insects [12], nematodes [13], and fungi [14], including antioxidant enzymes dealing with reactive oxygen species (oxidative stress response), heat shock proteins assisting with protein

Cichlidogyrus

Page 2 of 16

folding, and aquaporins dealing with osmotic stress. Animals may also respond to environmental stressors through behavioural changes. The *foraging* (*For*) gene of *Drosophila melanogaster* Meigen, 1830 and other species are among the best-known examples of genes determining behavioural differences [15].

Monogenean flatworms (recently split into Monopisthocotyla Brabec, Salomaki, Kolísko, Scholz & Kuchta, 2023 and Polyopisthocotyla Brabec, Salomaki, Kolísko, Scholz & Kuchta, 2023) offer several advantages for comparatively addressing the evolution of stress response in parasites. They have single-host life cycles and the host preferences of various monogenean groups have been studied in detail, ranging from host specialists to generalists. Here, we focus on closely related lineages Cichlidogyrus Paperna, 1960 and Kapentagyrus Kmentová, Gelnar & Vanhove, 2018 [16]. Species of Cichlidogyrus and Kapentagyrus infect host lineages (African cichlid vs. freshwater clupeid fishes) with contrasting species richness and ecological diversity (Fig. 1). Species of Cichlidogyrus are parasites of cichlid fishes, the one of the most species-rich and ecological diverse group of fishes [17]. One subclade (Cichlidogyrus spp. infecting East African cichlids) is reported from a host lineage that has undergone multiple rapid diversification events

Kapentagyrus



Fig. 1 The two flatworm parasite lineages *Cichlidogyrus* and *Kapentagyrus* differ substantially in species richness and host diversity. Species of *Cichlidogyrus* infect the gills of the hyperdiverse African cichlid fishes that include the adaptive radiations of Lake Tanganyika in East Africa [118]. Species of *Kapentagyrus* infect the gills of African freshwater clupeid fishes, an ecologically conserved group of 22 species inhabiting only pelagic environments of lakes and rivers [89]. Based on these differences, we hypothesise that stress responses of *Cichlidogyrus* have adapted to this enhanced ecological diversity of their hosts

(adaptive radiations) in its recent evolutionary history [18], coinciding with a high parasite species richness (Lake Tanganyika: n = 45, total: n = 144) (Fig. 1). Species of Kapentagyrus are parasites of African freshwater clupeids, all of which inhabit pelagic environments of rivers and lakes [19]. This niche conservatism is reflected in a much lower number of parasite species (n=14). Here, we explore the diversity and adaptive evolution of genes encoding antioxidant enzymes, heat shock proteins, aquaporins, and For orthologs. We hypothesise duplication and positive selection in stress genes of Cichlidogyrus compared to Kapentagyrus, which is species-poor and infects a species-poor, ecologically conserved host lineage. With whole-genome sequencing data of 11 species, our study provides the largest genomic dataset from a single flatworm lineage to date. With 345 single-copy orthologs and 48 stress gene models, we present the most extensive multi-species analysis of stress genes in parasitic flatworms. Our study highlights the role of stress responses in the adaptive evolution of parasites.

Results

Species trees

As a phylogenetic backbone for downstream analyses, we inferred the evolutionary history of the two monogenean parasite lineages through phylogenomic analyses of single-copy ortholog genes. We assembled the nucleotide sequences of conserved single copy genes via in silico exon bait capture using orthologs of *Scutogyrus longicornis* (Paperna & Thurston, 1969) [20] as bait (Fig. 2a, c; single-copy orthologs). After alignment filtering (Fig. 2d), we retained 277 (OMA tree, Fig. 3) and 86 (BUSCO tree, Additional File 1) gene alignments. *Cichlidogyrus* and *Kapentagyrus* form well-supported monophyletic groups (Fig. 3). High support is also found for a clade of species of *Cichlidogyrus* from Lake Tanganyika in East Africa (Fig. 3, Clade Lake Tanganyika).

Copy numbers and phylogenetic patterns of stress genes

Following preparation of the bait files using a genome annotation of *Cichlidogyrus casuarinus* Pariselle, Muterezi Bukinga & Vanhove, 2015 (see Additional File 2 [21-50]; Fig. 2a: single-copy orthologs) (WGS accession: JBJKFK000000000), search sequences of non-monogenean flatworms and other organisms (Fig. 2a: stress genes), and in-situ exon bait capture (Fig. 2c), we assembled nucleotide sequences of 48 putative stress genes of 11 monogenean species (Fig. 4a). The sequences of the 42 target genes included functional groups of their expected gene family (Fig. 4b, Additional File 3). A majority (63%) of the sequences matched with the reference transcriptome data of *S. longicornis* [51] (>95% identity and query coverage) (Fig. 4c), but three out of nine heat shock

Hsp70 genes and all glutathione peroxidase (Gpx) and aquaporin (Aqp) variants were not found in the transcriptome. It falls outside the scope of the current study why these genes were not detected in the reference transcriptome, and it can only be speculated whether this is for biological (real absence) or methodological reasons (e.g. sequencing depth).

For most targeted stress genes, we found a single copy per sequencing read pool (Additional File 4). We detected deviations regarding copy numbers in the draft genome annotation (C. casuarinus) compared to the search sequences of other flatworm parasites, i.e. two Gpx (+1 vs. Schistosoma mansoni), four glutathione Gstm (-6 vs. Echinococcus multilocularis), two peroxiredoxin Prx (-1 vs. S. mansoni), two aquaporin Aqp genes (+1 vs. S. mansoni). No copies of cytochrome P450 genes (Cyp) and several glutathione families (Gsta, Gsto, Gstp, Gsts, and Gstk) were detected neither in C. casuarinus (Additional File 3) and data produced in this study, nor other published monogenean genomes (see the 'Methods' section) using tblastn. Gpx, Gstm, Prx, and all heat shock protein orthologs except *Hsp10* were flagged by *HybPiper* for potential paralogs (Additional File 4). The read pools in this study were generated from pooled individuals. To avoid counting allelic variants in the population as paralogs, highly similar sequences from the same species were excluded from downstream analyses using phylogenetic inference and manual curation. The filtered variants are listed in Additional File 3.

For the Hsp70 family, a multitude of paralogs were flagged for all but two bait sequences (Additional File 4). After filtering, we detected seven well-supported groups of Hsp70 sequences in Cichlidogyrus (Fig. 5a), three of them not detected in Kapentagyrus (groups 3, 4a, and 4b). Through sequence similarity comparison with characterised Hsp70 genes, three Hsp70 groups were predicted to have highly specific functions: the hypoxia upregulated 1 (HYOU1) gene and the endoplasmatic reticulum chaperone binding proteins 1 (BIP1) and 2 (BIP2). Notably, group 4 constituted two orthologs for Cichlidogyrus, but only a single ortholog for Kapentagyrus. Group 3 appeared nested in group 4 (Fig. 5a), but this position might occur due to genetic saturation between the highly divergent Hsp70 groups causing longbranch attraction (see [52]).

For the *Gst* families, we detected seven phylogenetic clusters (Fig. 5b). The mu-class (*Gstm*) sequences did not group according to the four bait sequences of *C. casu-arinus*; hence, group names were reassigned according to the three *Gstm* clades inferred from the phylogenetic tree (Fig. 5b). Notably, *Gstm2* includes two copies for most species of *Cichlidogyrus* with identical gene ontology (GO) terms (Fig. 3) but only one for *Kapentagyrus*



Fig. 2 Schematic bioinformatic strategy for detecting single-copy orthologs (SCOs) and orthologs of stress genes in monogenean whole-genome short-reads. SCO sequences were used to infer the species tree and stress gene sequences for gene family trees and gene trees. **a** Bait sequences were chosen [*S. longicornis* was selected for SCOs; other organisms (non-monogenean flatworms, insects, vertebrates) were selected for stress genes due to lack of monogenean sequences]. **b** Orthologs of these sequences were detected in an annotated genome of *Cichlidogyrus casuarinus* (only stress genes). **c** The putative protein sequences of *S. longicornis* (SCOs)/*C. casuarinus* (stress genes) were used as baits for exon bait capture in the sequencing read pools of species of *Cichlidogyrus* and *Kapentagyrus* through *HybPiper* [99]. **d** Contaminant, variant, low-species coverage, and truncated sequences were filtered from the alignments. **e** Sequences were annotated through *PANNZER2* [103] (**e**)

(Fig. 5b). We consider this absence of a copy as informative as the high sequence similarities of *Gstm2a* and *Gstm2b* (71–77% identical nucleotides) suggest that orthologs of *Gstm2a* in *Kapentagyrus* should have been detected if present. Other than the absence of *Gsto* in *Kapentagyrus* (mentioned above), no further copy number differences between the target species were detected in *Gst*.

The species tree topologies (OMA vs. BUSCO orthologs) of *Cichlidogyrus* were highly similar to each

other (Kendall-Colijn distance: 0) in a multidimensional scaling (MDS) visualisation when excluding species of *Kapentagyrus* (Fig. 6). For all assembled stress genes, gene trees involving the target species were constructed. Some stress gene trees (Additional File 5) deviated from the species tree topology (but unrelated to selection pressures, see below). This topological variation of stress gene tree topologies followed no apparent patterns based on gene function or family (Fig. 6).



Fig. 3 Species tree of *Cichlidogyrus* and *Kapentagyrus* inferred from 277 single-copy orthologs based on a subset of genes selected by Caña-Bozada et al. [20], who used the OMA pipeline. For the tree inferred from orthologs from the BUSCO pipeline, see Additional File 1. Support values: ultrafast bootstraps (UF-Boot)/Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT) (see Methods), asterisks (*) indicate support below threshold (UF-Boot \leq 95, SH-aLRT \leq 80). Abbreviations: Ccas-*Cichlidogyrus casuarinus*, Ccir-*C. cirratus*, Chal-*C. halli*, Ckap-*C. sp.* 'kapembwa', Cscl-*C. sclerosus*, Cthu-*C. thurstonae*, Ctil-*C. tilapiae*, Czam-*C. zambezensis*, Slon-*Scutogyrus longicornis*, Klim-*Kapentagyrus limnotrissae*, Ktan-*K. tanganicanus*, Lake Tanganyika-species endemic to Lake Tanganyika. Scale bar: estimated number of substitutions per site

Detection of positive selection

Positive selection is the process by which gene variants that provide a fitness benefit dominate a population over time [53]. To investigate patterns of adaptive evolution in stress genes, we inferred positive selection from the ratio of substitution rates at nonsynonymous and synonymous sites in protein-coding sequences $(d_N/$ $d_{\rm s}$). We aimed to test whether the genes investigated here show signatures of positive selection regimes and whether differences are present between clades of Cichlidogyrus and Kapentagyrus. Our analyses revealed that seven stress genes had positively selected sites including one for and six Hsp genes (I, Fig. 4d). For cladespecific tests, we detected no differences between Cichlidogyrus and Kapentagyrus (IIa), but two Hsp genes had positively selected sites in East African species of *Cichlidogyrus* (IIb) (Fig. 4d), i.e. the mitochondrial molecular chaperone gene Hsp60 as well as a putative Hsp40 ortholog of the human DnaJ heat shock protein family (HSP40) member A1 (DNAJA1).

Discussion

Stress responses are key factors influencing the ability of parasites to infect their hosts. The current understanding of ecological drivers of the evolution of parasite stress genes is limited as most studies only compare phylogenetically and ecologically distant species. In particular, the role of stress-related genes in adaptive evolution and speciation of metazoan parasites has never been comprehensively addressed.

We detected several unique stress response features in the targeted monogenean genomes. The absence of the cytochrome P450 gene family (*Cyp*) and glutathione S-transferase (*Gst*) sigma- (*Gsts*) and kappaclasses (*Gstk*) is remarkable. CYP enzymes are mainly involved in the (oxidative) metabolism of various endogenous and exogenous compounds and are conserved across almost the entire tree of life [54]. All GST members serve for cellular protection as detoxification enzymes, and *Gsts* and *Gstk* genes were previously reported from all flatworm genomes [55], except Sod Cu

For1

For2

Prx1

Tgr

Gstz

0.50

Species

(a)

coverage

Hsp90 2

Hsp40 7

Hsp40 6

Hsp40_11 Sod_Mn

Hsp40_14

Hsp40 10

Gstm2a

Gstm2b

Gstm1 Hsp40 15

Aqp1

Mgst

Gpx1 Prx3

Hsp40_12

Hsp40 18

Hsp40 5

Hsp40 8

Gpx2

Aqp2

Hsp40_17 Hsp40_9

Hsp40_16 Hsp40_13

Hsp40 4

Hsp90 1

Hsp70 4a

Hsp70 4b

Hsp70 HYOU1

Hsp70_1

Hsp40 3

Hsp40_2

Hsp40_1

Hsp10 Hsp70 BiP2

Hsp60

Gsto1

Gsto2

Hsp70_2

Groups of putative orthologous protein sequences

Hsp70 BiP1



LRT

*
*
*

* * * *

*

* * *

*

*

*

*

*

*

*

*

.

M1a v M0
 M2a v M1a
 M8 v M7
 M1 v M0

Ib) M1 v M0

(c)

transcribed

(b)

1000

500

-500

-1000

-1500

0



functional monogenean Cyp, Gsts, and Gstk may have remained undetected (see also a proposed peroxin gene in a parasitic protozoan [61]). The loss of *Cyp* has hitherto only been reported from protozoan parasites [62-64]. In other parasitic flatworms, i.e. flukes and tapeworms [55], only a single variant was reported that still fulfils vital functions [65]. The gene family has also been reduced in parasitic crustaceans, e.g. salmon lice have the lowest known Cyp copy number of any arthropod [66], which may reflect an evolutionary trend of Cyp contractions in metazoan parasites. For Gsts and Gstk, no losses have been reported in parasitic flatworms so far, but Gsto genes were reported absent in tapeworms [67]. Evolutionary loss of genes and gene functions has also repeatedly been observed elsewhere in parasites, e.g. peroxisomal functions in parasitic protozoans, flatworms, roundworms [68], and crustaceans [66]. These losses have in part been attributed to the r-selected traits of many parasites (e.g. high fecundity, few resources for individual offspring) suggesting that some gene losses might be related to stress response mechanisms [68]. Cyp and Gst gene family contractions and losses in monogeneans may fit this pattern.

The detection of several stress gene families in our target species is the first for monogenean flatworms. We noticed two gene copies of glutathione peroxidase (Gpx), two peroxiredoxins (Prx), six cytosolic glutathione S-transferases (cGst), and two aquaporins (Aqp) (Additional File 3). Monogeneans, thus, differ from other parasitic flatworms, with tapeworms and flukes presenting one Gpx, three Prx, 12 cGst, and one to three Aqp [55, 69] copies. As the functions of these antioxidant enzymes are the reduction of hydrogen peroxide to water (GPX, PRX) and alkyl hydroperoxides to alcohol (PRX), detoxification (cGST), and osmoregulation (AQP), gene family contractions/expansions could provide valuable insight into the functional evolution of parasitic flatworms. For instance, increased *Gpx* copy numbers were linked to higher levels of oxidative stress in mammals [70]. The discussed examples for contractions/expansions of Cyp, Gpx, Prx, cGst, and Aqp in parasites indicate that these gene families are key for





Fig. 5 Maximum likelihood topologies of gene family trees and gene trees of species of *Cichlidogyrus* and *Kapentagyrus*. For abbreviation of species names, support values, and scale bars, see Fig. 3. **a** Gene models of the 70 kDa heat shock protein family (*Hsp70*). Group Hyou1 (hypoxia upregulated 1), BiP1 (endoplasmatic reticulum chaperone binding protein 1), and BiP2 refer to annotations assigned through *PANNZER2* (see Additional File 3); the remaining groups are numbered consecutively. **b** Gene models of the glutathione *S*-transferase (*Gst*) superfamily. Groups are named after *Gst* classes of the bait sequences and numbered consecutively. Group 4 *Hsp70* and Group *Gstm2* show potential duplication events (or gene losses) with two copies of the gene for species of *Cichlidogyrus* but only a single one for species of *Kapentagyrus*

the evolution of parasitism as a whole. Therefore, these genes likely play an important role in the adaptive parasite evolution of parasites.

We also provide the first report of *For* orthologs in flatworms, genes linked to behavioural traits in arthropods, nematodes, mammals, and amphibians [15]. Although fecundity, infection intensity [8], and drug resistance [71] in flatworms have been associated with the cGMPdependent protein kinase (PKG) family, to which the for-encoded protein belongs, the function of For in these organisms remains obscure. Our analyses indicate that one For ortholog has sites under positive selection specific to species of *Cichlidogyrus* from Lake Tanganyika. This observation correlates with the rapid expansion of host-parasites interactions in the lake in recent evolutionary history, but further studies are needed to understand the role of *For* in adaptive evolution [72, 73]. The importance of For and PKGs, in general, for other organisms, its potential role in driving infection intensities in parasitic flatworms, and its adaptive evolution in East African species warrant further studies into this gene family to better understand monogenean behavioural genetics.

Lake Tanganvika is a well-known biodiversity hotspot for several animal groups, particularly cichlid fishes. Indeed, the multiple explosive speciation events have made these fishes an established model system in evolutionary biology [74]. We provide the first evidence that their parasites belonging to *Cichlidogyrus* also present unique functional genetic adaptations compared to species of Cichlidogyrus elsewhere. Positively selected sites in Hsp60 and a putative Hsp40 ortholog of the DnaJ heat shock protein family (HSP40) member A1 (DNAJA1) gene suggest adaptations in the folding/assembly of proteins newly imported into the mitochondria (HSP60) and mitochondrial protein import (DNAJA1) [40] based on functions associated with these genes in other organisms (e.g. humans, D. melanogaster). Prior studies already suggested that at least some Lake Tanganyika monogenean lineages have evolved under evolutionary radiations [72, 73]. Although further evidence is needed, the present findings may indicate that this diversification



Fig. 6 First two axes (49% of total variation) of multidimensional scaling analysis of gene trees of antioxidant enzymes, for orthologs, aquaporins, and heat shock proteins, with some gene tree topologies deviating from the two species trees (highlighted in red) but not forming clusters based on gene function or family

might also be linked to functional adaptations. If true, the cichlid-*Cichlidogyrus* system would represent the first example of a specified genetic adaptation related to an adaptive radiation of flatworms (but see morphological evidence in free-living flatworms from the same lake [75]). To test this hypothesis, evolutionary rates and *Hsp* genetic diversity in more species of *Cichlidogyrus* in and outside of Lake Tanganyika should be analysed.

Species of *Cichlidogyrus* infect an ecologically diverse host lineage, whereas species of *Kapentagyrus* infect a host lineage with a conserved ecological niche (the pelagic zones of rivers and lakes). One may speculate that the gene duplication/loss of stress response genes observed in the *Cichlidogyrus-Kapentagyrus* comparison reflects the contrasting host ecology and evolutionary history and, thus, the adaptive potential of the parasites. Specifically, we identified two potential instances of stress response gene duplication/loss: in comparison with species of *Cichlidogyrus*, species of *Kapentagyrus* lack a gene copy of *Hsp70* (Fig. 5a: group 4b) and *Gstm* (Fig. 5b: *Gstm2a*), and all copies of Gsto. If the additional genes in Cichlidogyrus indeed resulted from duplication, the additional copies may have increased the adaptive potential to stressful conditions, e.g. their ability to adapt to new environments, as has been described for free-living nematodes [76], fungi [14], and invertebrate groups [77]. In metazoan parasites, prior studies detected gene family expansions in tapeworms [10] and aphids [78], but only rarely are these expansions linked to concrete environmental stressors because of unknown gene functions, but see cases among plant-pathogenic moths [79], and nematodes [13]. However, until a more detailed characterisation of the function of the potentially duplicated genes, the assumption of a role in adaptation to stressful conditions of new environments is hypothetical. Similarly, previous interpretations of expansions of Hsp70 among closely related lineages as adaptations to environmental stressors, e.g. in tapeworms [10], trypanosomatid protozoans [80], and invasive fishes [81], need to be taken with caution as these expansions were hypothesised to be expressed 'under certain conditions' [10] or were

only loosely associated with geographical or environmental gradients [80].

No doubt, our study also has conceptual and technological limitations. First, previous studies indicate that copy number evolution can occur between closely related animal species [78] and even strains [80]. No such differences were targeted here because the study focused on differences between lineages not species, an approach taken to avoid mistaking intraspecific allelic gene variants in each pooled DNA samples as paralogs (see the 'Methods' section). Future studies might use variant calling pipelines (e.g. [82]) or optimise techniques to sequence genomes from individual specimens to address this problem. The latter approach has recently been successful with monogenean mitochondrial genomes [58]. Another challenge lies in potentially highly divergent sequences that the in silico exon bait capture might fail to detect. Beyond gene copy numbers, we also found that the evolutionary relationships of the gene orthologs sometimes deviated from the evolutionary history of the species (Fig. 5), but this variation might be an artefact of inferring evolutionary histories from small datasets (243-2811 bp) in contrast to multi-gene phylogenies (279 and 78 kb). Furthermore, we only covered nine out of 144 described species of Cichlidogyrus. Nevertheless, species of Cichlidogyrus constitute a unique study system for host-parasite interactions that combines opportunities to investigate host repertoires and host switching [73], biological invasions [58], and speciation rates [72]. Second, gene models reveal little information on expression patterns. Some genes are only expressed under certain conditions (i.e. inducible genes) and may not be represented in reference transcriptomes [51], e.g. as evidenced for human-infecting flukes [83]. This might explain the absence of several Hsp70 and Gpx transcripts in reference transcriptome used here. Furthermore, environmental stress might not necessarily lead to upregulation (see Hsp in Antarctic animals [84, 85]). Therefore, future studies should also aim to quantify gene expression under different environmental conditions using experimental approaches.

Conclusions

Stress responses are key for the survival of organisms, yet their role in adaptive parasite evolution remains poorly understood. The present study addresses this knowledge gap by analysing the stress response gene presence, copy number variation, and adaptive selection in 11 genomes of two genera of parasitic flatworms. We also resolved the phylogenetic relationships between several lineages of *Cichlidogyrus*, which prior studies using nuclear ribosomal and mitochondrial DNA markers struggled to do [18].

We detected several cases of copy number differences and positively selected gene sites, indicating that alterations in stress response pathways may be a relevant aspect of parasite and disease evolution. Specifically, we highlighted the potential role of mitochondrial protein import and behaviour in parasite of a host lineage under adaptive radiation, which has far-reaching consequences for the many parasite lineages that infect such host lineages (see [86]). Additionally, the absence of cytochrome P450 and kappa and sigma-class glutathione S-transferases in monogenean flatworms is reported, highlighting the unique adaptations monogenean flatworms might present, which warrants further studies into their functional evolution. Consequently, we encourage researchers to not only replicate our approach in other species-rich and functionally diverse lineages but also explore other molecular pathways that might determine adaptive potential and, therefore, the evolution of parasitic diseases.

Methods

Sample collection and DNA sequencing

To analyse a representative selection of the species diversity of Cichlidogyrus, we collected at least one species from eight of the recently reported 11 main lineages [18] (Additional File 6). Fish hosts were collected as part of previous studies [58, 59] with the help of local fisherfolk and the gills were subsequently extracted from the fishes and stored in absolute ethanol. Individual flatworms were collected from the gills using entomological needles and morphologically identified to species level based on Kuchta [87], Pariselle and Euzet [88], and Vanhove et al. [89]. Morphological identification is considered a reliable method to differentiate species of both target taxa as highlighted by previous studies that highlighted consistency with DNA barcoding approaches [90, 91]. Total genomic DNA extraction was applied on species pools and followed a recently published protocol [60]. For whole-genome amplification, we used the Illustra Ready-To-Go Genomiphi V3 DNA amplification kit (Cytiva, United Kingdom), which was applied to two samples (see Additional File 6). Library preparation (Illumina TruSeq Nano, 350 bp target insert size) and shortread sequencing (151 bp, paired end, HiSeq X) were outsourced to Macrogen Korea (Seoul, South Korea) or Macrogen Europe (Amsterdam, The Netherlands) (for estimated coverages of genomic read-pools, see Additional File 6). Furthermore, we accessed whole-genome sequencing read pools of one species of Cichlidogyrus and two of Kapentagyrus from previous mitogenomic studies [59, 60] (SRA accessions: https://identifiers.org/ insdc.sra:SRX11523770, https://identifiers.org/insdc.sra: SRX18894998, https://identifiers.org/insdc.sra:SRX18

894989). We also attempted to use previously published genome short reads of different species of *Cichlidogyrus*/ *Scutogyrus* [92, 93]. However, the coverage of these reads proved to be too low for capturing gene sequences targeted here. Raw sequence reads were trimmed through *Trimmomatic* v0.39 [94] using a sliding window approach (settings: *SLIDINGWINDOW:4:28 HEADCROP:5 MIN-LEN:100 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:True*). The quality of filtered reads was checked in *FastQC* v0.11.8 [95]. Raw Illumina reads generated as part of this study were submitted to the NCBI Sequencing Read Archive (SRA) (accession numbers: https://identifiers. org/insdc.sra:SRR31400484–https://identifiers.org/insdc. sra:SRR31400491) under BioProject accession https:// identifiers.org/ncbi/bioproject:PRJNA1186934.

Gene selection for species tree estimation and stress response genes

We used single copy ortholog genes to infer the phylogenetic backbone (species tree) of the parasite species. To date, nuclear ribosomal genes (28S and 18S rDNA and the internal transcribed spacers) and mitochondrial genes have been used as phylogenetic markers across most animal taxa as their multi-copy nature increases the likelihood of successful amplification of the target loci [96]. However, both rDNA and mitochondrial DNA have high substitution rates in flatworms [97] that may cause sequence alignment errors, which in return create noise in phylogenetic analyses (e.g. long-branch attraction of rapidly evolving lineages). Genome data provide a large number of alternative markers for phylogenetic inference, e.g. large datasets composed of single-copy orthologous genes can be used to resolve phylogenetic relationships. Recently, this single-copy ortholog approach has been adopted for a neodermatan phylogeny [20] including several monogenean species-one of them belonging to Scutogyrus Pariselle & Euzet, 1995, a nested lineage within Cichlidogyrus. The resulting phylogeny was based on 137 and 479 orthologous groups of proteins inferred from the BUSCO v4 [29] and OMA v2.6.0 [98] pipelines, respectively. As S. longicornis is a member of one of the target lineages of this study, we used the previously assembled single copy protein sequences [20] as bait sequences for the downstream putative protein sequence assembly (Fig. 2a, single-copy orthologs). The assembly of the BUSCO and OMA genes through the pipeline Hyb-*Piper* v2.0 [99] may also include genes other than singlecopy genes in genomes of species of Cichlidogyrus, but this was considered unlikely because these conserved genes have previously been demonstrated to have only a single copy in S. longicornis and other flatworm lineages [20].

For the stress genes, we focused on 12 gene families from three different functional groups: antioxidant enzymes (Cyp, Gpx, Mgst, cGst, Gstk, Prx, Sod, and Tgr), heat shock proteins (Hsp10, Hsp40, Hsp60, Hsp70, and Hsp90), and foraging orthologs (see Fig. 2 and Additional File 3 for abbreviations). For cGst, we targeted all known gene families and classes [100]. For the other antioxidant enzymes, we included the main groups previously reported from parasitic flatworms [55]. For Hsp, we included the gene families investigated extensively in flatworms in a recent study [9]. An illustration of the ortholog selection process described below can be found in Fig. 6. As performance of exon bait capture (see section below: DNA sequence assembly) decreases with phylogenetic distance, we aimed to use bait sequences from species that are as closely related as possible to the target taxa. However, nucleotide and amino acid sequences of the targeted genes have rarely been explicitly targeted in genome assemblies of monogenean flatworms (except for the Hsp70 subfamily in Gyrodactylus salaris Malmberg, 1957 [11]). Therefore, we compiled a set of previously published protein sequences of other flatworm groups (Additional File 3). As no for orthologs have been reported in flatworms in previous studies, we included protein sequences of for isoforms of D. melanogaster. All bait sequences were used to detect putative protein orthologs in a draft annotation of a genome assembly of *Cichlidogyrus cas*uarinus (Additional File 3; Fig. 2a, stress genes) (WGS accession JBJKFK00000000) using BLAST+v2.13.0 [101] (Fig. 1b). These data were selected for being the only available genome annotation of a species of Cich*lidogyrus*, which is currently being optimised as part of a separate study. For gene families, for which we did not detect orthologs in C. casuarinus, we used the initial non-monogenean search sequences as baits (Additional File 3) and also verified the potential absence of these genes through a BLAST search of published monogenean genomes in NCBI GenBank [11, 32, 56, 57]. The assembly and annotation process of the draft genome of C. casuarinus is detailed in Additional File 2. For a list of accession numbers and the respective protein IDs in C. casuarinus, see Additional File 3. A total of 48 putative protein sequences of C. casuarinus with query coverages above 90% were considered highly likely to represent genuine orthologs and included as bait in downstream analyses. In case of multiple hits, all sequences were included as baits as they might present potential duplications. Following the selection procedure for the and stress genes, we used 479 (OMA) and 137 (BUSCO) bait sequences for single-copy orthologs and 48 for the stress genes.

DNA sequence assembly and paralog filtering

Target genes in individual samples were identified through an in-situ exon bait capture approach as implemented in the pipeline HybPiper v2.0 [99]. HybPiper uses a bait file to map the trimmed paired-end and unpaired reads of all analysed species against the bait sequences (Fig. 2c). The bait file for the single-copy orthologs were compiled as detailed above. The bait file for the stress genes was compiled through the sequences of C. casuarinus. For the target files of both the stress genes and the single-copy orthologs, we used the protein sequences rather than the nucleotides sequences as gene assemblies reportedly improve when using the former [99]. In HybPiper, we used default parameters for the assembly, contig alignment and stitching process, and flagging of potential paralogs (the contig with highest read depth is selected as 'main hit') and chimeric sequences, but we chose DIAMOND v2.0.15 for the rapid alignments of sequencing reads [48].

The pooling approach during the sample acquisition means that paralogs flagged by *HybPiper* may be both orthologs in the sampled population or 'real' paralogs. To exclude the former and to remove contaminant sequences (e.g. host DNA, microorganisms associated with host gills or flatworm parasites), we manually curated sequences by performing five filtering steps (Fig. 2d), one step for the single-copy orthologs (i) and four steps for the stress genes (ii–iv):

- (i) We excluded any single-copy ortholog alignments for which paralogs were flagged in *HybPiper* to minimise the risks of accidentally including any contaminant sequences. We also excluded single-copy ortholog alignments for which sequences were not recovered from all 11 target species to minimise the impact of missing data on the species tree.
- (ii) We applied a *BLAST*+search to all assembled protein sequences of stress genes against the NCBI protein database to exclude contaminants. Best-hit sequences with>90% identity with non-flatworm sequences were excluded.
- (iii) We performed phylogenetic analyses with all potential paralogous sequences of the stress genes in the *HybPiper* output (see details below) and identified groups of sequences with a lowest common ancestor (LCA) (i.e. the common ancestor furthest away from the root) [102] as orthologous groups. If these groups of sequences (i.e. potential orthologous groups) included sequences from all target species, they were immediately assembled into gene alignments for downstream analyses, using the main hits assembled by *HybPiper*. Groups for which genes were only recovered for some target

species were subjected to a second run in *HybPiper* to detect gene orthologs. A second phylogenetic analysis was applied to the resulting protein alignments combined with the sequences assembled in the first *HybPiper* run from the same gene family. The sequences from the second run were again

- filtered through the LCA approach. In cases for which *HybPiper* did not provide a main hit, the paralog sequence with the maximum read depth was retained in each orthologous group, supplanting the missing hit. This selection by read depth might create a bias but is done to minimise the effects of sequencing errors at lower read depths.
- (iv) We excluded orthologous groups of stress genes detected in less than three target species to further minimise the effects of variation in the sampled populations of each DNA read pool.
- (v) We checked whether alignments of stress gene models not targeted with the bait sequences (paralogs suggested by *HybPiper*) represented fragments of other assembled gene models using *BLAST* + and excluded such truncated sequences.

Following the filtering steps, we inferred functional descriptions and gene ontology (GO) classes for each orthologous group using *PANNZER2* [103] (Fig. 2e). These GO terms were only considered reliable if the annotations were assigned to orthologs of three or more species. We also verified the presence of the gene sequences through a *BLAST*+ search (*tblastn*) against a recently published transcriptome annotation of *S. lon-gicornis* [51], interpreting sequence identities and query coverage > 95% as confirmatory of transcription (Fig. 2e). Alignments of exon sequences can be accessed at Zenodo (https://doi.org/10.5281/zenodo.14236484).

Phylogenetic analyses

We performed phylogenetic analyses for three different sequence datasets: species trees (based on single-copy orthologs, 277 BUSCO and 86 OMA loci, respectively), gene family trees for sequence filtering and paralog identification (e.g. Gst and Hsp70), and gene trees (for each of the 48 groups of orthologs for the targeted gene families). Phylogenetic analyses of the nucleotide sequences were performed under the maximum likelihood (ML) criterion. Sequences of all genes were aligned and trimmed with codon awareness through MACSE v2.06 using the options trimNonHomologousFragments, alignSequences, and trimAlignment [104, 105]. For the gene family trees, we did not trim the alignments as many informative sites would be removed due to high divergence between genes of the same gene family. Codon substitution models were selected by gene through *ModelFinder* in *IQ-Tree* [106].

We estimated tree topologies through *IQ-Tree* v2.2.0 [107, 108], estimating branch support through ultrafast bootstraps [109] and Shimodaira-Hasegawa–like approximate likelihood ratio tests (SH-aLRT) [110] with 10,000 replicates. We considered nodes with an ultrafast bootstrap value (UF-Boot) \geq 95 and an SH-aLRT statistic \geq 80 as well-supported. Phylogenetic trees were visualised through *ggtree* v3.6.2 [111, 112] in *R* v4.3.2 [113].

Comparison of gene vs. species tree topologies

We employed two approaches to assess topological differences of the species tree, the gene family trees (oxidative stress, heat shock, aquaporin, and *foraging* genes), and the single gene trees: visual inspection and multidimensional scaling. First, we assessed the phylogenies of the gene families qualitatively through visual inspection to detect potential deletions and/or duplications of genes among the parasite species investigated here. We followed an approach based on the LCA (see above), where nodes are either considered speciation or duplication nodes [102] according to parsimony criteria (reconciliation). Groups of single sequences from different species that formed monophyletic clades were considered orthologous.

In the second step, we tested whether the tree topologies of each orthologous group of stress response genes deviated from the species trees of Cichlidogyrus using multi-dimensional scaling (MDS) based on Kendall-Colijn distances of the trees [114]. This analysis was performed to infer whether the evolution of the target genes showed concordance with the evolution of the lineage (species tree). To detect topological differences of gene trees regarding Cichlidogyrus, sequences of Kapentagyrus were dropped from these trees using the function *drop.tip* in the *R* package *ape* v5.7–1 [115]. All gene trees with missing taxa (less than nine species of Cich*lidogyrus*) were also excluded as MDS requires complete datasets. Finally, we performed the MDS analysis through the package treespace v1.1.4.2 [116] on all 43 remaining gene trees.

Positive selection of gene sites

To detect signals of adaptive evolution in the stress response genes, we analysed patterns of synonymous and non-synonymous changes (d_N/d_S) in each of the 48 sequence alignments. We tested (I) whether stress genes of *Cichlidogyrus* and *Kapentagyrus* present gene sites that show patterns of positive selection $(d_N/d_S>1)$ and (II) if positively selected sites were more prevalent in certain clades/species (branch-site tests). Specifically, we tested if stress genes of species of *Cichlidogyrus* outside of Lake Tanganyika have undergone positive selection (IIa) and if stress genes of East African species of *Cichlidogyrus*

infecting hosts that have undergone adaptive radiation (Lake Tanganyika clade, Fig. 3) have done so (IIb). These codon analyses were performed in *CODEML* in *PAML* v4.10 [53] using the OMA-based species tree (for its higher node support, see Fig. 3 and Additional File 1) and the average nucleotide frequencies at the three codon positions (*CodonFreq* = 2).

For (I), we performed pairwise likelihood ratio tests (LRTs) between models (M) with heterogeneous d_N/d_S across sites: M1a vs. M0 (rate heterogeneity), M2a vs. M1a (positive selection, test 1), and M8 vs. M7 (positive selection, test 2). The rate heterogeneity test serves to test variability in selective pressure across sites. The other two tests serve to detect positive selection. If both tests were positive, we considered this confirmation of strongly positively selected sites. If only M8 vs. M7 turned out positive, we interpreted this result as a sign of the presence of weakly (yet significantly) positively selected sites as the second test is less stringent [117].

For (II), we performed pairwise LRTs between models with heterogeneous d_N/d_S across sites and clades. In accordance with *PAML* guidelines [117], the clades with hypothesised positively selected sites were defined as foreground branches, and two models were applied to each case: M1 (site model M2a, see above) and M0 (site model M2a, but with d_N/d_S fixed to 0). We tested two selected clades based on our hypotheses: *Cichlidogyrus* without Lake Tanganyika (IIa) and *Cichlidogyrus*–Lake Tanganyika only (IIb). If LRTs were positive, we considered positively selected sites to be present in the tested clades. For all tests (I, IIa, IIb), we also performed a robustness analysis by varying the *CodonFreq* parameter (0, 1, 2, 3) and assessing differences in the outcome [53].

Abbreviations

Aqp	Aquaporin Dia dia grantzia
BIP	Binding protein
BUSCO	Senchmarking Universal Single-Copy Orthologs
CGIVIP	Cyclic guanosine monophosphate
CGST	Cytosolic glutathione S-transferase
Сур	Cytochrome P450
d _N /d _S	Ratio of non-synonymous to synonymous substitutions
DNAJA1	DnaJ heat shock protein family (Hsp40) member A1
For	Foraging
GO term	Gene ontology term
Gpx	Glutathione peroxidase
Gst	Glutathione S-transferase
Gsta	Glutathione S-transferase alpha-class
Gstk	Glutathione S-transferase kappa-class
Gstm	Glutathione S-transferase mu-class
Gsto	Glutathione S-transferase omega-class
Gstp	Glutathione S-transferase pi-class
Gsts	Glutathione S-transferase sigma-class
Gstz	Glutathione S-transferase zeta-class
Hsp10	Heat shock 10 kDa protein
Hsp40	Heat shock 40 kDa protein
Hsp60	Heat shock 60 kDa protein
Hsp70	Heat shock 70 kDa protein
Hsp90	Heat shock 90 kDa protein
Hyou1	Hypoxia upregulated protein 1

LCA	Lowest common ancestor
LRT	Likelihood ratio test
MDS	Multidimensional scaling
Mgst	Microsomal glutathione S-transferase
OMA	Orthologous MAtrix
PKG	CGMP-dependent protein kinase or protein kinase G
Prx	Peroxiredoxin
SH-aLRT	Shimodaira-Hasegawa approximate likelihood ratio test
Sod	Superoxide dismutase
Tgr	Thioredoxin glutathione reductase
UF-Boot	Ultrafast bootstrap

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s12915-024-02091-w.

Additional file 1. Species tree of *Cichlidogyrus* and *Kapentagyrus* inferred from 68 single-copy orthologs based on a subset of genes selected by Caña-Bozada et al. (2023), who used the BUSCO pipeline. Support values: ultrafast bootstraps (UF-Boot)/Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT) (see the 'Methods' section), asterisks (*) indicate support below threshold (UF-Boot \leq 95, SH-aLRT \leq 80). Abbreviations: Ccas-*Cichlidogyrus casuarinus*, Ccir-*C. cirratus*, Chal-*C. halli*, Ckap-C. sp. 'kapembwa', Cscl- *C. sclerosus*, Cthu- *C. thurstonae*, Ctil- *C. tilapiae*, Czam- *C. zambezensis*, Slon- *Scutogyrus longicornis*, Klim-*Kapentagyrus limnotrissae*, Ktan- *K. tanganicanus*. Scale bar: estimated number of substitutions per site.

Additional file 2. Draft assembly and annotation of genome of *Cichlidogyrus casuarinus*.

Additional file 3. Overview of sequences used for bait capture of target gene groups and baited sequences (hits) with annotations. Heat shock protein sequences can be accessed using the protein IDs [9] at *UniProt* [43]. Annotations were inferred from PANNZER2 [77] (see Fig. 2).

Additional file 4. Number of paralogs by parasite species flagged for each of the protein bait sequences of *Cichlidogyrus casuarinus* (see Supplementary File S2) for stress gene assembly. Gene names reflect bait sequences from draft annotation (*C. casuarinus*) and not the final assembled genes. Abbreviations: Ccas- *Cichlidogyrus casuarinus*, Ccir- *C. cirratus*, Chal- *C. halli*, Ckap- C. sp. 'kapembwa', Cscl- *C. sclerosus*, Cthu- *C. thurstonae*, Ctil- *C. tilapiae*, Czam- *C. zambezensis*, Slon- *Scutogyrus longicornis*, Klim- *Kap*-entagyrus limnotrissae, Ktan- *K. tanganicanus*.

Additional file 5. Stress gene tree topologies produced under the maximum likelihood criterion. Abbreviations: Ccas–*Cichlidogyrus casuarinus*, Ccir–*C. cirratus*, Chal–*C. halli*, Ckap–C. sp. 'kapembwa', Cscl–*C. sclerosus*, Cthu–*C. thurstonae*, Ctil–*C. tilapiae*, Czam–*C. zambezensis*, Slon–*Scutogyrus longicornis*, Klim–*Kapentagyrus limnotrissae*, Ktan–*K. tanganicanus*.

Additional file 6. Sampling data of collected specimens including reference for sampling campaigns and published whole-genome sequencing data.

Additional file 7. Extended version of Fig. 4. Detected putative stress gene orthologs (protein sequences) including species coverage (cyan = *Cich-lidogyrus*, red = *Kapentagyrus*) (a), gene ontology (GO) terms (black = term applies) (b), presence in transcriptome annotation (blue = present) (c), and hypothesis testing of different models for detecting positively selected gene sites (I, IIa, IIb) (d) with * indicating P < 0.05 for test results and the colour scale indicating the likelihood ratio test statistics (LRT) (see the 'Methods' section). Rows and columns of the GO heatmap are automatically sorted through Euclidean distances as implemented in ComplexHeatmap.

Acknowledgements

We dedicate this manuscript to Auguste Chocha Manda (1967–2023) and Charles F. Bilong Bilong (1958–2024) for their crucial support during the collection of the host specimens collected in the present study and their life-long contribution to ichthyological research. We appreciate the help of everyone involved in the field work and sampling procedure including Fidel Muterezi Bukinga, Clément Kalombo Kabalika, Jiří Vorel, Stephan Koblmüller, Holger Zimmermann, and several fishermen. We thank Stephan Koblmüller for his contribution to the genome annotation of C. casuarinus. We thank Natascha Steffanie (Hasselt University) for technical support in the laboratory. We thank Leona Milec (PXL University College, Hasselt University, Nord University) and Geert-Jan Bex (Hasselt University) for assistance with the use of the high-performance computing cluster. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Authors' contributions

AJCL conceptualised the study under supervision of NK and MPMV. AJCL, MG, MB, HB, ARBM, GKK, SN, and NK contributed to the collection of the host specimens. MPMV and TH supervised the sampling campaign. MG, AJCL, and NK performed analyses in the laboratory including microscopical examination of the fish gills, the collection and identification of the parasite specimens, and DNA extraction and genome pre-amplification. CH and PR produced the genome assembly and annotation used as a baseline. AJCL performed all statistical analyses and produced all graphs with input from NK and LB. AJCL and NK drafted the manuscript with substantial input from MPMV, LB, MB, HB, ARBM, MG, CH, TH, GKK, SN, PR, and KS. All authors read and approved the final manuscript.

Funding

The field work was partially funded by Czech Science Foundation (GAČR) project GA19-13573S. Molecular work was funded through research grant 1513419N of the Research Foundation—Flanders (FWO). AJCL (BOF19OWB02), MPMV (BOF20TT06, BOF21INCENT09), MG (BOF17NI02), and NK (BOF21PD01) were funded by the Special Research Fund of Hasselt University. AJCL is currently funded by the Research Foundation—Flanders (FWO) (12ABP24N) and NK by the Belgian Science Policy Office (BELSPO) (AfroWetMaP project). This research was funded in whole, or in part, by the Austrian Science Fund (FWF) [https://doi.org/10.55776/P32691]. Parts of the resources and services used in this work were provided by the VSC (Flemish Supercomputer Centre), funded by the Research Foundation—Flanders (FWO) and the Flemish Government, as well as by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic and the ELIXIR-CZ project (ID:90255), part of the international ELIXIR infrastructure. Part of the research leading to the results presented in this publication was carried out with infrastructure funded by the European Marine Biological Research Centre (EMBRC) Belgium, Research Foundation—Flanders (FWO) project GOH3817N.

Data availability

Raw Illumina reads were submitted to the NCBI Sequencing Read Archive (SRA) (accession numbers: https://identifiers.org/insdc.sra:SRR31400484– https://identifiers.org/insdc.sra:SRR31400491) under BioProject accession https://identifiers.org/ncbi/bioproject:PRINA1186934. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession https://identifiers.org/ncbigene:JBJKFK000000000. The version described in this paper is version https://identifiers.org/ncbigene:JBJKFK000000000. Phylogenetic trees and alignments were deposited in Zenodo (https://www. doi.org/10.5281/zenodo.14236484).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Sciences, Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, UHasselt – Hasselt University, Diepenbeek, Belgium. ²Systems Ecology and Resource Management Research Unit (SERM), Université Libre de Bruxelles-ULB, Brussels, Belgium. ³Natural History Museum, University of Oslo, Oslo, Norway. ⁴Department of Biological Sciences, University of Botswana, Gaborone, Botswana. ⁵Department of Biological Sciences, University of Ngaoundéré, Ngaoundéré, Cameroon. ⁶Department of Management of Fisheries and Aquatic Ecosystems, Institute of Fisheries, University of Douala, Douala, Cameroon. ⁷Department of Biology, Evolutionary Ecology Group – EVECO, University of Antwerp, Antwerp, Belgium. ⁸Institute of Biology, University of Graz, Graz, Austria. ⁹Department of Biology, Royal Museum for Central Africa, Tervuren, Belgium. ¹⁰Unité de Recherche en Biodiversité Et Exploitation Durable Des Zones Humides (BEZHU), Faculté Des Sciences Agronomiques, Université de Lubumbashi, Lubumbashi, Democratic Republic of the Congo. ¹¹Aquatic and Terrestrial Ecology, Operational Directorate Natural Environment, Royal Belgian Institute for Natural Sciences, Brussels, Belgium.

Received: 24 May 2024 Accepted: 9 December 2024 Published online: 13 January 2025

References

- 1. Bijlsma R, Loeschcke V. Environmental stress, adaptation and evolution: an overview. J Evol Biol. 2005;18:744–9.
- 2. Hotez PJ, Aksoy S, Brindley PJ, Kamhawi S. What constitutes a neglected tropical disease? PLoS Negl Trop Dis. 2020;14: e0008001.
- Ismail TNST, Kassim NFA, Rahman AA, Yahya K, Webb CE. Day biting habits of mosquitoes associated with mangrove forests in Kedah. Malaysia Trop Med Infect Dis. 2018;3:77.
- 4. Santi AMM, Ribeiro JM, Reis-Cunha JL, Burle-Caldas G de A, Santos IFM, Silva PA, et al. Disruption of multiple copies of the prostaglandin F2alpha synthase gene affects oxidative stress response and infectivity in *Trypanosoma cruzi*. PLoS Negl Trop Dis. 2022;16:e0010845.
- 5. Huyse T, Poulin R, Théron A. Speciation in parasites: a population genetics approach. Trends Parasitol. 2005;21:469–75.
- D'Bastiani E, Campiaõ KM, Boeger WA, Araújo SBL. The role of ecological opportunity in shaping host-parasite networks. Parasitology. 2020;147:1452–60.
- Hector TE, Gehman A-LM, King KC. Infection burdens and virulence under heat stress: ecological and evolutionary considerations. Philos Trans R Soc Lond B Biol Sci. 2023;378:20220018.
- Huyse T, Boon NAM, Van den Broeck F, Mbow M, Chaturvedi A, Meurs L, et al. Evolutionary epidemiology of schistosomiasis: linking parasite genetics with disease phenotype in humans. Int J Parasitol. 2018;48:107–15.
- Aguoru NA, Kirk RS, Walker AJ. Molecular insights into the heat shock proteins of the human parasitic blood fluke *Schistosoma mansoni*. Parasit Vectors. 2022;15:365.
- Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013;496:57–63.
- Hahn C, Fromm B, Bachmann L. Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparastic Neodermata. Genome Biol Evol. 2014;6:1105–17.
- Calla B, Noble K, Johnson RM, Walden KKO, Schuler MA, Robertson HM, et al. Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: birth, death and adaptation. Mol Ecol. 2017;26:6021–35.
- Zhang W, Yu H, Lv Y, Bushley KE, Wickham JD, Gao S, et al. Gene family expansion of pinewood nematode to detoxify its host defence chemicals. Mol Ecol. 2020;29:940–55.
- Wang B, Liang X, Gleason ML, Zhang R, Sun G. Comparative genomics of *Botryosphaeria dothidea* and *B. kuwatsukai*, causal agents of apple ring rot, reveals both species expansion of pathogenicity-related genes and variations in virulence gene content during speciation. IMA Fungus. 2018;9:243–57.
- 15. Anreiter I, Sokolowski MB. The *foraging* gene and its behavioral effects: pleiotropy and plasticity. Annu Rev Genet. 2019;53:373–92.
- Kmentová N, Cruz-Laufer AJ, Pariselle A, Smeets K, Artois T, Vanhove MPM. Dactylogyridae 2022: a meta-analysis of phylogenetic studies and generic diagnoses of parasitic flatworms using published genetic and morphological data. Int J Parasitol. 2022;52:427–57.
- Burress ED. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. Hydrobiologia. 2014;748:7–27.

- Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steenberge M, et al. Somewhere I belong: phylogeny and morphological evolution in a species-rich lineage of ectoparasitic flatworms infecting cichlid fishes. Cladistics. 2022;38:465–512.
- 19. Wilson AB, Teugels GG, Meyer A. Marine incursion: the freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. PLoS ONE. 2008;3: e1979.
- Caña-Bozada V, Robinson MW, Hernández-Mena DI, Morales-Serna FN. Exploring evolutionary relationships within Neodermata using putative orthologous groups of proteins, with emphasis on peptidases. Trop Med Infect Dis. 2023;8:59.
- Leeming SJ, Hahn C, Koblmüller S, McAllister CT, Vanhove MPM, Kmentová N. Amended diagnosis, mitochondrial genome, and phylogenetic position of *Sphyranura euryceae* (Neodermata, Monogenea, Polystomatidae), a parasite of the Oklahoma salamander. Parasite. 2023;30:27.
- 22. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.
- 23. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinformatics. 2020;70: e102.
- 24. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19:1117–23.
- Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014;24:1384–95.
- 27. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014;30:31–7.
- Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018;34:i142–50.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021;38:4647–54.
- Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Res. 2017;6:1287.
- 31. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics. 2012;28:2520–2.
- Vorel J, Kmentová N, Hahn C, Bureš P, Kašný M. An insight into the functional genomics and species classification of *Eudiplozoon nipponicum* (Monogenea, Diplozoidae), a haematophagous parasite of the common carp *Cyprinus carpio*. BMC Genomics. 2023;24:363.
- Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23:1061–7.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUS-TUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 2006;34 WEB. SERV. ISS:W435–9.
- 36. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
- Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2023. http:// www.repeatmasker.org/. Accessed 31 Jan 2023.
- Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33:6494–506.
- Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31.
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50:D20–6.
- Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite

 a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2–10.

- Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.
- 44. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterPro-Scan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
- 46. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46:D624–32.
- Käll L, Krogh A, Sonnhammer ELL. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res. 2007;35(SUPPL. 2):W429–32.
- Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-oflife scale using DIAMOND. Nat Methods. 2021;18:366–8.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309-14.
- Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34:2115–22.
- Caña-Bozada V, Morales-Serna FN, Fajer-Ávila EJ, Llera-Herrera R. De novo transcriptome assembly and identification of G-Protein-Coupled-Receptors (GPCRs) in two species of monogenean parasites of fish. Parasite. 2022;29:51.
- 52. Talavera G, Vila R. What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. BMC Evol Biol. 2011;11:315.
- Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
- 54. Nelson DR. Cytochrome P450 diversity in the tree of life. Biochim Biophys Acta Proteins Proteom. 2018;1866:141–54.
- Martínez-González J de J, Guevara-Flores A, Del Arenal Mena IP. Evolutionary adaptations of parasitic flatworms to different oxygen tensions. Antioxidants. 2022;11:1102.
- Baeza JA, González MT. A first look at the 'repeatome' of *Benedenia humboldti*, a major pathogen in yellowtail aquaculture: repetitive element characterization, nuclear rRNA operon assembly, and microsatellite discovery. Mar Genomics. 2021;58: 100848.
- Konczal M, Przesmycka KJ, Mohammed RS, Phillips KP, Camara F, Chmielewski S, et al. Gene duplications, divergence and recombination shape adaptive evolution of the fish ectoparasite *Gyrodactylus bullatarudis*. Mol Ecol. 2020;29:1494–507.
- Geraerts M, Huyse T, Barson M, Bassirou H, Bilong Bilong CF, Bitja-Nyom AR, et al. Mosaic or melting pot: the use of monogeneans as a biological tag and magnifying glass to discriminate introduced populations of Nile tilapia in sub-Saharan Africa. Genomics. 2022;114: 110328.
- Kmentová N, Thys K, Hahn C, Vorel J, Koblmüller S, Van Steenberge M, et al. Comparative population mitogenomics of fish parasites reveals contrasting geographic pattern in the pelagic zone of Lake Tanganyika. Authorea. 2023. https://doi.org/10.22541/au.167274604.41657501/v1.
- Kmentová N, Hahn C, Koblmüller S, Zimmermann H, Vorel J, Artois T, et al. Contrasting host-parasite population structure: morphology and mitogenomics of a parasitic flatworm on pelagic deepwater cichlid fishes from Lake Tanganyika. Biology (Basel). 2021;10:797.
- Saveria T, Halbach A, Erdmann R, Volkmer-Engert R, Landgraf C, Rottensteiner H, et al. Conservation of PEX19-binding motifs required for protein targeting to mammalian peroxisomal and trypanosome glycosomal membranes. Eukaryot Cell. 2007;6:1439–49.
- 62. Wisedpanichkij R, Grams R, Chaijaroenkul W, Na-Bangchang K. Confutation of the existence of sequence-conserved cytochrome P450 enzymes in *Plasmodium falciparum*. Acta Trop. 2011;119:19–22.
- Pazdzior R, Yang Z, Mesbahuddin MS, Yee J, van der Est A, Rafferty S. Low reduction potential cytochrome b₃ isotypes of *Giardia intestinalis*. Exp Parasitol. 2015;157:197–201.
- Shaheen S, Barrett KF, Subramanian S, Arnold SLM, Laureanti JA, Myler PJ, et al. Solution structure for an *Encephalitozoon cuniculi* adrenodoxinlike protein in the oxidized state. Protein Sci. 2020;29:809–17.

- 65. Pakharukova MY, Vavilin VA, Sripa B, Laha T, Brindley PJ, Mordvinov VA. Functional analysis of the unique cytochrome P450 of the liver fluke *Opisthorchis felineus*. PLoS Negl Trop Dis. 2015;9: e0004258.
- Skern-Mauritzen R, Malde K, Eichner C, Dondrup M, Furmanek T, Besnier F, et al. The salmon louse genome: copepod features and parasitic adaptations. Genomics. 2021;113:3666–80.
- Iriarte A, Arbildi P, La-Rocca S, Musto H, Fernández V. Identification of novel glutathione transferases in *Echinococcus granulosus*. An evolutionary perspective Acta Trop. 2012;123:208–16.
- Žárský V, Tachezy J. Evolutionary loss of peroxisomes not limited to parasites. Biol Direct. 2015;10:74.
- Ni Z-X, Cui J-M, Zhang N-Z, Fu B-Q. Structural and evolutionary divergence of aquaporins in parasites. Mol Med Rep. 2017;15:3943–8.
- Tian R, Geng Y, Yang Y, Seim I, Yang G. Oxidative stress drives divergent evolution of the glutathione peroxidase (GPX) gene family in mammals. Integr Zool. 2021;16:696–711.
- Pirozhkova D, Katokhin A. *Opisthorchis felineus* genes differentially expressed under praziquantel shed light on the nature of tegument disruption and indicate the adaptive role of cGMP-dependent protein kinase. Parasitol Res. 2020;119:2695–702.
- 72. Vanhove MPM, Pariselle A, Van Steenberge M, Raeymaekers JAM, Hablützel PI, Gillardin C, et al. Hidden biodiversity in an ancient lake: phylogenetic congruence between Lake Tanganyika tropheine cichlids and their monogenean flatworm parasites. Sci Rep. 2015;5:13669.
- 73. Cruz-Laufer AJ, Artois T, Koblmüller S, Pariselle A, Smeets K, Van Steenberge M, et al. Explosive networking: the role of adaptive host radiations and ecological opportunity in a species-rich host–parasite assembly. Ecol Lett. 2022;25:1795–812.
- Jordan A, Taborsky B, Taborsky M. Cichlids as a model system for studying social behaviour and evolution. In: Abata ME, Noakes DLG, editors. The behavior, ecology and evolution of cichlid fishes. Dordrecht, The Netherlands: Springer Nature; 2021. p. 587–635.
- 75. Brand JN. Support for a radiation of free-living flatworms in the African Great Lakes region and the description of five new *Macrostomum* species. Front Zool. 2023;20:31.
- Markov GV, Baskaran P, Sommer RJ. The same or not the same: lineagespecific gene expansions and homology relationships in multigene families in nematodes. J Mol Evol. 2015;80:18–36.
- 77. Makino T, Kawata M. Invasive invertebrates associated with highly duplicated gene content. Mol Ecol. 2019;28:1652–63.
- Lin R, Yang M, Yao B. The phylogenetic and evolutionary analyses of detoxification gene families in *Aphidinae*[sic] species. PLoS ONE. 2022;17:e0263462.
- Gouin A, Bretaudeau A, Nam K, Gimenez S, Aury JM, Duvic B, et al. Two genomes of highly polyphagous lepidopteran pests (*Spodop-tera frugiperda*, Noctuidae) with different host-plant ranges. Sci Rep. 2017;7:11816.
- Drini S, Criscuolo A, Lechat P, Imamura H, Skalický T, Rachidi N, et al. Species- and strain-specific adaptation of the HSP70 super family in pathogenic trypanosomatids. Genome Biol Evol. 2016;8:1980–95.
- Stanley TR, Kim Guisbert KS, Perez SM, Oneka M, Kernin I, Higgins NR, et al. Stress response gene family expansions correlate with invasive potential in teleost fish. J Exp Biol. 2022;225:jeb243263.
- Kofler R, Pandey RV, Schlötterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics. 2011;27:3435–6.
- Neumann S, Ziv E, Lantner F, Schlechter I. Regulation of HSP70 gene expression during the life cycle of the parasitic helminth *Schistosoma* mansoni. Eur J Biochem. 1993;212:589–96.
- Clark MS, Fraser KPP, Peck LS. Lack of an HSP70 heat shock response in two Antarctic marine invertebrates. Polar Biol. 2008;31:1059–65.
- Bilyk KT, Vargas-Chacoff L, Cheng CHC. Evolution in chronic cold: varied loss of cellular response to heat in Antarctic notothenioid fish. BMC Evol Biol. 2018;18:143.
- Vanhove MPM, Kmentová N, Faes C, Fernandes JMO, Hahn C, Hens N, et al. Understanding the influence of host radiation on symbiont speciation through parasites of species flocks. Cold Spring Harb Perspect Biol. 2024; https://doi.org/10.1101/cshperspect.a041450.
- Kuchta R, Basson L, Cook C, Fiala I, Bartošová-Sojková P, Řehulková E. A systematic survey of the parasites of freshwater fishes in Africa. In: Scholz T, Vanhove MPM, Smit N, Jayasundera Z, Gelnar M, editors. A

Guide to the parasites of African freshwater fishes AbcTaxa , vol. 18. Brussels: CEBioS, Royal Belgian Institute of Natural Sciences; 2018. p. 137–402.

- Pariselle A, Euzet L. Systematic revision of dactylogyridean parasites (Monogenea) from cichlid fishes in Africa, the Levant and Madagascar. Zoosystema. 2009;31:849–98.
- Vanhove MPM, Hermans R, Artois T, Kmentová N. From the Atlantic coast to Lake Tanganyika: gill-infecting flatworms of freshwater pellonuline clupeid fishes in West and Central Africa, with description of eleven new species and key to *Kapentagyrus* (Monogenea, Dactylogyridae). Animals. 2021;11:3578.
- Geraerts M, Huyse T, Vanhove MPM, Artois T. A new species of *Cichlidogyrus* Paperna, 1960 (Platyhelminthes: Monogenea: Dactylogyridae) infecting tilapias in Lake Kariba (Zimbabwe), with a discussion on its phylogenetic position. Syst Biodivers. 2022;20:2143594.
- Kmentová N, Van Steenberge M, Raeymaekers JAM, Koblmüller S, Hablützel PI, Bukinga FM, et al. Monogenean parasites of sardines in Lake Tanganyika: diversity, origin and intraspecific variability. Contributions to Zoology. 2018;87:105–32.
- Vanhove MPM, Briscoe AG, Jorissen MWP, Littlewood DTJ, Huyse T. The first next-generation sequencing approach to the mitochondrial phylogeny of African monogenean parasites (Platyhelminthes: Gyrodactylidae and Dactylogyridae). BMC Genomics. 2018;19:520.
- Caña-Bozada V, Llera-Herrera R, Fajer-Ávila EJ, Morales-Serna FN. Mitochondrial genome of *Scutogyrus longicornis* (Monogenea: Dactylogyridea), a parasite of Nile tilapia *Oreochromis niloticus*. Parasitol Int. 2021;81: 102281.
- 94. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
- Andrews S. FastQC. 2018. http://www.bioinformatics.babraham.ac.uk/ projects/fastqc. Accessed 8 May 2023.
- Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B. A plea for standardized nuclear markers in metazoan DNA taxonomy. Trends Ecol Evol. 2020;35:336–45.
- Vanhove MPM, Tessens B, Schoelinck C, Jondelius U, Littlewood DTJ, Artois T, et al. Problematic barcoding in flatworms: a case-study on monogeneans and rhabdocoels (Platyhelminthes). Zookeys. 2013;365:355–79.
- Altenhoff AM, Levy J, Zarowiecki M, Tomiczek B, Vesztrocy AW, Dalquen DA, et al. OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome Res. 2019;29:1152–63.
- Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, et al. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Appl Plant Sci. 2016;4:1600016.
- Bae YA, Kim JG, Kong Y. Phylogenetic characterization of *Clonorchis* sinensis proteins homologous to the sigma-class glutathione transferase and their differential expression profiles. Mol Biochem Parasitol. 2016;206:46–55.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
- Swenson KM, El-Mabrouk N. Gene trees and species trees: irreconcilable differences. BMC Bioinformatics. 2012;13(Suppl 19):S15.
- Törönen P, Holm L. PANNZER—a practical tool for protein function prediction. Protein Sci. 2022;31:118–28.
- Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol Biol Evol. 2018;35:2582–4.
- Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: multiple alignment of coding SEquences accounting for frameshifts and stop codons. PLoS One. 2011;6:e22594.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le Vinh S. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
- 111. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36.
- 112. Yu G, Lam TT-Y, Zhu H, Guan Y. Two methods for mapping and visualizing associated data on phylogeny using ggtree. Mol Biol Evol. 2018;35:3041–3.
- 113. R Core Team. R: a language and environment for statistical computing. 2023. https://www.R-project.org/. Accessed 11 Jun 2023.
- 114. Hillis DM, Heath TA, St. John K. Analysis and visualization of tree space. Syst Biol. 2005;54:471–82.
- 115. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8.
- Jombart T, Kendall M, Almagro-Garcia J, Colijn C. treespace: statistical exploration of landscapes of phylogenetic trees. Mol Ecol Resour. 2017;17:1385–92.
- 117. Álvarez-Carretero S, Kapli P, Yang Z. Beginner's guide on the use of PAML to detect positive selection. Mol Biol Evol. 2023;40:msad041.
- 118. Moons T, Kmentová N, Pariselle A, Artois T, Bert W, Vanhove MPM, et al. All quiet on the western front? The evolutionary history of monogeneans (Dactylogyridae: *Cichlidogyrus, Onchobdella*) infecting a West and Central African tribe of cichlid fishes (Chromidotilapiini). Parasite. 2023;30:25.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.