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

Armando J. Cruz-Laufer[1,2], Maarten P. M. Vanhove[1], Lutz Bachmann[3], Maxwell Barson[4,5,6], Hassan Bassirou[7], Arnold R. Bitja Nyom[7,8], Mare Geraerts[1,9], Christoph Hahn[10], Tine Huyse[11], Gyrhaiss Kapepula Kasembele[1,12], Samuel Njom[7], Philipp Resl[10], Karen Smeets[1], Nikol Kmentová[1,13]

1 UHasselt – Hasselt University, Faculty of Sciences, Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, Diepenbeek, Belgium.

2 Systems Ecology and Resource Management Research Unit (SERM), Université Libre de Bruxelles-ULB, Brussels, Belgium.

3 Natural History Museum, University of Oslo, Oslo, Norway.

4 Department of Biological Sciences, University of Zimbabwe, Harare, Zimbabwe.

5 Department of Biological Sciences, University of Botswana, Gaborone, Botswana.

6 Lake Kariba Research Station, University of Zimbabwe, Kariba, Zimbabwe.

7 Department of Biological Sciences, University of Ngaoundéré, Ngaoundéré, Cameroon.

8 Department of Management of Fisheries and Aquatic Ecosystems, Institute of Fisheries, University of Douala, Douala, Cameroon.

9 Evolutionary Ecology Group – EVECO, Department of Biology, University of Antwerp, Antwerp, Belgium

10 Institute of Biology, University of Graz, Graz, Austria.

11 Department of Biology, Royal Museum for Central Africa, Tervuren, Belgium.

12 Unité de Recherche en Biodiversité et Exploitation durable des Zones Humides (BEZHU), Faculté des Sciences Agronomiques, Université de Lubumbashi, Lubumbashi, Democratic Republic of the Congo.

13 Aquatic and Terrestrial Ecology, Operational Directorate Natural Environment, Royal Belgian Institute for Natural Sciences, Brussels, Belgium.

CORRESPONDING AUTHORS

Christoph Hahn, christoph.hahn@uni-graz.at

Armando J. Cruz-Laufer, armando.cruzlaufer@uhasselt.be

Abstract

Stress responses are key for parasite survival and, consequently, also the evolutionary success of these organisms. Despite this importance, our understanding of the molecular pathways dealing with environmental stressors remains limited for parasitic animals. Here, we targeted the molecular pathways dealing with environmental stressors and comparatively investigated antioxidant, heat shock, osmoregulatory, and behaviour-related genes (*foraging*) in two parasitic flatworm lineages with contrasting species and ecological diversity, *Cichlidogyrus* and *Kapentagyrus* (Platyhelminthes: Monogenea), through whole-genome sequencing of 11 species. Using an *in silico* exon bait capture approach, 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—consistent with a potential adaptation linked to a co-radiation of parasites and hosts. Accordingly, this study potentially identifies the first molecular function linked to a flatworm radiation. 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.

Key words

Comparative genomics, positive selection, Monogenea, heat shock proteins, oxidative stress

Introduction

Evolutionary theory predicts that a species' ability to maintain homeostasis against environmental stressors fundamentally affects its adaptive potential (Bijlsma & Loeschcke 2005). Although metazoan parasites cause many neglected tropical diseases in humans (Hotez et al. 2020), they are often overlooked as groups of pathogens. Research on parasite stress responses remains largely limited to few well-known human-infecting species, for the purpose of drug development (Huyse et al. 2018; Aguoru et al. 2022), or focuses on macroevolutionary adaptations of major parasite clades (Tsai et al. 2013; Hahn et al. 2014). Stress response pathways are rarely comparatively analysed below the level of major lineages, e.g. flatworm classes or insect orders, although new parasite species are formed at this level. As they deal with host defences, effective stress responses can increase fitness of individuals and populations (microevolution) and, therefore, permit species to expand host repertoires and geographical ranges, which may give rise to new parasite species and diseases (macroevolution) (Huyse et al. 2005). Stress responses can determine infectivity and virulence of parasites (Ismail et al. 2018; Santi et al. 2022). Moreover, recent studies suggest that human activity promotes the rise of emerging infectious diseases as environmental disturbance creates ecological opportunities for parasite species with an adequate adaptive potential (D'Bastiani et al. 2020; Hector et al. 2023). Here, we address this knowledge gap: how do 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 (Calla et al. 2017), nematodes (Zhang et al. 2020), and fungi (Wang et al. 2018), including antioxidant enzymes dealing with reactive oxygen species (oxidative stress response), heat shock proteins assisting with protein folding, and aquaporins dealing with osmotic stress. Animals may also respond to environmental stressors through behavioural changes. The *foraging* (*for*) gene of *Drosophila melanogaster* and other species are among the best-known examples of genes determining behavioural differences (Anreiter & Sokolowski 2019).

Here, we explore the diversity and adaptive evolution of genes encoding antioxidant enzymes, heat shock proteins, aquaporins, and *for* orthologs in the closely related flatworm ectoparasite lineages, *Cichlidogyrus* and *Kapentagyrus* (Kmentová et al. 2022). The two groups infect host lineages (African

cichlid vs. freshwater clupeid fishes) with contrasting species richness (Fig. 1) and ecological diversity (Burress 2014; Wilson et al. 2008), with one subclade (*Cichlidogyrus* spp. infecting East African cichlids) infecting a host lineage that has undergone multiple rapid diversification events (adaptive radiations) in its recent evolutionary history (Cruz-Laufer et al. 2022), coinciding with a high parasite species richness (Fig. 1). Therefore, we hypothesise duplication and positive selection in stress genes of parasites infecting cichlids (*Cichlidogyrus*), compared to *Kapentagyrus*, which is specialised on a relatively species-poor, ecologically conserved host lineage (i.e. African freshwater clupeids all occupy pelagic environments of rivers and lakes) (Wilson et al. 2008). With whole-genome sequencing data of 11 species, our study provides the largest genomic dataset from a single flatworm genus 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* (Caña-Bozada et al. 2023) as bait (Fig. 2a, c; single-copy orthologs). After alignment filtering (Fig. 2d), we retained 277 (OMA tree, Fig. 3) and 86 (BUSCO tree, Supplementary Fig. S1) 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 (see Supplementary File S2; Fig. 2a: single-copy orthologs), 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, Supplementary Table S3). A majority (63%) of the sequences matched with the transcriptome data of *S. longicornis* (Caña-Bozada et al. 2022) (> 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.

For most targeted stress genes, we found only a single copy per sequencing read pool (Supplementary Fig. S4). 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* (Supplementary Table S3) and data produced in this study, nor other published monogenean genomes (see Methods) using *tblastn*. *Gpx*, *Gstm*, *Prx*, and all heat shock protein orthologs except *Hsp10* were flagged by *HybPiper* for potential paralogs (Supplementary Fig. S4). 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 Supplementary Table S3.

For the *Hsp70* family, a multitude of paralogs were flagged for all but two bait sequences (Supplementary Fig. S4). 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). Three *Hsp70* groups were assigned highly specific functions: the hypoxia up-regulated 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 long-branch attraction.

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. casuarinus*, 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. 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 (Supplementary File S5) 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).

Detection of positive selection

Positive selection is the process by which gene variants that provide a fitness benefit dominate a population over time (Yang 2007). 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_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 clade-specific 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), ie. 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. Here, we compared 48 putative stress genes between two metazoan parasite lineages (*Cichlidogyrus* and *Kapentagyrus*) infecting host lineages with contrasting levels of niche diversity. In a multi-species study system [i.e. more than three species, e.g. Aguoru et al. (2022)], we showed for the first time that stress responses are associated with the adaptive evolution of closely related parasite lineages. With nine species of *Cichlidogyrus*, we provide the most species-

rich set of whole-genome sequencing data from a flatworm genus to date [but see blood fluke genus *Schistosoma* with eight species (Ebbs et al. 2022)]. With 68/277 single-copy orthologs and 48 stress genes, this study is also the most extensive genomic analysis of monogenean flatworms to date. Furthermore, we resolved the relationships between several lineages of *Cichlidogyrus* through phylogenomic analyses, which prior studies recovering these lineages using nuclear ribosomal and mitochondrial DNA markers struggled to do (Cruz-Laufer et al. 2022).

We detected several unique stress response features in the targeted monogenean parasites. The absence of the cytochrome P450 gene family (*Cyp*) and glutathione *S*-transferase (*Gst*) sigma- (*Gsts*) and kappa-classes (*Gstk*) in monogenean genomes is remarkable. CYP enzymes are mainly involved in the (oxidative) metabolism of various endogenous and exogenous compounds; and in these reactions, they can cause oxidative stress. They are conserved across almost the entire tree of life (Nelson 2018). All GST members serve for cellular protection as detoxification enzymes, and *Gsts* and *Gstk* genes were previously reported from all flatworm genomes investigated so far (Martínez-González et al. 2022), except for monogenean species. Yet orthologs were not only absent in *Cichlidogyrus* and *Kapentagyrus*, but could also not be detected in previously published monogenean genome data (Baeza & González 2021; Hahn et al. 2014; Vorel et al. 2023; Konczal et al. 2020). The recovery and assembly of single-copy orthologs without any problem indicates the absence of *Cyp*, *Gsts*, and *Gstk* orthologs in our data is real and not caused by low sequencing depth (see also estimated coverages, Supplementary Table S6). Hence, functional monogenean *Cyp*, *Gsts*, and *Gstk* are either highly divergent from those of other organisms, and thus not annotated [as reported for a proposed peroxin gene in a parasitic protozoan (Saveria et al. 2007)], or they are indeed absent in monogeneans. The loss of *Cyp* has only been reported from protozoan parasites (Wisedpanichkij et al. 2011; Pazdzior et al. 2015; Shaheen et al. 2020). In other parasitic flatworms, i.e. flukes and tapeworms (Martínez-González et al. 2022), only a single variant was reported, unlike the multiple copies present in most organisms (Nelson 2018). However, CYP still fulfils vital functions in flatworm parasites (Pakharukova et al. 2015). The gene family has also been reduced in parasitic crustaceans, e.g. salmon lice have the lowest known *Cyp* copy number of any arthropod(Skern-Mauritzen et al. 2021), which may reflect an evolutionary trend of *Cyp* contractions in metazoan parasites. For *Gst*s and *Gstk*, no losses have been reported in parasitic flatworms. However, the absence of *Gsto* genes in tapeworms might indicate a functional replacement by another *Gst* class, but due to lack of phylogenetic evidence, the authors remained cautious about interpreting these results (Iriarte et al. 2012). Evolutionary loss of genes and gene functions has also repeatedly been observed elsewhere

in parasites, e.g. peroxisomal functions in parasitic protozoans, flatworms, roundworms (Žárský & Tachezy 2015), and crustaceans (Skern-Mauritzen et al. 2021). These losses have in part been attributed to the r-selected traits of many parasites (high fecundity, few resources for individual offspring) leading them to stress response mechanisms (Žárský & Tachezy 2015). Accordingly, *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*) (Supplementary Table S3). Monogeneans, thus, differ from other parasitic flatworms, with tapeworms and flukes presenting one *Gpx*, three *Prx*, 12 *cGst*, and one to three *Aqp* (Martínez-González et al. 2022; Ni et al. 2017) 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 (Tian et al. 2021). The discussed examples for contractions/expansions of *Cyp*, *Gpx*, *Prx*, *cGst*, and *Aqp* in parasites, indicate that these gene families are key for 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 (Anreiter & Sokolowski 2019). Although fecundity, infection intensity (Huyse et al. 2018), and drug resistance (Pirozhkova & Katokhin 2020) in flatworms have been associated with the cGMP-dependent 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 (Vanhove et al. 2015; Cruz-Laufer et al. 2022). 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 Tanganyika 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 (Jordan et al. 2021). In the present study, we provide the first evidence that their parasites belonging to *Cichlidogyrus* also present unique 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) (Bateman et al. 2023) 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 (Vanhove et al. 2015; Cruz-Laufer et al. 2022). Our present findings indicate that this diversification might be linked to functional adaptations, which would make the cichlid-*Cichlidogyrus* system the first adaptive radiation of flatworms with a specified genetic adaptation [but see morphological evidence in free-living flatworms from the same lake (Brand 2023)]. 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). Hence, instances of gene duplication/loss of stress response genes in *Cichlidogyrus* and *Kapentagyrus* might reflect their contrasting host ecology and evolutionary history, specifically the adaptive potential of the parasites. We identified two potential instances of gene duplication/loss. Species of *Kapentagyrus* lack a gene copy of *Gstm* and *Hsp70* and all copies of *Gsto*, which indicates either gene duplication in *Cichlidogyrus* or gene loss in *Kapentagyrus*. If the assumption of gene duplications in *Cichlidogyrus* was correct, it might suggest a higher potential to adapt to stressful conditions. Expansions of the *Gst* gene superfamily in closely related organisms were previously also observed in free-living nematodes (Markov et al. 2015). In fungi, gene family expansion, e.g. of *Cyp*, had been linked to habitat/host diversity (Wang et al. 2018) and, in several invertebrate animal groups, the proportion of gene duplication in genomes has been linked to their invasive potential (Makino & Kawata 2019), i.e. their ability to adapt to new environments. In metazoan parasites, prior studies detected gene family expansions in tapeworms (Tsai et al. 2013) and aphids (Lin et al. 2022), but only rarely are these expansions linked to concrete environmental stressors because of unknown gene functions. Notable exceptions include the plant-pathogenic moths of the *Spodoptera frugiperda* species

complex, where polyphagous representatives possess an expanded set of detoxification genes compared to their specialist relatives (Gouin et al. 2017), and the plant-parasitic nematode *Bursaphelenchus xylophilus*, which upregulates members of an expanded groups of *Gst* to detoxify xenobiotics in pinewood (Zhang et al. 2020). Expansions of *Hsp70* among closely related lineages were previously interpreted as adaptations to environmental stressors, e.g. in invasive fishes (Stanley et al. 2022). Amongst parasitic animals, *Hsp70* expansions have been reported for tapeworms (Tsai et al. 2013) and trypanosomatid protozoans (Drini et al. 2016). These expansions were hypothesised to be expressed only 'under certain conditions' (Tsai et al. 2013) or were loosely associated with geographical and environmental gradients (Drini et al. 2016). Similar to *Gst* genes, *Hsp70* copy numbers alone provide no definitive evidence of environmental adaptation without detailed knowledge on gene functions. For instance, some cold-adapted Antarctic organisms have lost their inducible heat shock responses despite the presence of *Hsp70* sequences in their genomes (Bilyk et al. 2018; Clark et al. 2008). Nevertheless, *Cichlidogyrus* and *Kapentagyrus* provide the first evidence of divergent adaptations of stress response pathways in closely related flatworm lineages occupying similar functional niches (as freshwater epithelial-feeding gill parasites).

The positive selection of stress gene sites and gene duplications/losses provide new insights into molecular mechanisms underpinning the adaptive evolution of parasites, which contributes to an improved understanding of mechanisms of host selection and switching. Nonetheless, our study also has conceptual and technological limitations. First, previous studies indicate that copy number evolution can occur between closely related animal species (Lin et al. 2022) and even strains (Drini et al. 2016), but no such differences were detected here with confidence because the focus was on differences between lineages not species, an approach taken to avoid mistaking intraspecific allelic gene variants in each pooled DNA samples as paralogs (see Methods). Future studies might use variant calling pipelines [e.g. Kofler et al. (2011)] or optimise techniques to sequence genomes from individual specimens to address this problem. The latter approach has recently been successful with monogenean mitochondrial genomes (Geraerts et al. 2022a) and recent advances in sequencing technology for medical research means that whole-genome sequencing can be applied to single cells, albeit limited to model organisms (Alfieri et al. 2022). 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 presented some deviation 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 (Cruz-Laufer et al. 2022), biological invasions (Geraerts et al. 2022a), and speciation rates (Vanhove et al. 2015). Second, gene models reveal no information on expression patterns. Some genes are only expressed under certain conditions (i.e. inducible genes) and may not be represented in reference transcriptomes (Caña-Bozada et al. 2022), e.g. as evidenced for human-infecting flukes (Neumann et al. 1993). This might explain the absence of several *Hsp70* and *Gpx* transcripts in reference transcriptome used here. Furthermore, environmental stress might not necessarily lead to up-regulation [see *Hsp* in Antarctic animals (Clark et al. 2008; Bilyk et al. 2018)]. Therefore, future studies should also aim to quantify gene expression under different environmental conditions using experimental studies.

Stress responses are key for the survival of organisms, yet their role in adaptive parasite evolution remains poorly understood. A strong bias in parasite genomics towards few human-relevant pathogens results in a lack of DNA and RNA sequence data of closely related and functionally similar parasite lineages. The present study addresses this bias by analysing the stress response gene presence, copy number variation, and adaptive selection in 11 genomes of two genera of parasitic flatworms. Even though these data only represent a fraction of the known species from these lineages, we detected several cases of copy number differences and positively selected gene sites, confirming that alterations in stress response pathways are a relevant aspect of parasite and disease evolution. However, stress responses are not the only mechanisms that might determine the adaptive potential of parasites. Therefore, we encourage researchers to not only replicate our approach in other species-rich and functionally diverse lineages, but also to explore other molecular pathways that might determine adaptive potential and, therefore, the evolution of parasitic diseases.

MATERIAL AND 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 (Cruz-Laufer et al. 2022) (Supplementary Table S6). Fish hosts were collected as part of previous studies (Geraerts et al. 2022a; Kmentová et al. 2023) 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 (Geraerts et al. 2022b; Kmentová et al. 2018). Total genomic DNA extraction was applied on species pools and followed a recently published protocol (Kmentová et al. 2021). 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 Supplementary Table S6). Library preparation (Illumina TruSeq Nano, 350 bp target insert size) and short-read 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 Supplementary Table S6). Furthermore, we accessed whole-genome sequencing read pools of one species of *Cichlidogyrus* and two of *Kapentagyrus* from previous mitogenomic studies (Kmentová et al. 2023, 2021) (Accession BioProjects: PRJNA749051, XXXXX, XXXXX). We also attempted to use previously published genome short reads of different species of *Cichlidogyrus/Scutogyrus* (Vanhove et al. 2018; Caña-Bozada et al. 2021). However, the coverage of these reads proved to be too low for capture gene sequences targeted here. Raw sequence reads were trimmed through *Trimmomatic* v0.39(Bolger et al. 2014) using a sliding window approach (settings: *SLIDINGWINDOW:4:28 HEADCROP:5 MINLEN:100 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:True*). The quality of filtered reads was checked in *FastQC* v0.11.8(Andrews 2018). Raw Illumina reads generated as part of this study were submitted to the NCBI Sequencing Read Archive (SRA) (accession numbers: XXXXXXXXX-XX) under BioProject accession XXXXXXXXXX.

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 (Eberle et al.

2020). However, both rDNA and mitochondrial DNA have high substitution rates in flatworms(Vanhove et al. 2013) 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 (Caña-Bozada et al. 2023) including several monogenean species—one of them belonging to *Scutogyrus*, a nested lineage within *Cichlidogyrus*. The resulting phylogeny was based on 137 and 479 orthologous groups of proteins inferred from the *BUSCO* v4 (Manni et al. 2021) and *OMA* v2.6.0 (Altenhoff et al. 2019) 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 (Caña-Bozada et al. 2023) 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 *HybPiper* v2.0 (Johnson et al. 2016) may also include genes other than single-copy 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 (Caña-Bozada et al. 2023).

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 Supplementary Table S3 for abbreviations). For *cGst*, we targeted all known gene families and classes (Bae et al. 2016). For the other antioxidant enzymes, we included the main groups previously reported from parasitic flatworms (Martínez-González et al. 2022). For *Hsp*, we included the gene families investigated extensively in flatworms in a recent study (Aguoru et al. 2022). An illustration of the ortholog selection process described below can be found in Figure 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 targeted have rarely been explicitly targeted in genome assemblies of monogenean flatworms [except for the *Hsp70* subfamily in *Gyrodactylus salaris* (Hahn et al. 2014)]. Therefore, we compiled a set of previously published protein sequences of other flatworm groups (Supplementary Table S3). As no *for* orthologs have been reported in flatworms in previous studies, we included protein sequences of *for* isoforms of *D. melanogaster*. For

gene families, for which we did not detect orthologs in *C. casuarinus*, we used the initial non-monogenean search sequences as baits (Supplementary Table S3), and also verified the potential absence of these genes through a *BLAST* search of published monogenean genomes in NCBI Genbank (Konczal et al. 2020; Hahn et al. 2014; Vorel et al. 2023; Baeza & González 2021). All bait sequences were used to detect putative protein orthologs in a draft annotation of a genome assembly of *Cichlidogyrus casuarinus* (Supplementary Table S3; Fig. 2a, stress genes) using *BLAST*+ v2.13.0 (Camacho et al. 2009) (Fig. 1b). The assembly and annotation process of the draft genome of *C. casuarinus* is detailed in Supplementary File S2. For a list of accession numbers and the respective protein IDs in *C. casuarinus*, see Supplementary Table S3. 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 (Johnson et al. 2016). *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 (Johnson et al. 2016). 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 (Buchfink et al. 2021).

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) (Swenson & El-Mabrouk 2012) 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.

(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* (Törönen & Holm 2022) (Fig. 2e). These GO terms were only considered reliable and included in Table 2 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. longicornis* (Caña-Bozada et al. 2022), interpreting sequence identities and query coverage > 95% as confirmatory of transcription (Fig. 2e). We deposited assembled exon sequences in NCBI GenBank (XXXXXXXX–XXXXXXXX).

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* (Ranwez et al. 2018, 2011). 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* (Kalyaanamoorthy et al. 2017). We estimated tree topologies through *IQ-Tree* v2.2.0 (Minh et al. 2020; Nguyen et al. 2015), estimating branch support through ultrafast bootstraps (Hoang et al. 2018) and Shimodaira-Hasegawa–like approximate likelihood ratio tests (SH-aLRT) (Guindon et al. 2010) with 10,000 replicates. We considered nodes with an ultrafast bootstrap value (UF-Boot) ≥ 95 and an SH-aLRT statistic ≥ 80 as well-supported. Phylogenetic trees were visualised through *ggtree* v3.6.2 (Yu et al. 2017, 2018) in *R* v4.3.2 (R Core Team 2023).

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(Swenson & El-Mabrouk 2012) 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 (Hillis et al. 2005). 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 (Paradis & Schliep 2019). All gene trees with missing taxa (less than nine species of *Cichlidogyrus*) were also excluded as MDS requires complete datasets. Finally, we performed the MDS analysis through the package *treespace* v1.1.4.2 (Jombart et al. 2017) 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 (Yang 2007) using the OMA-based species tree (for its higher node support, see Fig. 3 and Supplementary Fig. S1) 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 (Álvarez-Carretero et al. 2023).

For (II), we performed pairwise LRTs between models with heterogeneous d_N/d_S across sites and clades. In accordance with *PAML* guidelines (Álvarez-Carretero et al. 2023), 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 (Yang 2007).

Data and resource availability

The assembled DNA sequence data are available in the GenBank Nucleotide Database at https://www.ncbi.nlm.nih.gov/genbank, accession numbers (XXXXXXX–XXXXXXX). Phylogenetic trees and data matrices were deposited in Zenodo (DOI: XXXXXXXX). Raw Illumina reads were submitted to the NCBI Sequencing Read Archive (SRA) (accession numbers: XXXXXXXXX-XX) under BioProject accession XXXXXXXXXX.

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. 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). NK is currently funded by the Belgian Science Policy Office (BELSPO) (AfroWetMap project). This research was funded in part by the Austrian Science Fund (FWF) [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. 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. 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.

Author 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.

Ethics declaration

The authors declare no competing interests.

References

Aguoru NA, Kirk RS, Walker AJ. 2022. Molecular insights into the heat shock proteins of the human parasitic blood fluke *Schistosoma mansoni*. *Parasit Vectors*. 15:365. https://www.doi.org/10.1186/s13071-022-05500-7.

Alfieri JM et al. 2022. A primer for single-cell sequencing in non-model organisms. *Genes (Basel)*. 13:380. https://www.doi.org/10.3390/genes13020380.

Altenhoff AM et al. 2019. OMA standalone: orthology inference among public and custom genomes and transcriptomes. *Genome Res*. 29:1152–1163. https://www.doi.org/10.1101/gr.243212.118.

Álvarez-Carretero S, Kapli P, Yang Z. 2023. Beginner's guide on the use of PAML to detect positive selection. *Mol Biol Evol*. 40:msad041. https://www.doi.org/10.1093/molbev/msad041.

Andrews S. 2018. FastQC. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed May 8, 2023).

Anreiter I, Sokolowski MB. 2019. The *foraging* gene and its behavioral effects: pleiotropy and plasticity. *Annu Rev Genet*. 53:373–392. https://www.doi.org/10.1146/annurev-genet-112618.

Bae YA, Kim JG, Kong Y. 2016. Phylogenetic characterization of *Clonorchis sinensis* proteins homologous to the sigma-class glutathione transferase and their differential expression profiles. *Mol Biochem Parasitol*. 206:46–55. https://www.doi.org/10.1016/j.molbiopara.2016.01.002.

Baeza JA, González MT. 2021. 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*. 58:100848. https://www.doi.org/10.1016/j.margen.2021.100848.

Bateman A et al. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 51:D523–D531. https://www.doi.org/10.1093/nar/gkac1052.

Bijlsma R, Loeschcke V. 2005. Environmental stress, adaptation and evolution: an overview Bijlsma, R & Loeschcke, V, editors. *J Evol Biol*. 18:744–749. https://www.doi.org/10.1111/j.1420-9101.2005.00962.x.

577 Bilyk KT, Vargas-Chacoff L, Cheng CHC. 2018. Evolution in chronic cold: varied loss of cellular
578 response to heat in Antarctic notothenioid fish. *BMC Evol Biol*. 18:143.
579 https://www.doi.org/10.1186/s12862-018-1254-6.

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

582 Brand JN. 2023. Support for a radiation of free-living flatworms in the African Great Lakes region
583 and the description of five new *Macrostomum* species. *Front Zool*. 20.
584 https://www.doi.org/10.1186/s12983-023-00509-9.

585 Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using
586 DIAMOND. *Nat Methods*. 18:366–368. https://www.doi.org/10.1038/s41592-021-01101-x.

587 Burress ED. 2014. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and
588 consequences. *Hydrobiologia*. 748:7–27. https://www.doi.org/10.1007/s10750-014-1960-z.

589 Calla B et al. 2017. Cytochrome P450 diversification and hostplant utilization patterns in specialist
590 and generalist moths: birth, death and adaptation. *Mol Ecol*. 26:6021–6035.
591 https://www.doi.org/10.1111/mec.14348.

592 Camacho C et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.
593 https://www.doi.org/10.1186/1471-2105-10-421.

594 Caña-Bozada V, Llera-Herrera R, Fajer-Ávila EJ, Morales-Serna FN. 2021. Mitochondrial genome of
595 *Scutogyrus longicornis* (Monogenea: Dactylogyridea), a parasite of Nile tilapia *Oreochromis*
596 *niloticus*. *Parasitol Int*. 81:102281. https://www.doi.org/10.1016/j.parint.2020.102281.

597 Caña-Bozada V, Morales-Serna FN, Fajer-Ávila EJ, Llera-Herrera R. 2022. De novo transcriptome
598 assembly and identification of G-Protein-Coupled-Receptors (GPCRs) in two species of monogenean
599 parasites of fish. *Parasite*. 29:51. https://www.doi.org/10.1051/parasite/2022052.

600 Caña-Bozada V, Robinson MW, Hernández-Mena DI, Morales-Serna FN. 2023. Exploring
601 evolutionary relationships within Neodermata using putative orthologous groups of proteins, with
602 emphasis on peptidases. *Trop Med Infect Dis*. 8:59.
603 https://www.doi.org/10.3390/tropicalmed8010059.

Clark MS, Fraser KPP, Peck LS. 2008. Lack of an HSP70 heat shock response in two Antarctic marine invertebrates. *Polar Biol*. 31:1059–1065. https://www.doi.org/10.1007/s00300-008-0447-7.

Cruz-Laufer AJ et al. 2022. Explosive networking: the role of adaptive host radiations and ecological opportunity in a species-rich host–parasite assembly. *Ecol Lett*. 25:1795–1812. https://www.doi.org/10.1111/ele.14059.

Cruz-Laufer AJ et al. 2022. Somewhere I belong: phylogeny and morphological evolution in a species-rich lineage of ectoparasitic flatworms infecting cichlid fishes. *Cladistics*. 38:465–512. https://www.doi.org/10.1111/CLA.12506.

D'Bastiani E, Campiaõ KM, Boeger WA, Araújo SBL. 2020. The role of ecological opportunity in shaping host-parasite networks. *Parasitology*. 147:1452–1460. https://www.doi.org/10.1017/S003118202000133X.

Drini S et al. 2016. Species- and strain-specific adaptation of the HSP70 super family in pathogenic trypanosomatids. *Genome Biol Evol*. 8:1980–1995. https://www.doi.org/10.1093/gbe/evw140.

Ebbs ET et al. 2022. Phylogenomics and diversification of the Schistosomatidae based on targeted sequence capture of ultra-conserved elements. *Pathogens*. 11:769. https://www.doi.org/10.3390/pathogens11070769.

Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B. 2020. A plea for standardized nuclear markers in metazoan DNA taxonomy. *Trends Ecol Evol*. 35:336–345. https://www.doi.org/10.1016/j.tree.2019.12.003.

Geraerts M et al. 2022a. 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*. 114:110328. https://www.doi.org/10.1016/j.ygeno.2022.110328.

Geraerts M, Huyse T, Vanhove MPM, Artois T. 2022b. 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*. 20:2143594. https://www.doi.org/10.1080/14772000.2022.2143594.

631 Gouin A et al. 2017. Two genomes of highly polyphagous lepidopteran pests (*Spodoptera*
632 *frugiperda*, Noctuidae) with different host-plant ranges. *Sci Rep*. 7:11816.
633 https://www.doi.org/10.1038/s41598-017-10461-4.

634 Guindon S et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies:
635 assessing the performance of PhyML 3.0. *Syst Biol*. 59:307–321.
636 https://www.doi.org/10.1093/sysbio/syq010.

637 Hahn C, Fromm B, Bachmann L. 2014. Comparative genomics of flatworms (Platyhelminthes)
638 reveals shared genomic features of ecto- and endoparastic Neodermata. *Genome Biol Evol*.
639 6:1105–1117. https://www.doi.org/10.1093/gbe/evu078.

640 Hector TE, Gehman A-LM, King KC. 2023. Infection burdens and virulence under heat stress:
641 ecological and evolutionary considerations. *Philos Trans R Soc Lond B Biol Sci*. 378:20220018.
642 https://www.doi.org/10.1098/rstb.2022.0018.

643 Hillis DM, Heath TA, St. John K. 2005. Analysis and visualization of tree space. *Syst Biol*. 54:471–
644 482. https://www.doi.org/10.1080/10635150590946961.

645 Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le Vinh S. 2018. UFBoot2: improving the
646 ultrafast bootstrap approximation. *Mol Biol Evol*. 35:518–522.
647 https://www.doi.org/10.1093/molbev/msx281.

648 Hotez PJ, Aksoy S, Brindley PJ, Kamhawi S. 2020. What constitutes a neglected tropical disease?
649 *PLoS Negl Trop Dis*. 14:e0008001. https://www.doi.org/10.1371/JOURNAL.PNTD.0008001.

650 Huyse T et al. 2018. Evolutionary epidemiology of schistosomiasis: linking parasite genetics with
651 disease phenotype in humans. *Int J Parasitol*. 48:107–115.
652 https://www.doi.org/10.1016/j.ijpara.2017.07.010.

653 Huyse T, Poulin R, Théron A. 2005. Speciation in parasites: a population genetics approach. *Trends*
654 *Parasitol*. 21:469–475. https://www.doi.org/10.1016/j.pt.2005.08.009.

655 Iriarte A, Arbildi P, La-Rocca S, Musto H, Fernández V. 2012. Identification of novel glutathione
656 transferases in *Echinococcus granulosus*. An evolutionary perspective. *Acta Trop*. 123:208–216.
657 https://www.doi.org/10.1016/j.actatropica.2012.05.010.

Ismail TNST, Kassim NFA, Rahman AA, Yahya K, Webb CE. 2018. Day biting habits of mosquitoes associated with mangrove forests in Kedah, Malaysia. *Trop Med Infect Dis*. 3:77. https://www.doi.org/10.3390/tropicalmed3030077.

Johnson MG et al. 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Appl Plant Sci*. 4:1600016. https://www.doi.org/10.3732/apps.1600016.

Jombart T, Kendall M, Almagro-Garcia J, Colijn C. 2017. treespace: statistical exploration of landscapes of phylogenetic trees. *Mol Ecol Resour*. 17:1385–1392. https://www.doi.org/10.1111/1755-0998.12676.

Jordan A, Taborsky B, Taborsky M. 2021. Cichlids as a model system for studying social behaviour and evolution. In: The behavior, ecology and evolution of cichlid fishes. Abata, ME & Noakes, DLG, editors. Springer Nature: Dordrecht, The Netherlands pp. 587–635. https://www.doi.org/10.1007/978-94-024-2080-7_16.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14:587–589. https://www.doi.org/10.1038/nmeth.4285.

Kmentová N et al. 2023. Comparative population mitogenomics of fish parasites reveals contrasting geographic pattern in the pelagic zone of Lake Tanganyika [PREPRINT]. *Authorea*. https://www.doi.org/10.22541/au.167274604.41657501/v1.

Kmentová N et al. 2021. Contrasting host-parasite population structure: morphology and mitogenomics of a parasitic flatworm on pelagic deepwater cichlid fishes from Lake Tanganyika. *Biology (Basel)*. 10:797. https://www.doi.org/10.3390/biology10080797.

Kmentová N et al. 2022. Dactylogyridae 2022: a meta-analysis of phylogenetic studies and generic diagnoses of parasitic flatworms using published genetic and morphological data. *Int J Parasitol*. 52:427–457. https://www.doi.org/10.1016/j.ijpara.2022.01.003.

Kmentová N et al. 2018. Monogenean parasites of sardines in Lake Tanganyika: diversity, origin and intraspecific variability. *Contributions to Zoology*. 87:105–132. https://www.doi.org/10.1163/18759866-08702004.

686 Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between
687 populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*. 27:3435–3436.
688 https://www.doi.org/10.1093/bioinformatics/btr589.

689 Konczal M et al. 2020. Gene duplications, divergence and recombination shape adaptive evolution
690 of the fish ectoparasite *Gyrodactylus bullatarudis*. *Mol Ecol*. 29:1494–1507.
691 https://www.doi.org/10.1111/mec.15421.

692 Lin R, Yang M, Yao B. 2022. The phylogenetic and evolutionary analyses of detoxification gene
693 families in *Aphidinae*[sic] species. *PLoS One*. 17:e0263462.
694 https://www.doi.org/10.1371/journal.pone.0263462.

695 Makino T, Kawata M. 2019. Invasive invertebrates associated with highly duplicated gene content.
696 *Mol Ecol*. 28:1652–1663. https://www.doi.org/10.1111/mec.15019.

697 Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and
698 streamlined workflows along with broader and deeper phylogenetic coverage for scoring of
699 eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 38:4647–4654.
700 https://www.doi.org/10.1093/molbev/msab199.

701 Markov G V., Baskaran P, Sommer RJ. 2015. The same or not the same: lineage-specific gene
702 expansions and homology relationships in multigene families in nematodes. *J Mol Evol*. 80:18–36.
703 https://www.doi.org/10.1007/s00239-014-9651-y.

704 Martínez-González J de J, Guevara-Flores A, Del Arenal Mena IP. 2022. Evolutionary adaptations of
705 parasitic flatworms to different oxygen tensions. *Antioxidants*. 11:1102.
706 https://www.doi.org/10.3390/antiox11061102.

707 Minh BQ et al. 2020. IQ-TREE 2: new Models and efficient methods for phylogenetic inference in
708 the genomic era. *Mol Biol Evol*. 37:1530–1534. https://www.doi.org/10.1093/molbev/msaa015.

709 Moons T et al. 2023. All quiet on the western front? The evolutionary history of monogeneans
710 (Dactylogyridae: *Cichlidogyrus*, *Onchobdella*) infecting a West and Central African tribe of cichlid
711 fishes (Chromidotilapiini). *Parasite*. 30:25.

712 Nelson DR. 2018. Cytochrome P450 diversity in the tree of life. *Biochim Biophys Acta Proteins*
713 *Proteom*. 1866:141–154. https://www.doi.org/10.1016/j.bbapap.2017.05.003.

714 Neumann S, Ziv E, Lantner F, Schlechter I. 1993. Regulation of HSP70 gene expression during the
715 life cycle of the parasitic helminth *Schistosoma mansoni*. *Eur J Biochem*. 212:589–596.
716 https://www.doi.org/10.1111/j.1432-1033.1993.tb17697.x.

717 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic
718 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32:268–274.
719 https://www.doi.org/10.1093/molbev/msu300.

720 Ni Z-X, Cui J-M, Zhang N-Z, Fu B-Q. 2017. Structural and evolutionary divergence of aquaporins in
721 parasites. *Mol Med Rep*. 15:3943–3948. https://www.doi.org/10.3892/mmr.2017.6505.

722 Pakharukova MY et al. 2015. Functional analysis of the unique cytochrome P450 of the liver fluke
723 *Opisthorchis felineus*. *PLoS Negl Trop Dis*. 9:e0004258.
724 https://www.doi.org/10.1371/journal.pntd.0004258.

725 Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary
726 analyses in R. *Bioinformatics*. 35:526–528. https://www.doi.org/10.1093/bioinformatics/bty633.

727 Pazdzior R et al. 2015. Low reduction potential cytochrome b_5 isotypes of *Giardia intestinalis*. *Exp*
728 *Parasitol*. 157:197–201. https://www.doi.org/10.1016/j.exppara.2015.08.004.

729 Pirozhkova D, Katokhin A. 2020. *Opisthorchis felineus* genes differentially expressed under
730 praziquantel shed light on the nature of tegument disruption and indicate the adaptive role of
731 cGMP-dependent protein kinase. *Parasitol Res*. 119:2695–2702.
732 https://www.doi.org/10.1007/s00436-020-06764-7.

733 R Core Team. 2023. R: a language and environment for statistical computing. https://www.R-
734 project.org/ (Accessed June 10, 2023).

735 Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: toolkit for the
736 alignment of coding sequences accounting for frameshifts and stop codons. *Mol Biol Evol*. 35:2582–
737 2584. https://www.doi.org/10.1093/molbev/msy159.

738 Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: multiple alignment of coding SEquences
739 accounting for frameshifts and stop codons. *PLoS One*. 6.
740 https://www.doi.org/10.1371/journal.pone.0022594.

741 Santi AMM et al. 2022. Disruption of multiple copies of the Prostaglandin F2alpha synthase gene
742 affects oxidative stress response and infectivity in *Trypanosoma cruzi*. *PLoS Negl Trop Dis*.
743 16:e0010845. https://www.doi.org/10.1371/JOURNAL.PNTD.0010845.

744 Saveria T et al. 2007. Conservation of PEX19-binding motifs required for protein targeting to
745 mammalian peroxisomal and trypanosome glycosomal membranes. *Eukaryot Cell*. 6:1439–1449.
746 https://www.doi.org/10.1128/EC.00084-07.

747 Shaheen S et al. 2020. Solution structure for an *Encephalitozoon cuniculi* adrenodoxin-like protein
748 in the oxidized state. *Protein Science*. 29:809–817. https://www.doi.org/10.1002/pro.3818.

749 Skern-Mauritzen R et al. 2021. The salmon louse genome: copepod features and parasitic
750 adaptations. *Genomics*. 113:3666–3680. https://www.doi.org/10.1016/j.ygeno.2021.08.002.

751 Stanley TR et al. 2022. Stress response gene family expansions correlate with invasive potential in
752 teleost fish. *J Exp Biol*. 225:jeb243263. https://www.doi.org/10.1242/jeb.243263.

753 Swenson KM, El-Mabrouk N. 2012. Gene trees and species trees: irreconcilable differences. *BMC*
754 *Bioinformatics*. 13(Suppl 19):S25. https://www.doi.org/10.1186/1471-2105-13-s19-s15.

755 Tian R, Geng Y, Yang Y, Seim I, Yang G. 2021. Oxidative stress drives divergent evolution of the
756 glutathione peroxidase (GPX) gene family in mammals. *Integr Zool*. 16:696–711.
757 https://www.doi.org/10.1111/1749-4877.12521.

758 Törönen P, Holm L. 2022. PANNZER—A practical tool for protein function prediction. *Protein*
759 *Science*. 31:118–128. https://www.doi.org/10.1002/pro.4193.

760 Tsai IJ et al. 2013. The genomes of four tapeworm species reveal adaptations to parasitism.
761 *Nature*. 496:57–63. https://www.doi.org/10.1038/nature12031.

762 Vanhove MPM et al. 2015. Hidden biodiversity in an ancient lake: phylogenetic congruence between
763 Lake Tanganyika tropheine cichlids and their monogenean flatworm parasites. *Sci Rep*. 5:13669.
764 https://www.doi.org/10.1038/srep13669.

765 Vanhove MPM et al. 2013. Problematic barcoding in flatworms: a case-study on monogeneans and
766 rhabdocoels (Platyhelminthes). *Zookeys*. 365:355–379.
767 https://www.doi.org/10.3897/zookeys.365.5776.

768 Vanhove MPM, Briscoe AG, Jorissen MWP, Littlewood DTJ, Huyse T. 2018. The first next-generation
769 sequencing approach to the mitochondrial phylogeny of African monogenean parasites
770 (Platyhelminthes: Gyrodactylidae and Dactylogyridae). *BMC Genomics*. 19:520.
771 https://www.doi.org/10.1186/s12864-018-4893-5.

772 Vanhove MPM, Hermans R, Artois T, Kmentová N. 2021. From the Atlantic coast to Lake
773 Tanganyika: gill-infecting flatworms of freshwater pellonuline clupeid fishes in West and Central
774 Africa, with description of eleven new species and key to *Kapentagyrus* (Monogenea,
775 Dactylogyridae). *Animals*. 11:3578. https://www.doi.org/10.3390/ani11123578.

776 Vorel J, Kmentová N, Hahn C, Bureš P, Kašný M. 2023. An insight into the functional genomics and
777 species classification of *Eudiplozoon nipponicum* (Monogenea, Diplozoidae), a haematophagous
778 parasite of the common carp *Cyprinus carpio*. *BMC Genomics*. 24:363.
779 https://www.doi.org/10.1186/s12864-023-09461-8.

780 Wang B, Liang X, Gleason ML, Zhang R, Sun G. 2018. Comparative genomics of *Botryosphaeria*
781 *dothidea* and *B. kuwatsukai*, causal agents of apple ring rot, reveals both species expansion of
782 pathogenicity-related genes and variations in virulence gene content during speciation. *IMA*
783 *Fungus*. 9:243–257. https://www.doi.org/10.5598/imafungus.2018.09.02.02.

784 Wilson AB, Teugels GG, Meyer A. 2008. Marine incursion: the freshwater herring of Lake
785 Tanganyika are the product of a marine invasion into West Africa. *PLoS One*. 3:e1979.
786 https://www.doi.org/10.1371/journal.pone.0001979.

787 Wisedpanichkij R, Grams R, Chaijaroenkul W, Na-Bangchang K. 2011. Confutation of the existence
788 of sequence-conserved cytochrome P450 enzymes in *Plasmodium falciparum*. *Acta Trop*. 119:19–
789 22. https://www.doi.org/10.1016/j.actatropica.2011.03.006.

790 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
791 https://www.doi.org/10.1093/molbev/msm088.

792 Yu G, Lam TT-Y, Zhu H, Guan Y. 2018. Two methods for mapping and visualizing associated data on
793 phylogeny using ggtree. *Mol Biol Evol*. 35:3041–3043.
794 https://www.doi.org/10.1093/molbev/msy194.

795 Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for visualization and
796 annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol*
797 *Evol*. 8:28–36. https://www.doi.org/10.1111/2041-210X.12628.

798 Žárský V, Tachezy J. 2015. Evolutionary loss of peroxisomes - not limited to parasites. *Biol Direct*.
799 10:74. https://www.doi.org/10.1186/s13062-015-0101-6.

800 Zhang W et al. 2020. Gene family expansion of pinewood nematode to detoxify its host defence
801 chemicals. *Mol Ecol*. 29:940–955. https://www.doi.org/10.1111/mec.15378.

802

FIGURE LEGENDS

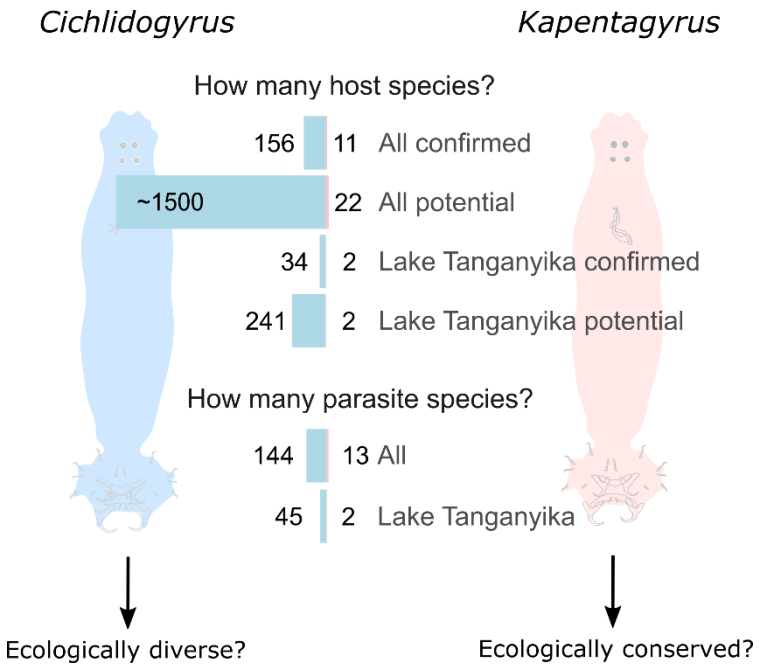

Figure 1. The two flatworm parasite lineages *Cichlidogyrus* and *Kapentagyrus* differ substantially in species richness and host diversity. Species of *Cichlidogyrus* infect the gills of the hyperdiverse African cichlid fishes that include the adaptive radiations of Lake Tanganyika in East Africa (Moons et al. 2023). 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 (Vanhove et al. 2021). Based on these differences, we hypothesise that stress responses of *Cichlidogyrus* have adapted to this ecological diversity.

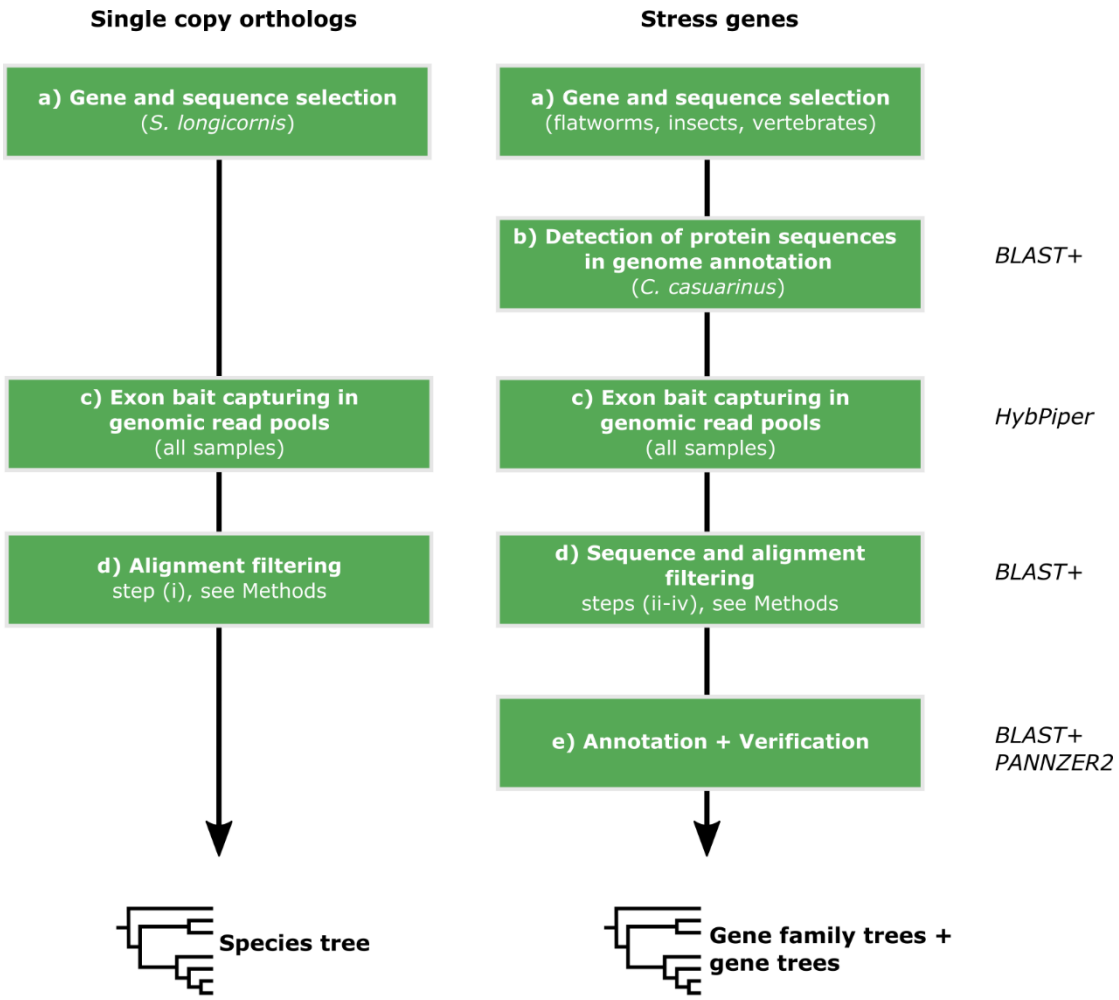

Figure 2. Schematic bioinformatic strategy for detecting single-copy orthologs (SCOs) and orthologs of stress genes in monogenean whole-genome short-reads. SCO sequences were used to infer the species tree and stress gene sequences for gene family trees and gene trees. (a) Bait sequences [*S. longicornis* was selected for SCOs; other organisms (non-monogenean flatworms, insects, vertebrates) 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* (Johnson et al. 2016). (d) Contaminant, variant, low–species coverage, and truncated sequences were filtered from the alignments. (e) Sequences were annotated through *PANNZER2* (Törönen & Holm 2022) (e).

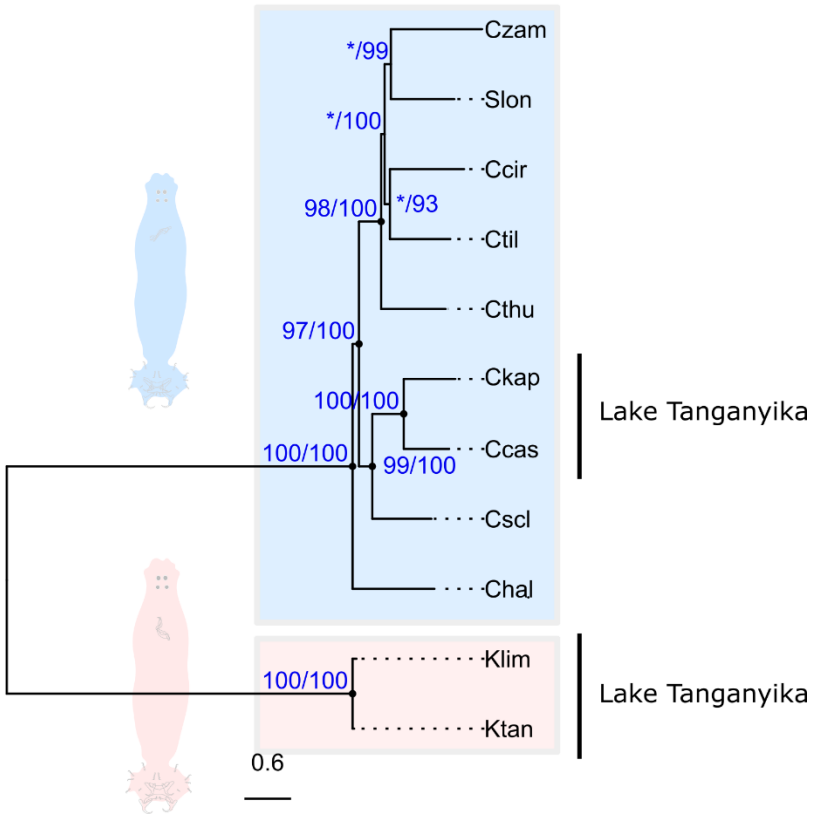

Figure 3. Species tree of *Cichlidogyrus* and *Kapentagyrus* inferred from 277 single-copy orthologs based on a subset of genes selected by Caña-Bozada et al.(Caña-Bozada et al. 2023), who used the OMA pipeline. For the tree inferred from orthologs from the BUSCO pipeline, see Supplementary Fig. S1. Support values: ultrafast bootstraps (UF-Boot)/Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT) (see Methods), asterisks (*) indicate support below threshold (UF-Boot ≤ 95, SH-aLRT ≤ 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.

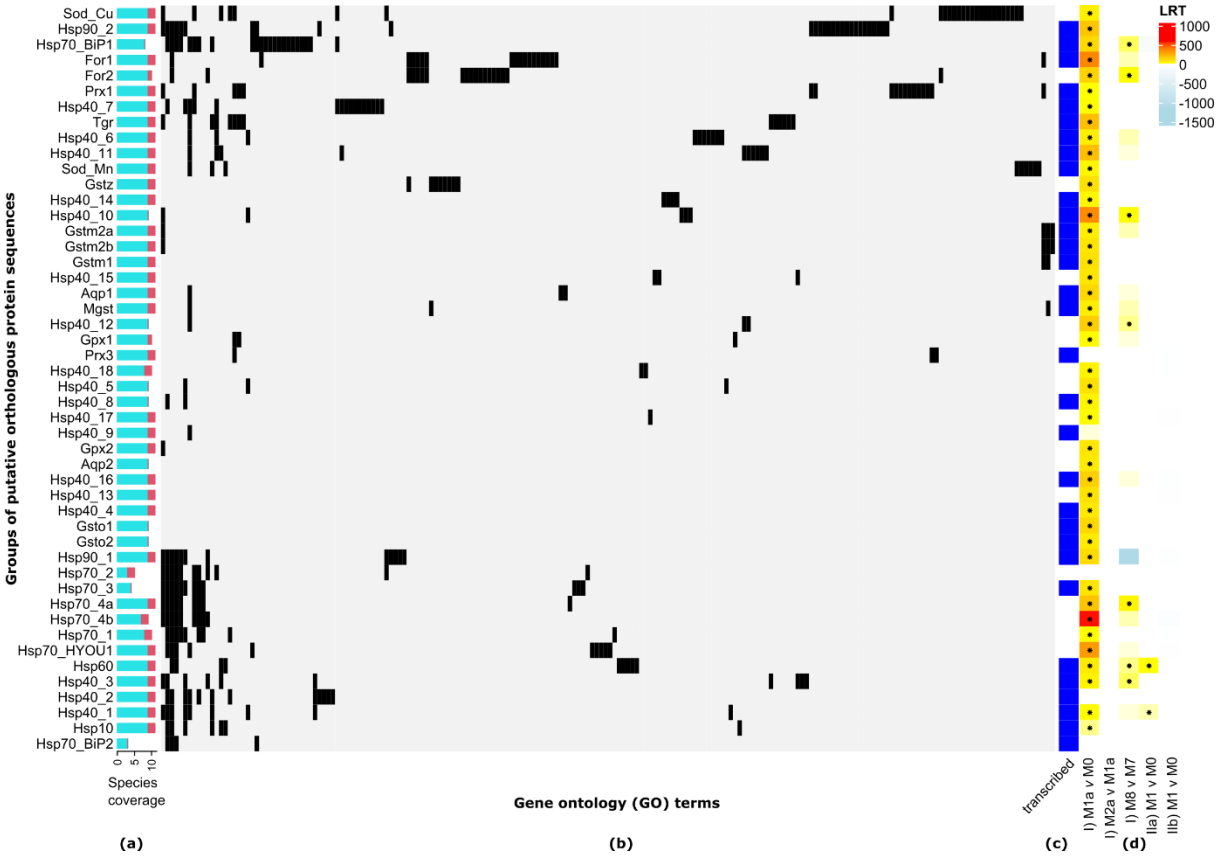

Figure 4. Detected putative stress gene orthologs (protein sequences) including species coverage (cyan = *Cichlidogyrus*, red = *Kapentagyrus*) (a), gene ontology (GO) terms (black = term applies) (b), presence in transcriptome annotation (blue = present) (c), and hypothesis testing of different models 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 Methods). Rows and columns of the GO heatmap are automatically sorted through Euclidean distances as implemented in *ComplexHeatmap*. For an extended version of this figure with GO term labels, see Supplementary Fig. S6.

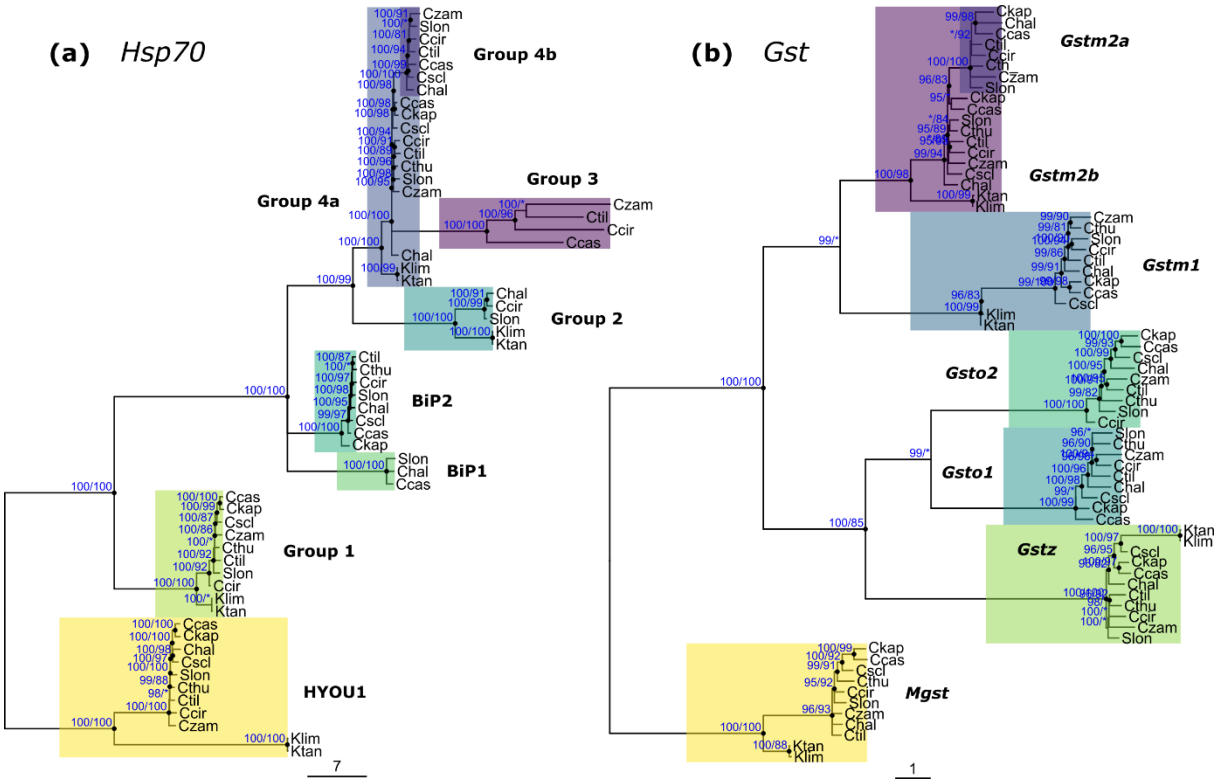

Figure 5. Maximum likelihood topologies of gene family trees and gene trees of species of *Cichlidogyrus* and *Kapentagyrus*. For abbreviation of species names, support values, and scale bars, see Figure 3. (a) Gene models of the 70 kDa heat shock protein family (*Hsp70*). Group Hyou1 (hypoxia up-regulated 1), BiP1 (endoplasmatic reticulum chaperone binding protein 1), and BiP2 refer to annotations assigned through *PANNZER2* (see Supplementary Table S3); 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*.

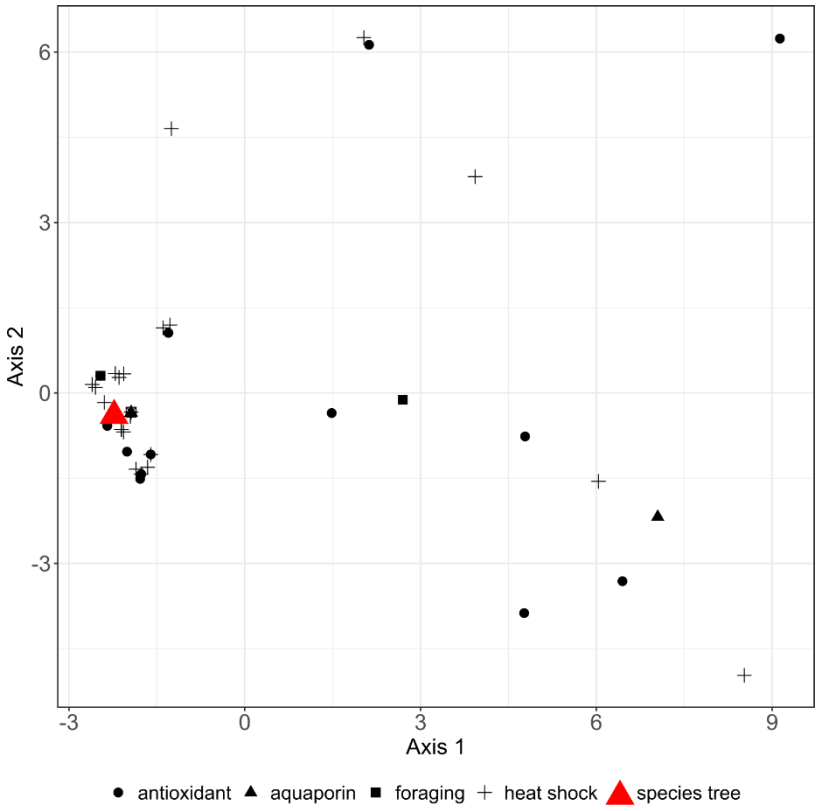

Figure 6. First two axes (49% of total variation) of multidimensional scaling analysis of gene trees of antioxidant enzymes, *for* orthologs, aquaporins, and heat shock proteins, with some gene tree topologies deviating from the two species trees (highlighted in red) but not forming clusters based on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

INVENTORY OF SUPPORTING INFORMATION

Supplementary Figure S1. 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 Methods), asterisks (*) indicate support below threshold (UF-Boot ≤ 95, SH-aLRT ≤ 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.

Supplementary File S2. Draft assembly and annotation of genome of *Cichlidogyrus casuarinus*.

Supplementary Table S3. 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 (Aguoru et al. 2022) at *UniProt* (Bateman et al. 2023). Annotations were inferred from *PANNZER2* (Törönen & Holm 2022) (see Fig. 2).

Supplementary Fig. S4. Number of paralogs flagged for each of the 48 assembled stress genes by parasite species. Gene names reflect bait sequences from draft annotation of *Cichlidogyrus casuarinus* (see Supplementary File S2). 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*.

Supplementary File S5. 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*.

Supplementary Table S6. Sampling data of collected specimens including reference for sampling campaigns and published whole-genome sequencing data.

Supplementary Figure S7. Extended version of Fig. 4. Detected putative stress gene orthologs (protein sequences) including species coverage (cyan = *Cichlidogyrus*, 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 Methods). Rows and columns of the GO heatmap are automatically sorted through Euclidean distances as implemented in *ComplexHeatmap*.