Copy numbers of stress response genes reflect difference in adaptive potential of metazoan parasites

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

All authors have approved the current version of the manuscript.

1 UHasselt – Hasselt University, Faculty of Sciences, Centre for Environmental Sciences, Research Group Zoology: Biodiversity and Toxicology, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium.

2 Natural History Museum, University of Oslo, 0318 Oslo, Norway.

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

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

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

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

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

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

9 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.

Abstract

Stress responses are key for parasite survival and, thus, also the evolutionary success of these organisms. However, the evolution of the molecular pathways dealing with environmental stressors are poorly understood as most research focuses either on few selected human-relevant pathogens or major parasite clades. Here, we comparatively investigate, for the first time, antioxidant, heat shock, and behaviour-related genes in the two parasite lineages *Cichlidogyrus* and *Kapentagyrus* from the same family Dactylogyridae through whole-genome sequencing data of 11 species. The two lineages differ concerning their species and ecological diversity, which is expected to affect the diversity of their stress responses and, hence, their adaptive potential. Through an exon bait capture approach, we assembled the putative protein sequences of 43 stress-related genes. We discovered that *Cichlidogyrus* presented higher copy numbers of stress genes (70 kDA heat shock protein, glutathione *S*-transferase genes) than *Kapentagyrus*. This difference might explain the ability of species of *Cichlidogyrus* to colonise various cichlid and non-cichlid lineages. In comparison to most other organisms studied so far, we also observed a previously unreported absence of cytochrome P450 and sigma class glutathione *S*-transferase in monogenean flatworms. This pattern aligns with previously published genome annotations of monogeneans.

6.1 Introduction

Evolutionary theory predicts that a species' ability to maintain homeostasis can fundamentally affect its adaptive potential in the face of environmental stressors (Bijlsma & Loeschcke, 2005). Metazoan parasites are one of the most overlooked groups of pathogens and cause a majority of neglected human diseases recognised by WHO (Hotez et al., 2020). Previous research on parasite stress responses remains largely limited to two main areas: first, few well-known human-infecting species, e.g., *Schistosoma* sp. (Huyse et al., 2018; Aguoru et al., 2022) for the purpose of drug development and, second, differences in stress response pathways in major clades of metazoan parasites (Tsai et al., 2013; Hahn et al., 2014) to infer macroevolutionary adaptations to their parasitic lifestyle. This focus leaves an important knowledge gap: how do adaptive potentials in terms of stress response systems differ between parasites that are more closely related and functionally alike, i.e. what makes one parasite lineage more stress-tolerant than the other?

Understanding the link between stress responses and adaptive potential of parasites is of major interest to society. Stress responses might affect and determine infectivity, virulence, and disease burden of parasites (Ismail et al., 2018; Santi et al., 2022). Minor differences between parasite populations and species in dealing with host defences can determine their chances of survival. Stress responses are a major way of dealing with host defences, thus, effective stress responses matter not only for the fitness of individuals and populations (microevolution) but also enable parasite species to expand their geographical ranges or colonise new host species, eventually giving rise to new parasite species and parasitic diseases (macroevolution) (McCoy, 2003; Huyse et al., 2005). Moreover, anthropogenic impact on the environment may promote the rise of emerging infectious diseases (Brooks et al., 2019) as environmental disturbance can create ecological opportunity for parasite species with higher adaptive potential (D'Bastiani et al., 2020; Hector et al., 2023). Despite this importance, the link between stress responses and the adaptive potential of parasite lineages remains unresolved as differences in these pathways are rarely studied below the level of major clades, e.g. between species of the same flatworm class or insect order.

One major question that arises when investigating stress responses is which stress response mechanisms are most likely to mediate the adaptive potential of parasites. Indicative of adaptive potential is the ability of some lineages to use a broad spectrum of resources, i.e. host species. In this context, several stress-related proteins have been characterised as major factors in determining resource usage, e.g. in parasitic insects, nematodes, and fungi (Coleman et al., 2009; Calla et al., 2017; Gouin et al., 2017; Wang et al., 2018; Zhang et al., 2020). The first group of these proteins are antioxidant enzymes. Maintaining an equilibrium of oxidation and reduction processes (redox) is a crucial aspect of life (Sies et al., 2017). Oxidative stress is caused by an imbalance in the production and accumulation of reactive oxygen species (ROS)—a product of redox reactions in cells, which damages the cell and its functions. Increased ROS production can result from a variety of environmental and physiological stressors such as oxygen tension, heat, toxic compounds, and radiation (Sies et al., 2017). Antioxidant enzymes, which deactivate ROS, are mostly studied in well-established model organisms, e.g. humans, mice, zebrafish, and fruit flies (Sies et al., 2017). Antioxidant enzymes are conserved across the animal kingdom with well-known enzymes such as catalase (CAT) (*note on protein/gene nomenclature: protein names—uppercase, equivalent genes—lowercase and in italics*), cytochrome P450 (CYP), glutathione peroxidases (GPx), glutathione-S-transferases (GST), superoxide dismutases (SOD), peroxiredoxines (Prx), and thioredoxin (TRX). Accordingly, these enzymes have been reported from a range of metazoan parasite taxa including flatworms (Chiumiento & Bruschi, 2009; Martínez-González et al., 2022), nematodes (Chiumiento & Bruschi, 2009), phytophagous (Mittapalli et al., 2007; Łukasik et al., 2011) and parasitoid (Sabadin et al., 2019) insects, and blood-feeding arthropods (Sabadin et al., 2019). However, both in parasitic flatworms (Martínez-González et al., 2022) and nematodes (Xu et al., 2020a) CAT is absent. Differences in terms of copy numbers have also been reported for parasites as compared to free-living organisms. In parasitic flatworms, *cyp* and *gpx* are reduced to a single copy (Martínez-González et al., 2022) and parasitic nematodes possess an additional extracellular SOD variant unlike their free-living relatives (Xu et al., 2020a). Furthermore, the expression of antioxidant enzymes is key for the colonisation of new host species, e.g. SOD and CAT in aphids (Durak et al., 2018; Łukasik et al., 2022). Accordingly, antioxidant

enzymes are potentially important targets for understanding the molecular mechanisms determining the adaptive potential of parasites.

Heat shock proteins (HSP) form another group of proteins produced under stressful conditions [summarised in Calderwood et al. (2019)], which function mostly as molecular chaperones assisting with protein folding and refolding. As their name suggests, HSPs are a key part of the organism's heat shock response, but they are also involved in cold, radiation, and wound healing responses [see Calderwood et al. (2019) and references therein]. Therefore, HSPs are considered important determinants for the adaptive potential of parasite lineages alongside antioxidant enzymes. The functions of heat shock genes (*hsp*) and proteins (HSP) has been studied in several human parasites, albeit again mostly in the context of responses to antihelmintic drugs, e.g. in the human-infecting blood flukes (Abou-El-Naga, 2020; Xu et al., 2020b). Unsurprisingly, the expression of many HSP families is also induced by thermal stress in several parasite lineages, e.g. in copepods (Borchel et al., 2018), nematodes (Chen et al., 2014), ticks (Busby et al., 2012), and phytophagous (Yuan et al., 2014), parasitoid (Pan et al., 2018), and blood-feeding (Pereira et al., 2017) insects. High degrees of thermal tolerance in some parasites might pose a human health risk. Global warming is expected to provide suitable habitats for some pathogens and disease vectors with increased thermal tolerance (Morley & Lewis, 2014; Carlson et al., 2017) and might lead to the rise of emerging and re-emerging infectious diseases (Marcogliese, 2008; Brooks et al., 2019). Several studies also highlight the importance of HSPs for parasite lineages to avoid host defences through the expansion of *hsp* gene families, e.g. work on tapeworms (Tsai et al., 2013), lung flukes (Rosa et al., 2020), and liver flukes (Choi et al., 2020). Beyond copy number differences, HSP expression levels also likely determine the adaptive potential of parasites as parasite species infecting multiple host species need to deal with variations of host defence systems (see examples from phytophagous insects: Roy et al., 2018; Liu et al., 2019). Furthermore, HSPs enable parasites to cope with environmental stressors associated with different life stages, see studies on liver flukes (Yoo et al., 2011). Heat shock responses are key determinants of the adaptive potential of parasite species.

In addition to physiological stress responses, animals may also address environmental stressors through behavioural changes. Animal behaviour is often

thought to be influenced by numerous genes, making it inherently challenging to decipher the connection between genetic and behavioural traits (Anreiter & Sokolowski, 2019). The *foraging* (*for*) gene of *Drosophila melanogaster* Meigen, 1830 is one of the best-known examples of a gene with major behavioural effects (Sokolowski, 1980). The gene codes for a cGMP-dependent protein kinase (PKG) and was found to group larval and adult flies into two distinct clusters according to feeding behaviour, i.e. rovers vs. sitters (Anreiter & Sokolowski, 2019). Following its discovery in *D. melanogaster*, the role of *for* in parasite behaviour has been studied mostly in phytophagous insects. For instance, aphids infecting multiple host species showed higher expression of *for*, which was linked to an increased learning capacity, i.e. 'remembering' their initial host preference even after a longer period of exposure to a different host species, unlike specialist congeners, which quickly lose attraction to their original host (Tapia et al., 2015). Even though the role of *for* in parasite evolution was mostly studied in insects including plant-parasitic insects (Lucas et al., 2010; Chardonnet et al., 2014; Gallot et al., 2022) and human disease vectors (Keating et al., 2013; Marliére et al., 2020), the gene is evolutionarily conserved influencing human social behaviour (Struk et al., 2019) as well as feeding behaviour and satiety quiescence in nematodes (Raizen et al., 2008; Kroetz et al., 2012). Therefore, orthologs are also likely found among other metazoan parasites such as in blood flukes, where a PKG has been linked to fecundity of the parasites (Huyse et al., 2018).

In the present study, we explore the link between adaptive potential, copy numbers, and diversity of orthologs of antioxidant enzymes, heat shock proteins, and *for* in two closely related lineages of monogenean flatworms infecting teleost fish hosts. *Cichlidogyrus* Paperna, 1968 is the most species-rich genus of monogeneans on the African continent with currently 144 described species [including the nested genus *Scutogyrus* Pariselle & Euzet, 1995] (Cruz-Laufer et al., 2022; Geraerts et al., 2022b; Moons et al., 2023). *Cichlidogyrus* provides a model system for evolution and ecology of host-parasite interactions (Pariselle et al., 2003; Vanhove et al., 2016; Cruz-Laufer et al., 2022). Species of *Cichlidogyrus* infect the gills of cichlid fishes among those the hyperdiverse cichlid lineages in the East African lakes, one of the best-known model systems of adaptive radiation, behavioural, and genomic research (Jordan et al., 2021; Wagner, 2021). Cichlids cover various ecological niches including herbivory, invertivory, piscivory and

omnivory (Burress, 2014), and inhabit lacustrine, riverine, and estuarine habitats across Africa (Chapman, 2021). This ecological diversity exposes cichlid ectoparasites to various environmental stressors. In contrast, *Kapentagyrus* Kmentová, Gelnar & Vanhove, 2018 currently only contains 13 species (Vanhove et al., 2021). These parasites infect African freshwater sardines (Pellonulini), which all inhabit the pelagic zones of African lakes and rivers (Wilson et al., 2008). Pellonuline clupeids like their parasites are much less species-rich than their cichlid counterparts. *Cichlidogyrus* and *Kapentagyrus* both belong to the same subfamily Dactylogyrinae sensu Kmentová et al. (2022) (Monogenea: Dactylogyridae), yet the niche conservatism of the hosts and the much lower species richness suggest that species of *Kapentagyrus* have a lower adaptive potential than species of *Cichlidogyrus* (Fig. 6.1).

The contrast in habitat diversity of the two monogenean lineages provides a playing ground to test differences in stress response and adaptive potential in parasites (Fig. 6.1). We investigate differences (copy numbers and gene evolution) in the orthologs of 43 antioxidant enzymes, heat shock proteins, and *for* using whole-genome sequencing data of species of *Cichlidogyrus* from eight different lineages and two species of *Kapentagyrus*. As shown above, stressful environments can lead to changes in copy numbers of stress response genes in parasites. However, rapid evolution can also lead to discordance in the evolution of stress genes with the evolution of the lineage [see examples in plants (Pease et al., 2016; Wu et al., 2018)]. We hypothesise that:

(i) Species of *Cichlidogyrus* have higher copy numbers of stress gene families than species of *Kapentagyrus*.

(ii) The evolution of stress genes in species of *Cichlidogyrus* deviates from the genome level evolutionary history of the lineage.

We assess potential gene family expansions/contractions through phylogenetic analyses of the putative protein sequences of respective gene families and subfamilies (i). We also explore the variation of the gene trees compared to species tree using multidimensional scaling (ii). Through these approaches, we aim to address how differences in copy numbers and stress gene evolution reflect the adaptive potential of parasite lineages.

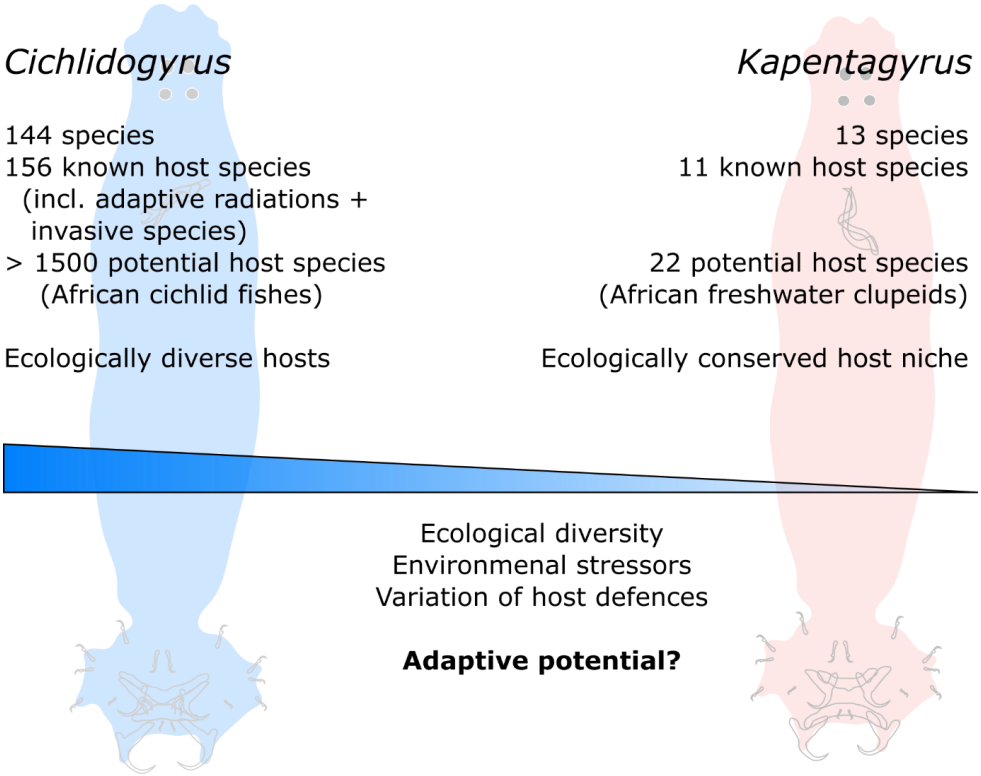


Figure 6.1. Overview of species richness and ecological diversity of species of two dactylogyrid monogenean genera: *Cichlidogyrus* and *Kapentagyrus*.

6.2 Material and methods

6.2.1 Sample collection and DNA sequencing

To analyse a representative selection of the species diversity of the monogenean genus *Cichlidogyrus*, we aimed to collect at least one species from each of the 11 main lineages that were recently reported in a phylogenetic study of nuclear ribosomal and mitochondrial genes (Cruz-Laufer et al., 2022). Using samples collected during recent expeditions (Geraerts et al., 2022a; Kmentová et al., 2023), we compiled material from species from eight lineages of *Cichlidogyrus* (Table 6.1). Hosts were collected with the help of local fisherfolk and the gills were extracted from the fishes and stored in absolute ethanol. Parasites were collected from the gills using entomological needles. Most parasite specimens were identified to species level according to Geraerts et al. (2022a) except for *C. zambezensis* and *C.* sp. 'kapembwa', which were identified/characterised in the present study based on their sclerotised attachment and reproductive organs (see

footnotes in Table 6.1). Genomic DNA extraction followed protocols published by 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 Table 6.1). 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 sequence coverages, see Table 6.1). Furthermore, we accessed genomic read pools of one species of *Cichlidogyrus* and two of *Kapentagyrus* from previous mitogenomic studies (Kmentová et al., 2021b, 2023) (Accession BioProjects: PRJNA749051, XXXXX, XXXXX). We also attempted to use genome short reads of different species of *Cichlidogyrus/Scutogyrus* published by Vanhove et al. (2018) and Caña-Bozada et al. (2021). However, the coverage of these reads proved to be too low for capturing the genes targeted in the present study. 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).

Table 6.1. Sampling data of collected specimens including reference for sampling campaigns and published whole-genome sequencing data.

Species	Host	Country	Locality	LAT	LON	Date	Pool size	Estimated coverage[1]	Sampling campaign	Genome sequences
Cichlidogyrus casuarinus Pariselle, Muterezi Bukinga & Vanhove, 2015	*Hemibates stenosoma* (Boulenger, 1901)	Zambia	Lake Tanganyika, Mpulungu	-08.7667	31.1167	09/2018 09/2019	80	825	Kmentová et al. (2021)	Hahn et al. (unpublished)
Cichlidogyrus cirratus Paperna, 1964	*Oreochromis niloticus* (L., 1758)	Cameroon	Yaoundé, Ponds at Ecoparque	3.78582	11.4883	13/02/2022	50	643	Geraerts et al. (2022)	This study
Cichlidogyrus halli (Price & Kirk, 1967)	*Oreochromis niloticus* (L., 1758)	Cameroon	Yaoundé, Fishing ponds of Obili, 'Projet de Promotion de l'Entreprenariat Aquacole (P.P.E.A.)	3.85672	11.4958	14/02/2020	50	1374	Geraerts et al. (2022)	This study
Cichlidogyrus sclerosus Paperna & Thurston, 1969	*Oreochromis niloticus* (L., 1758)	Zimbabwe	Kariba, Lake Kariba at discharge channel of crocodile farm	-16.5495	28.8616	27/10/2019	50	365	Geraerts et al. (2022)	This study
Cichlidogyrus sp. 'kapembwa'[2,3]	*Callochromis macrops* (Boulenger, 1898)	Zambia	Lake Tanganyika, Kalambo Lodge	-8.6539	31.1954	01/10/2019	52	1510	Kmentová et al. (2023)[4]	This study
Cichlidogyrus thurstonae Ergens, 1981	*Oreochromis niloticus* (L., 1758)	Cameroon	Mbalmayo, River Fala (affluent of the river So'o)	3,318972	11,48519	17/02/2020	50	556	Geraerts et al. (2022)	This study
Cichlidogyrus tilapiae Paperna, 1960	*Oreochromis niloticus* (L., 1758)	Cameroon	Mbalmayo, Ponds Capfort	3.532733	11.51642	17/02/2020	50	482	Geraerts et al. (2022)	This study
Cichlidogyrus zambezensis Douëllou, 1993[3,5]	*Serranochromis* cf. *thumbergi*	Dem. Rep. Congo	Province Haut-Katanga, Kiniama, Kafubu River, near Catholic Mission	-11.4777	28.3089	27/07/2019	1	768	Geraerts et al. (2022)	This study
Scutogyrus longicornis (Paperna & Thurston, 1969)	*Oreochromis niloticus* (L., 1758)	Cameroon	Édéa, small affluent stream (Mboue) of Sanaga River running through the village	3.79215	10.13303	11–12/02/2022	50	638	Geraerts et al. (2022)	This study
Kapentagyrus limnotrissae (Paperna, 1973)	*Limnothrissa miodon* (Boulenger, 1906)	Zambia	Lake Tanganyika, Mpulungu	-08.7667	31.1167	08/2018	50	5451	Kmentová et al. (2023)	Kmentová et al. (2023)
Kapentagyrus tanganicanus Kmentová, Gelnar & Vanhove, 2018	*Limnothrissa miodon* (Boulenger, 1906)	Zambia	Lake Tanganyika, Mpulungu	-08.7667	31.1167	12/04/2018	85	8292	Kmentová et al. (2023)	Kmentová et al. (2023)

[1] Based on genome size of genome assembly for *C. casuarinus* (66.086 MB) (Hahn et al., unpublished). [2] Species currently under description (part of *EAR* clade sensu Chapter 2. [3] Whole-genome pre-amplification was applied. [4] Sample location not mentioned in cited publication. [5] Species identified in present studies based on description/re-description of *C. zambezensis* and *C. pseuodzambezensis* (Jorissen et al., 2018; Geraerts et al., 2020).

6.2.2 Gene selection for gene tree estimation

We focused on 12 gene families from three different functional groups: antioxidant enzymes (CYP, GPx, GSTM, GSTS, Prx, SOD, and TGR), HSPs (HSP10, HSP40, HSP60, HSP70, and HSP90), and *for* (see Table 6.2 for abbreviations). For the antioxidant enzymes, we included the main groups previously reported from parasitic flatworms (Martínez-González et al., 2022). For HSPs, we included the best-studied gene families in flatworms (Aguoru et al., 2022). An illustration of the ortholog selection process described below can be found in Figure 2. As performance of exon bait capture (see section below: *protein sequence assembly*) decreases with phylogenetic distance (Yuan et al., 2019), we aimed to use bait sequences from species that are as closely related as possible to the target taxa. However, neither nucleotide nor amino acid sequences of the functional genes targeted in the present study have ever been assembled and annotated for monogenean flatworms [except for the *hsp70* subfamily in the distantly related *Gyrodactylus salaris* Malmberg, 1957 (Hahn et al., 2014)]. Therefore, we used protein sequences of other flatworm species (Fig. 6.2a) to detect putative protein orthologs in an annotation of an assembled genome of *Cichlidogyrus casuarinus* (Hahn et al. unpublished) using *BLAST*+ v2.13.0 (Camacho et al., 2009) (Fig. 6.2b). For a list of accession numbers and the respective protein IDs in *C. casuarinus*, see Table 6.2. Not a single *for* ortholog has been reported in flatworms to date. Therefore, we used protein sequences of *for* isoforms of *D. melanogaster* for this purpose (Table 6.2, Fig. 6.2a). We included matches with query coverages above 90% to only include matches that were highly likely to present orthologs.

6.2.3 Gene selection for species tree estimation

To compare patterns of gene evolution against the evolutionary history of *Cichlidogyrus*, we inferred a species tree based on single-copy nuclear orthologs. 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 gene alignment errors, which in return create noise in phylogenetic analyses (e.g. long-branch attraction of rapidly evolving lineages)

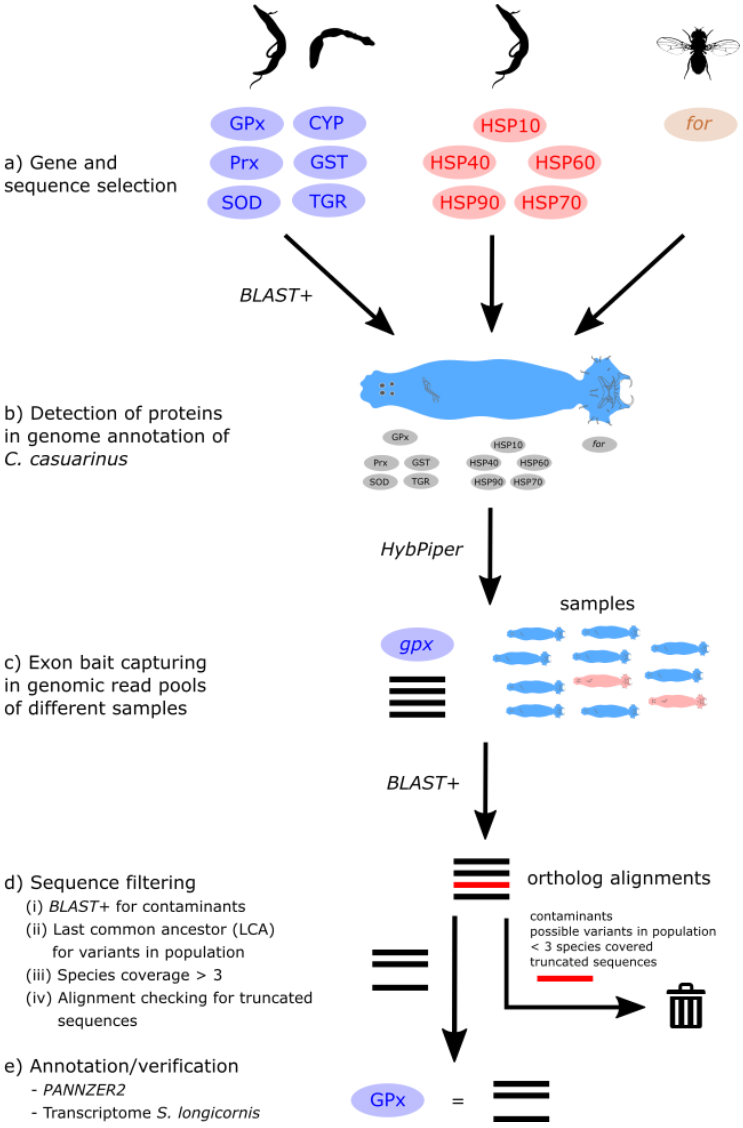


Figure 6.2. Steps of ortholog detection in monogenean whole-genome short-reads for stress response genes. First, antioxidant enzymes and the heat shock proteins sequences of other flatworms (*Schistosoma* spp. and *Echinococcus multilocularis*, depicted by their silhouette) and *foraging* gene sequences of *D. melanogaster* were selected (a). Second, orthologs of these sequences were detected in an annotated genome of *Cichlidogyrus casuarinus* (Hahn et al. unpublished) using *BLAST+* (b). Third, the putative protein sequences of *C. casuarinus* were used as baits for exon bait capturing in the different whole-genome sequencing reads of species of *Cichlidogyrus* (blue) and *Kapentagyrus* (pink) through *HybPiper* (Johnson et al., 2016) (c). Fourth, contaminant, variant, low–species coverage, and truncated sequences (red) were filtered from the alignments (black) (d). Fifth, the resulting sequences were annotated through *PANNZER2* (Törönen & Holm, 2022)

(Philippe et al., 2005). Genome data provide a large number of alternative markers for phylogenetic inference, e.g. large datasets composed of single-copy orthologous genes have been used to resolve phylogenetic relationships of plant, fungal, and animal taxa (Philippe et al., 2005; Aguileta et al., 2008; Lu et al., 2014). Recently, this approach has been adopted for the first time with a neodermatan phylogeny (Caña-Bozada et al., 2023) including several monogenean species—one of them belonging to *Scutogyrus*, a nested lineage of *Cichlidogyrus*. This phylogeny was based on 137 and 479 orthologous groups of proteins inferred from the *BUSCO* (Waterhouse et al., 2018) and *OMA* (Dylus et al., 2020) pipelines, respectively. As *S. longicornis* is a member of one of the target lineages of this study, we omitted the ortholog detection step in *C. casuarinus* implemented for the target gene groups (see Fig. 6.2b). Instead, we used these putative protein sequences of *S. longicornis* assembled by Caña-Bozada et al. (2023) as bait sequences for the downstream putative protein sequence assembly (not shown in Fig. 6.2). While the assembly of the *BUSCO* and *OMA* genes through *HybPiper* provides no guarantee that these gene have indeed only a single-copy in genomes of species of *Cichlidogyrus*, duplications events in a substantial number of these conserved genes appear unlikely due to their single-copy nature in *S. longicornis* and other flatworm lineages (Caña-Bozada et al., 2023).

Table 6.2. Overview of sequences used for bait capturing of target gene groups and baited sequences (hits) with annotations. Protein sequences from Aguoru et al. (2022) can be accessed via the protein ID at UniProt (Bateman et al., 2023). Annotations were inferred from PANNZER2 (Törönen & Holm, 2022) (see Fig. 6.2).

Bait sequence				**Hits**		**Annotations**		
Name	Family	Species	*GenBank* Accession	Protein ID (*C. casuarinus*)[1]	*HybPiper*	Description	GO terms	Transcriptome (*S. longicornis*)
Cytochrome P450 (CYP)		*Schistosoma mansoni* Sambon, 1907		—	—	—	—	—
Glutathione peroxidase (GPx)		*S. mansoni*	Q00277.2	XXXXX	Group 1: all but *K. limnotrissae*	Glutathione peroxidase	GO:0006979 GO:0098869 GO:0004602	
				XXXXX	Group 2: all	Glutathione peroxidase	GO:0005829	
Glutathione *S*-transferase (GST)	Microsomal (MAPEG)	*Echinococcus multilocularis* Leuckart, 1863	CDS41239	XXXXX	All	Microsomal glutathione S-transferase 3	GO:0004364 GO:0016740 GO:0016020	✓
	mu-class (GSTM)	*E. multilocularis*	CDS38059.1 CDS38060.1 CDS38083.1 CAA59739.1	XXXXX	Group 1: all	Glutathione S-transferase class-mu 26 kDa isozyme	GO:0004364 GO:0042802	✓
			CDS38090.2 CDS38091.1 CDS38092.1 CDS38086.2	XXXXX XXXXX XXXXX	Group 2a: *Cichlidogyrus* spp. without *C. sclerosus*	Glutathione transferase	GO:0042221 GO:0004364 GO:0042802 GO:0005829	✓
			CDS38094.2 CDI96481.2		Group 2b: all	Glutathione transferase		✓
	sigma-class (GSTS)	*E. multilocularis*	CDS39356.1 CDS37347.1	—	—	—	—	—
Peroxiredoxin (Prx)	1	*S. mansoni*	AAD17299.1	XXXXX		Peroxiredoxin	GO:0098869 GO:0006979 GO:0042744 GO:0006801 GO:1901701 GO:0048144 GO:0042267 GO:0030101 GO:0045454	✓

							GO:0051920 GO:0042802 GO:0045296 GO:0003723 GO:0005829 GO:0000502 GO:0005634 GO:0070062 GO:0042470	
	2	*S. mansoni*	XP_018645129.1	—	—	—	—	—
	3	*S. mansoni*	AAG15506.1	XXXXX		Thioredoxin-dependent peroxide reductase, mitochondrial	GO:0098869 GO:0034599 GO:0051920	✓
Superoxide dismutase (SOD)	Cu-Zn	*S. mansoni*	Q01137.1	XXXXX		Superoxide dismutase [Cu-Zn]	GO:0006801 GO:0071451 GO:0098869 GO:0071450 GO:0000303 GO:0045541 GO:0050665 GO:0051881 GO:0060087 GO:0051597 GO:0070050 GO:0009410 GO:0010038 GO:0005507 GO:0004784 GO:0030346 GO:0051087 GO:0008270 GO:0005615 GO:0005829 GO:0005634 GO:0032839 GO:0043025 GO:0031410 GO:0005777 GO:0005739 GO:0032991 GO:0032991	

	Mn	*Schistosoma japonicum* (Katsurada, 1904)	AAW26480.1	XXXXX		Superoxide dismutase	GO:0019430 GO:0051597 GO:0001780 GO:0046686 GO:0045087 GO:0004784 GO:0046872 GO:0005759 GO:0016020	✓
Thioredoxin glutathione reductase (TGR)		*S. mansoni*	XP_018649018.1	XXXXX		thioredoxin-disulfide reductase	GO:0045454 GO:0098869 GO:0010035 GO:0006979 GO:0007283 GO:0030154 GO:0004791 GO:0050660 GO:0046872 GO:0005654 GO:0005829 GO:0005739 GO:0005783 GO:0016020	✓
Heat shock protein 10 kDa		*S. mansoni*	XP_018653498.1	XXXXX		10 kDa heat shock protein, mitochondrial	GO:0006457 GO:0044183 GO:0005524 GO:0051087 GO:0051082 GO:0046872 GO:0005759	✓
Heat shock protein 40 kDa	31 members	*S. mansoni*	see Aguoru et al. (2022)	XXXXX	Group 1	DnaJ homolog subfamily A member 1	GO:0009408 GO:0006457 GO:0030544 GO:0051082 GO:0005524 GO:0046872 GO:0016874 GO:0016020 GO:0005829	✓
				XXXXX	Group 2	Tumorous imaginal discs, mitochondrial	GO:0009408 GO:0006457 GO:0043066 GO:0007005	✓

				GO:0006807 GO:0031072 GO:0051082 GO:0005524 GO:0046872 GO:0008168 GO:0005739 GO:0016020	
	XXXXX	Group 3	DnaJ heat shock protein family (Hsp40) member B4	GO:0006457 GO:0000122 GO:0006986 GO:0009408 GO:0051082 GO:0001671 GO:0051087 GO:0005654 GO:0005829 GO:0005886	✓
	XXXXX	Group 4	J domain-containing protein	—	✓
	XXXXX	Group 5	DnaJ-like protein subfamily B member 8	GO:0061077 GO:0030544 GO:0051082	
	XXXXX	Group 6	DnaJ heat shock protein family (Hsp40) member B14	GO:0051085 GO:0065003 GO:0071218 GO:0030544 GO:0005783 GO:0016020 GO:0031984 GO:0005635 GO:0042175 GO:0031090	✓
	XXXXX	Group 7	DnaJ homolog shv/DnaJ homolog subfamily B member 11	GO:0006457 GO:0051604 GO:0036098 GO:0060250 GO:0090129 GO:2001046 GO:0007155 GO:0051082 GO:0051787 GO:0005178	✓

				GO:0005783 GO:0016020 GO:0031594 GO:0005615 GO:0005634 GO:0071944	
	XXXXX	Group 8	unknown	GO:0006457 GO:0051082	✓
	XXXXX	Group 9	DnaJ homolog subfamily C member 1	GO:0016020	✓
	XXXXX	Group 10	DnaJ homolog subfamily C member 2	GO:0051083 GO:0006450 GO:0030544 GO:0043022 GO:0005829	✓
	XXXXX	Group 11	DnaJ homolog subfamily C member 3/putative dsrna-activated protein kinase inhibitor p58	GO:0016310 GO:0034975 GO:0006415 GO:0016301 GO:0051787 GO:0051087 GO:0003747 GO:0005783 GO:0016020 GO:0061617	✓
	XXXXX	Group 12	DnaJ heat shock protein family (Hsp40) member C7	GO:0016310 GO:0016301 GO:0016020	
	XXXXX	Group 13	DnaJ homolog subfamily C member 8	—	
	XXXXX	Group 14	DnaJ homolog subfamily C member 9	GO:0016705 GO:0004497 GO:0005506 GO:0020037	✓
	XXXXX	Group 15	Mitochondrial import inner membrane translocase subunit TIM14	GO:0030150 GO:0001671 GO:0001405	
	XXXXX	Group 16	unknown	—	✓

				XXXXX	Group 17	DnaJ subfamily C member 17	GO:0003676	
				XXXXX	Group 18	DnaJ sub C member 27, variant 2	GO:0003924 GO:0005525	
Heat shock protein 60 kDa	HSP60	*S. mansoni*	XP_018645622.1	XXXXX		Heat shock protein 60	GO:0042026 GO:0034514 GO:0008637 GO:0045041 GO:0140662 GO:0005524 GO:0051087 GO:0005759 GO:0005743	✓
Heat shock protein 70 kDa	HSPA9 protein-like	*S. mansoni*	see Aguoru et al. (2022)	XXXXX	Group 1: All except *C. halli* and *C.* sp. 'kapembwa'	Heat shock protein cognate 5	GO:0006457 GO:0006950 GO:0140662 GO:0051082 GO:0005524 GO:0031072 GO:0032440 GO:0016887 GO:0005739	
				XXXXX	Group 1bis: All except *C. halli* and *C.* sp. 'kapembwa'	Heat shock protein 70	GO:0006457 GO:0140662 GO:0005524	
	Hypoxia up-regulated protein 1 precursor-like	*S. mansoni*	see Aguoru et al. (2022)	XXXXX	Group HYOU1: All	Molecular chaperone grp170/sil1 hsp70 superfamily protein + Hypoxia up-regulated protein 1	GO:0006457 GO:1903297 GO:1900038 GO:0031204 GO:0071456 GO:2001243 GO:0140662 GO:0005524 GO:0034663 GO:0016020	
	Endoplasmic reticulum chaperone BiP precursor-like	*S. mansoni*	XP_018649109.1	XXXXX	Group BiP1: *Cichlidogyrus* spp. except *C. halli*	Heat shock 70 kDa protein cognate 3	GO:0006457 GO:0036335 GO:0034976 GO:0035194 GO:0006782 GO:0034620	✓

						GO:0010243 GO:0043161 GO:0035332 GO:0046621 GO:0007165 GO:0140662 GO:0005524 GO:0004655 GO:0031072 GO:0016887 GO:0046872 GO:0015450 GO:0005788 GO:0034663 GO:0005615 GO:0005634 GO:0016020	
			XXXXX	Group BiP2: *C. casuarinus* *C. halli* *S. longicornis*	Heat shock 70 kDa protein cognate 3	GO:0006457 GO:0140662 GO:0005524 GO:0005788	✓
Others (4 members)	*S. mansoni*	see Aguoru et al. (2022)	XXXXX XXXXX XXXXX XXXXX	Group 2: *C. casuarinus* *C. halli* *S. longicornis* *Kapentagyrus* spp.	Heat shock protein 70	GO:0006457 GO:0140662 GO:0005524 GO:0031072 GO:0016887 GO:0005783 GO:0005634 GO:0005829 GO:0070013 GO:0032991 GO:0005886	
				Group 3: *C. casuarinus* *C. cirratus* *C. sclerosus* *C. thurstonae* *C. tilapiae* *C. zambezensis* *S. longicornis*	Heat shock protein 70	GO:0006457 GO:0006950 GO:0140662 GO:0005524 GO:0031072 GO:0019899 GO:0051082 GO:0016887 GO:0008180 GO:0005814 GO:0005829	✓

					Group 4a: All	Heat shock protein 70	GO:0005634 GO:0006457 GO:0006950 GO:0009266 GO:0140662 GO:0005524 GO:0031072 GO:0016887 GO:0005829 GO:0005634	
					Group 4b: *Cichlidogyrus* spp.	Heat shock protein 70	GO:0005634 GO:0006457 GO:0006950 GO:0140662 GO:0005524 GO:0031072 GO:0016887 GO:0005829 GO:0005634	
Heat shock protein 90 kDa	HSP 90-alpha isoform 2-like	*S. mansoni*	see Aguoru et al. (2022)	XXXXX	Group 1: all	Heat shock protein 83	GO:0005886 GO:0006457 GO:0050821 GO:0034605 GO:0140662 GO:0051082 GO:0016887 GO:0005524 GO:0097718 GO:0048471 GO:0005829 GO:0032991	✓
				XXXXX	Group 2: all	Endoplasmin	GO:0005886 GO:0006457 GO:0030433 GO:0031247 GO:0071318 GO:1903513 GO:0032527 GO:0001666 GO:0043666 GO:0043066 GO:0035889 GO:0006880	✓

							GO:0140662 GO:0051082 GO:0016887 GO:0005524 GO:0050750 GO:0019903 GO:0003723 GO:0005509 GO:0042470 GO:0005788 GO:0048471 GO:0030496 GO:0016529 GO:0005789 GO:0005829 GO:0097524 GO:0034663	
	TRAP1-like	*S. mansoni*	XP_018652104.1	—	—	—	—	—
	Endoplasmin precursor-like	*S. mansoni*	see Aguoru et al. (2022)	—	—	—	—	—
foraging (*for*)		*Drosophila melanogaster*	NP_001356955.1 NP_001356896.1 NP_001356892.1 NP_001334731.1 NP_001162858.1 NP_001014464.1 NP_995629.1 NP_995628.1 NP_995626.1 NP_599146.1 NP_477490.1 NP_477489.1 NP_477487.1	XXXXX	Group 1: all	cGMP-dependent protein kinase	GO:0006468 GO:0007165 GO:2001226 GO:0045794 GO:0071476 GO:0032332 GO:0032922 GO:0072659 GO:0018209 GO:1903829 GO:0004692 GO:0030553 GO:0051019 GO:0106310 GO:0005524 GO:0042802 GO:0031965 GO:0016324 GO:0005737	✓
				XXXXX	Group 2: all except *K. tanganicanus*	cGMP-dependent protein kinase	GO:0006468 GO:1904753 GO:0090331 GO:1904706	

	GO:0060087
	GO:0048668
	GO:0019934
	GO:0043087
	GO:0016358
	GO:0001764
	GO:0004692
	GO:0030553
	GO:0106310
	GO:0005524
	GO:0005246
	GO:0005737
	GO:0012505
	GO:0005886
	GO:0043231

6.2.4 Protein sequence assembly and paralog filtering

We assembled the exon sequences of the three target gene groups (43 genes) and the single-copy genes coding for the 137 and 479 orthologous groups of proteins. We applied an exon bait capturing approach as implemented in the pipeline *HybPiper* v2.0 (Johnson et al., 2016) that has previously been successfully applied to whole-genome and transcriptome sequencing short reads (Yuan et al., 2019; Layton et al., 2020).

For the target gene groups, the bait protein sequences of *C. casuarinus* were combined into a bait file for *HybPiper* to map the trimmed paired-end and unpaired reads of all analysed species against the bait sequences (Fig 2c). We also included the protein sequence of *cyp* of *S. mansoni* (Table 6.2) in the bait files as we found no match in the annotated putative protein database of *C. casuarinus*. The target files for the species tree were compiled as mentioned above. For the target files of both the target genes and the species tree, we used the protein sequences rather than the nucleotides sequences as the developers of *HybPiper* reported that gene assemblies improved when using the former (Johnson et al., 2016). In *HybPiper*, we used *DIAMOND* v2.0.15 for the rapid alignments of sequencing reads (Buchfink et al., 2021). The remaining pipeline was left at default parameters including the assembly through *SPAdes* v3.15.5 (Prjibelski et al., 2020), a contig alignment and stitching process through *Exonerate* v2.4.0 (Slater & Birney, 2005), and the flagging of potential paralogs and chimeric sequences during the *Exonerate* searches (Johnson et al., 2016).

The pooling approach during the library preparation 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, we performed four filtering steps:

(i) We applied a *BLAST+* search on all protein sequences that we assembled to exclude contaminants. Sequences with closest hits above 90% identity with non-flatworm sequences were excluded (Fig. 6.2d).

(ii) We performed phylogenetic analyses with all potential paralogous sequences 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. Groups that covered all species used in the present analysis were immediately

assembled into gene alignments for downstream analyses, using the main sequence assembled by *HybPiper*. Groups, for which only some species were captured as main orthologs by *HybPiper*, were included in a second run in *HybPiper* to detect the orthologs of these genes. 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 sequences were again filtered through the LCA approach.

(iii) We only excluded orthologous groups of putative protein sequences if these covered less than three species to further minimise the effects of variation in the sampled populations in each DNA read pool.

(iv) We checked whether alignments of gene models not targeted with the bait sequences (paralogs suggested by *HybPiper*) represented fragments of other assembled gene models. These sequence fragments (truncated sequences) were excluded from downstream analyses.

Following the filtering steps, we inferred functional descriptions and gene ontology (GO) classes for the remaining sequences using *PANNZER2* (Törönen & Holm, 2022) for each orthologous group (Fig. 6.2e). These GO terms were only included in Table 6.2 if the annotations were assigned to orthologs in three or more species to avoid less reliable terms. We also verified the presence of the putative protein sequences through a *BLAST+* search (*tblastn*) in a recently published transcriptome annotation of *S. longicornis* (Caña-Bozada et al., 2022). We interpreted sequence identities and query coverage > 95% as confirmation of presence in the transcriptome (Fig. 6.2e).

In case of the single-copy orthologs, we excluded any putative protein alignments, for which paralogs were flagged in *HybPiper*. We took this approach to minimise the risks of accidentally including any contaminant sequences. We also excluded genes, for which sequences were not recovered from all 11 species included in the present analysis to minimise the impact of missing data on the species tree.

6.2.5 Phylogenetic analyses

We performed phylogenetic analyses for three different sets of trees: gene family trees for sequence filtering and paralog identification (e.g. GSTM and

HSP70), gene trees (for each of the 43 group of orthologs for the targeted gene groups), and species trees (based on OMA and BUSCO orthologs, 277 and 86 loci respectively). Phylogenetic analyses were performed under the maximum likelihood (ML) criterion using the protein sequences rather than nucleotide sequences as we were interested in functional differences between the species (and not silent nucleotide substitutions). Furthermore, the higher stability of protein sequences means these genes are more suitable to infer phylogenetic relationships between proteins of the same gene family (e.g. Aguoru et al., 2022). The protein sequences of the genes were aligned through *L-INS-i* in *MAFFT* v7.427 (Katoh & Standley, 2013) and trimmed in *Gblocks* v0.91b under less stringent parameters (option: *-b2=6 -b3=10 -b4=5 -b5=h*) (Talavera & Castresana, 2007). For the gene family trees, we did not trim the alignments as many informative sites would be removed by *Gblocks* due to relatively high divergence between genes of the same gene family. Substitution models for amino acids were selected by gene through *ModelFinder* in *IQ-Tree* (Kalyaanamoorthy et al., 2017). We estimated tree topologies through *IQ-Tree* v2.2.0 (Nguyen et al., 2015; Minh et al., 2020), and branch support through ultrafast bootstrap estimation (Hoang et al., 2018) and Shimodaira-Hasegawa–like approximate likelihood ratio tests (SH-aLRT) (Guindon et al., 2010) with 1000 replicates. We considered nodes with an ultrafast bootstrap value (UFBoot) ≥ 95 and a SH-aLRT statistic ≥ 80 as well-supported. Phylogenetic trees were visualised through the packages *ggplot2* v3.4.1 (Wickham, 2016) and *ggtree* v3.6.2 (Yu et al., 2017, 2018) in *R* v4.2.2 (R Core Team, 2022).

6.2.6 Comparison of gene vs. species tree topologies

We employed two approaches to assess topological differences of the gene trees (oxidative stress, heat shock, and foraging genes) and the species tree: 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 monogeneans 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 oxidative stress, heat shock, and *for* genes deviate from the species trees of *Cichlidogyrus* using multi-dimensional scaling (MDS) based on unweighted Robinson-Foulds 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. As the focus was not the comparison between the two monogenean genera, sequences of *Kapentagyrus* were dropped from the gene trees for this purpose using the function *drop.tip* in the *R* package *ape* v5.7-1 (Paradis & Schliep, 2019). Finally, we performed a multidimensional scaling (MDS) approach implemented in the *R* package *treespace* v1.1.4.2 (Jombart et al., 2017). The results were visualised in *ggplot2*.

6.3 Results

6.3.1 Species trees

After the exclusion of all potential multi-copy protein sequences and those with missing data for one or more species, we assembled two datasets of 277 (OMA) and 86 (BUSCO) single-copy orthologous sequence alignments, respectively. The subsequent phylogenomic analyses of the two datasets produced rather similar tree topologies (Fig. 6.3). As expected, species of *Cichlidogyrus* and *Kapentagyrus* form well-supported (OMA–BUSCO: 100/100–100/100) monophyletic groups. Furthermore, both trees support two major lineages within *Cichlidogyrus*: *C. casuarinus, C. sclerosus*, *C.* sp. 'kapembwa' (Clade A), and *C. cirratus*, *C. tilapiae*, *C. thurstonae*, *S. longicornis*, and *C. zambezensis* (Clade B). The phylogenetic position of *C. halli* remains unresolved although the *OMA* phylogeny (Fig. 6.3b) suggests that *C. halli* forms part of an early diverging lineage, with Clade A and B clustering monophyletically with moderate support (*/99). In Clade A, *C. sclerosus* forms the sister taxon to *C. casuarinus* and *C.* sp. 'kapembwa' (100/100–100/100). In Clade B, we detected three well supported clades: *C. cirratus* and *C. tilapiae* (98/100–*/100), *C. thurstonae* and *S. longicornis* (*/99–100/100), and *C. zambezensis*. The relationship between these clades remains unresolved with

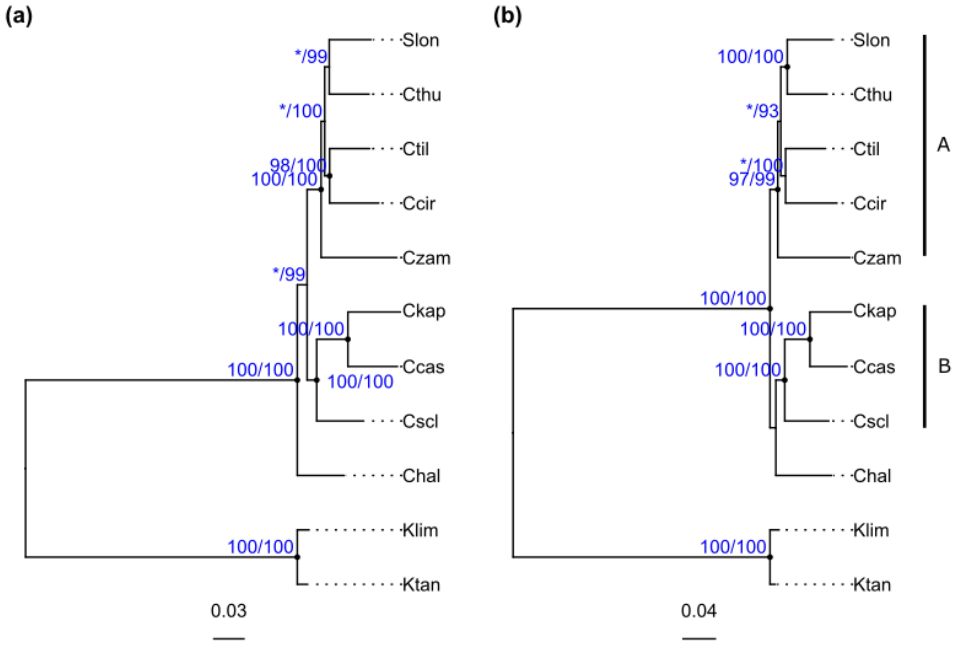


Figure 6.3. Species trees of *Cichlidogyrus* inferred from sequences of orthologs of single-copy genes selected by Caña-Bozada et al. (2023) through (a) the OMA (277 genes) and (b) the BUSCO (86 genes) pipelines. Abbreviations: Ccas–*Cichlidogyrus casuarinus*, Ccir–*C. cirratus*, Chal–*C. halli*, Ckap–*C.* sp. 'kapembwa', Cscl–*C. sclerosus*, Cthu–*C. thurstonae*, Ctil–*C. tilapiae*, Czam–*C. zambezensis*, Klim–*Kapentagyrus limnotrissae*, Ktan–*K. tanganicanus*, A–Clade A, B–Clade B.

only moderate support for *C. zambezensis* as sister group to the other two clades (82/93–88/100). Nonetheless, the species tree resolved almost all relationships of the main lineages of *Cichlidogyrus* that nuclear ribosomal and mitochondrial DNA markers were unable to resolve (Chapter 2; Rahmouni et al., 2022).

6.3.2 Gene copy numbers

The putative protein sequences of all 43 target genes included functional groups of their respective gene family and subfamilies both inferred from *PANNZER2*. For a full overview of the annotations, see Table 6.2. A majority (67%) of the putative protein sequences matched with the transcriptome data from *S. longicornis* (> 95% identity and query coverage) (Table 6.2). However, only three out of nine of the putative HSP70 variants and none of the GPx sequences were detected. In general, we found no species-specific differences in stress gene copy numbers detectable through the visual inspections of the gene family trees.

In the annotations of the genome of *C. casuarinus*, we detected deviations in terms of copy numbers from the search sequences of the other non-monogenean flatworms. Specifically, we found two copies of GPx (+1 as compared to *S. mansoni*), four copies of GSTM (-6 as compared to *E. multilocularis*), and two copies of Prx (-1 as compared to *S. mansoni*). No hits were found for CYP and GSTS (Table 6.2). For most bait sequences, we found only a single hit in the genome read pools. However, *gpx*, *gstm*, *prx*, and all the heat shock protein families except for *hsp10* were flagged by *HybPiper* for potential paralogs (results not shown). Phylogenetic analyses of these genes revealed that, for *hsp40* and *hsp60*, these paralogous sequences clustered with protein sequences from the same species and, therefore, could be excluded from downstream phylogenetic analyses. We detected one cluster of *gpx* and *hsp90* each. An alignment of these sequences with the main target sequences revealed that the former were fragments (truncated versions) of the latter. These truncated sequences were also excluded from downstream analyses.

For the *hsp70* subfamily, two of the bait sequences resulted in single hits. For the remaining bait sequences, a multitude of paralogs were flagged by *HybPiper*. A phylogenetic analysis revealed a complex array of well-supported clusters of *hsp70* genes. After the removal of contaminant sequences and sequences with orthologs in less than three species, we detected overall eight well-supported clades of sequences of putative HSP70 (Fig. 6.4). *HybPiper* failed to assemble orthologs for species of *Kapentagyrus* in three of the clades. Three groups were assigned highly specific functions: hypoxia up-regulated 1 (HYOU1) and endoplasmatic reticulum chaperone binding protein 1 (BiP1) and 2 (BiP2). Notably, group 4 consisted of two subgroups for species of *Cichlidogyrus*, but only a single subgroup for species of *Kapentagyrus*. However, the relationship between these subgroups was not resolved (Fig. 6.4).

For *gstm*, we detected three major sequence clusters, which did, however, group differently than the four sequences of *C. casuarinus*. The assembly of orthologs and paralogs in different species based on the same bait sequence suggests a high similarity of these genes. The resulting sequences were assigned to groups according to the three clades inferred from the phylogenetic tree of

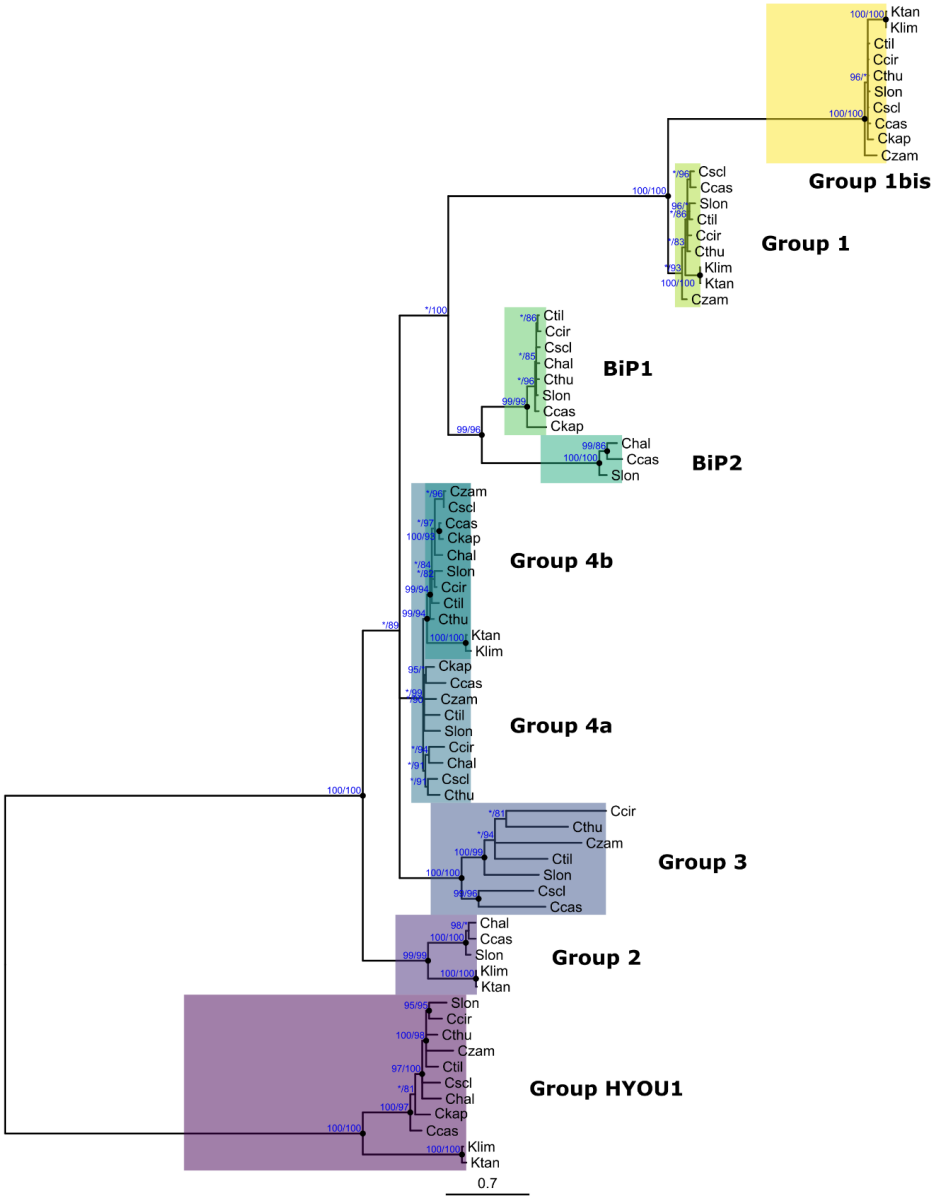


Figure 6.4. Maximum likelihood phylogram of gene models of the 70 kDa heat shock protein family (HSP70) of species of *Cichlidogyrus* and *Kapentagyrus*. For abbreviation of species names, see Fig. 6.3. Group HYOU1 (hypoxia up-regulated 1) and BiP1 (endoplasmatic reticulum chaperone binding protein 1) and BiP2 refer to annotations assigned through *PANNZER2* (see Table 6.2). The remaining groups are numbered consecutively. Group 4 shows a potential duplication event (or gene loss) with two copies of the gene for species of *Cichlidogyrus* but only a single one for species of *Kapentagyrus*.

gstm (Fig. 6.5). Notably, Group 2 includes two copies for most species of

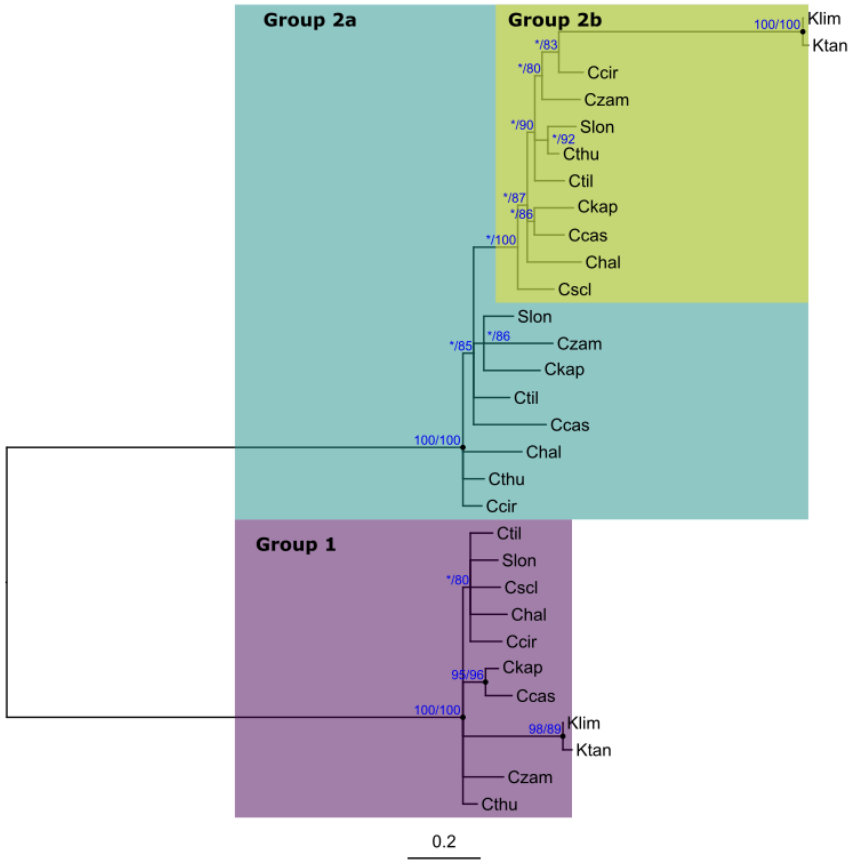


Figure 6.5. Maximum likelihood phylogram of gene models of the glutathione *S*-transferase mu-class (GSTM) of species of *Cichlidogyrus* and *Kapentagyrus*. For abbreviation of species names, see Fig. 6.3. Group 2 shows a potential duplication event (or gene loss) with two copies of the gene for species of *Cichlidogyrus* but only a single one for species of *Kapentagyrus*.

Cichlidogyrus with identical gene ontology (GO) terms (Table 6.2) but only one for *Kapentagyrus* (Fig. 6.5). We consider the absence of one copy in *Kapentagyrus* as informative as the high sequence similarities of Group 2a and b suggests that the ortholog of Group 2a should have been detected through *DIAMOND* if present.

6.3.3 Gene tree topologies

The species trees had a high level of similarity with each other, effectively overlapping in the MDS plot (Fig. 6.6). The gene trees (Appendix S6.2) of the three studied gene groups (oxidative stress, heat shock, foraging) deviated from

this topology. Although the gene trees showed some variation, their topologies formed no apparent clusters (Fig. 6).

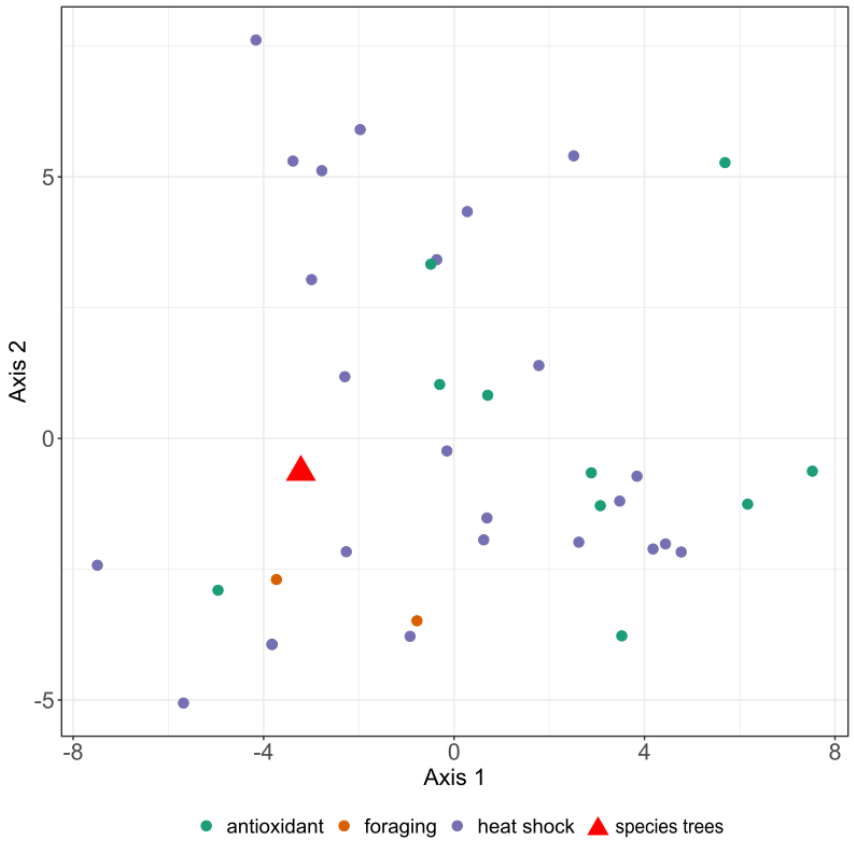


Figure 6.6. First two axes of multidimensional scaling analysis of gene trees of antioxidant enzymes, *for* orthologs, and heat shock proteins (see Appendix S1). The genes tree topologies deviate from the two species trees (highlighted in red), but show no apparent pattern or clustering.

6.4 Discussion

We compared the putative adaptive potential of two metazoan parasite lineages—a species-rich one (*Cichlidogyrus*) and a species-poor one (*Kapentagyrus*)—based on three groups of stress response genes. For this purpose, we performed a comparative genomic analysis of 43 antioxidant enzymes, heat shock proteins, and *foraging* gene ortholog sequences. The expression of 67% of these genes was confirmed through a previously published

transcriptome annotation (Caña-Bozada et al., 2022) and 41 genes included functional groups of their expected gene family (Table 6.2). This study is an innovative approach to test whether stress responses are related to adaptive potentials of parasite lineages as it uses a multi-species study system [i.e. more than three species, e.g. Aguoru et al. (2022)]. With nine species of *Cichlidogyrus*, the study provides the most species-rich set of whole-genome sequencing data from a single genus of flatworms to date, followed only by the blood fluke genus *Schistosoma* with eight species (Ebbs et al., 2022). Furthermore, with 68/277 single-copy ortholog and 43 stress genes, this study is by far the most extensive genomic analysis of monogenean flatworms. Therefore, our study and dataset present important contributions for comparative genomics on parasitic flatworms as well as metazoan parasites in general.

6.4.1 Potential gene expansions, contractions, and losses in monogeneans

The apparent absence of the cytochrome P450 gene family (*cyp*) in monogenean genomes is remarkable. Copies were found neither in the annotated genome of *C. casuarinus* (Hahn, unpublished), the transcriptome of *S. longicornis* (Caña-Bozada et al., 2022) nor in the genome reads of 11 species in the present study. Furthermore, no *cyp* orthologs have been reported for the genomes of other monogenean species published to date including those of *Benedenia humboldti* Baeza, Sepúlveda & González, 2019 (Baeza & González, 2021), *Eudiplozoon nipponicum* (Goto, 1891) (Vorel et al., 2023), *Gyrodactylus bullatarudis* Turnbull, 1956 (Konczal et al., 2020), and *G. salaris* Malmberg, 1957 (Hahn et al., 2014). The absence of *cyp* in all species analysed so far cannot be interpreted as an artefact of low sequencing coverage due to the success in capturing 363 likely single-copy orthologs (see also estimated coverages in Table 6.1). We suggest that either the DNA/amino acid sequences are highly divergent from *cyp*/CYP sequences of other metazoan parasites, such as was reported for a proposed peroxin gene in *Trypanosoma brucei* (Saveria et al., 2007), or *cyp* is indeed absent in monogeneans. Both cases represent absolute novelties as *cyp* is conserved across almost the entire tree of life (Nelson, 2018). The loss of *cyp* was previously only reported from protozoan parasites, such as *Plasmodium falciparum* (Welch, 1897) (Wisedpanichkij et al., 2011), *Giardia intestinalis* (Lambl, 1859) (Pazdzior et al., 2015), and *Encephalitozoon cuniculi* Levaditi, Nicolau & Schoen, 1923

(Shaheen et al., 2020). In other flatworms, e.g. flukes (Pakharukova et al., 2015; Ziniel et al., 2015) and cestodes (Tsai et al., 2013), only a single variant of *cyp* was reported, unlike the multiple copies that are usually present in most organisms (Nelson, 2018), but functional analyses highlighted that *cyp* still fulfils vital functions in flukes and cestodes, e.g. the detoxification of xenobiotics (Pakharukova et al., 2015; Ziniel et al., 2015). The gene family has also been reduced in parasitic crustaceans, e.g. the salmon louse has the lowest known number of *cyp* copies of any arthropod (Skern-Mauritzen et al., 2021), suggesting a trend of contracting the *cyp* gene superfamily in metazoan parasites. Evolutionary loss of genes and gene functions is a phenomenon that has also been observed elsewhere in parasites, e.g. peroxisomal functions in parasitic protozoans, flatworms, roundworms (Žárský & Tachezy, 2015), and crustaceans (Skern-Mauritzen et al., 2021). Therefore, we suggest that *cyp* gene family contraction and loss should be further investigated to understand the magnitude of this trend in parasites.

We reported multiple gene orthologs in flatworms and monogeneans for the first time. We detected two gene copies of glutathione peroxidase (GPx), two of peroxiredoxin (Prx), and three of cytosolic glutathione *S*-transferase (cGST) from the mu-class (Table 6.1). All of the copy numbers differ from previously studied flatworms with cestodes and trematodes presenting a single *gpx*, three *prx*, and around 12 cytosolic *gst* copies (Martínez-González et al., 2022). As the functions of these antioxidant enzymes are the reduction of hydrogen peroxide to water (GPx, Prx) and alkyl hydroperoxides to alcohol (Prx) as well as detoxification (cGST), these gene family contractions/expansions could provide valuable insight into the functional evolution of parasitic flatworms. For instance, increased *gpx* copy numbers were linked to lineages exposed to higher levels of oxidative stress in mammals (Tian et al., 2021). Furthermore, we detected three *cgst* copies of the mu-class (see detailed discussion below) but none of the sigma-class. The absence of omega class *gst* in cestodes was speculated to result from 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). The discussed examples for copy number variation in *cyp*, *gpx*, *prx*, and *gst* in other organisms highlight the potential importance of stress response pathways for parasite evolutionary biology.

We also provide the first report of orthologs of the *foraging* gene in flatworms, a gene linked to behavioural traits in arthropods, mammals, amphibians, and nematodes. Although fecundity and 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 untested. We suggest that the importance of *for* and PKGs in other organisms and its potential role in driving infection intensities in parasitic flatworms (Huyse et al., 2018) warrant further studies into this gene family in these organisms to better understand their behavioural genetics.

6.4.2 How they differ: do species of Cichlidogyrus *have a higher adaptive potential than species of* Kapentagyrus*?*

Differences between species of *Cichlidogyrus* and *Kapentagyrus* in copy numbers of stress response genes might provide insight into their diverging ecology and evolutionary history, specifically their adaptive potential. Based on a visual inspection of the phylogenies of the gene families, we identified two potential instances of either gene duplication or loss. Species of *Kapentagyrus* present at least one less gene copy of the glutathione *S*-transferase mu-class (*gstm*) and the 70-kDA heat shock protein family (*hsp70*), which indicates either gene duplication in species of *Cichlidogyrus* or gene loss in species of *Kapentagyrus*. Higher copy numbers in species of *Cichlidogyrus* might suggest a higher potential to adapt to adverse conditions. Expansions of the *gst* gene family in closely related organisms were previously also observed in free-living nematodes (Markov et al., 2015). In fungi, gene expansion, e.g. of *cyp*, was linked to habitat/host diversity (Coleman et al., 2009; Wang et al., 2018) and, in invertebrates, the proportion of gene duplication in genomes was linked to the invasive potential (Liu et al., 2018; 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 such knowledge requires a detailed understanding of gene functions. Notable exceptions include the plant-pathogenic moths of the *Spodoptera frugiperda* (Smith, 1797) 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* (Steiner & Buhrer, 1934), 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 and extremophile bivalves (Guerin et al., 2019) and invasive fishes (Stanley et al., 2022). In parasites, *hsp70* expansions within parasite lineages have been reported for tapeworms (Tsai et al., 2013) and trypanosomatid protozoans (Drini et al., 2016). Tsai et al. (2013) theorised that the expanded gene copies of the tapeworms are expressed only 'under certain conditions' based on their low expression levels in their transcriptomes, but did not elaborate further. Furthermore, Drini et al. (2016) linked copy number variations in strains of *Leishmania* Ross, 1903 to geographical and environmental gradients. Yet similar to *gst*, *hsp70* copy numbers alone provide no definitive evidence of environmental adaptation without more detailed knowledge about gene functions. For instance, some cold-adapted Antarctic organisms have lost their inducible heat shock responses despite the presence of *hsp70* (Clark et al., 2008; Bilyk et al., 2018). Nevertheless, *Cichlidogyrus* and *Kapentagyrus* provide the first evidence of divergent stress response pathways in flatworm lineages of the same family.

Although the function of the expanded gene families remains unknown, the ecology and evolutionary history of species of *Cichlidogyrus* and *Kapentagyrus* provide ample arguments why the observed copy number differences might be related to their adaptive potential. *Cichlidogyrus* is a highly diverse monogenean lineage that infects African cichlid fishes, which cover a range of ecological niches including herbivores, invertivores, piscivores, and omnivores (Burress, 2014) and inhabit lacustrine, riverine, and estuarine habitats across Africa (Chapman, 2021). This ecological diversity exposes cichlid ectoparasites to a series of environmental stressors. In contrast, the hosts of species of *Kapentagyrus* are freshwater clupeids. These host fishes only occupy the pelagic zone of African rivers and lakes (Wilson et al., 2008) and are, therefore, much more ecologically conserved. This niche conservatism likely creates a relatively stable environment for their parasites, which is why recent studies found only a weak population structure in both hosts and parasites (De Keyzer et al. 2019; Kmentová et al. 2020). It should

be noted that these studies and the present study only included species of *Kapentagyrus* from Lake Tanganyika, thus, omitting all riverine species (Vanhove et al., 2021). These other species should be sequenced to infer whether the detected copy number differences are a general patterns between *Cichlidogyrus* and *Kapentagyrus* or specific to species from Lake Tanganyika.

6.4.3 Limitations: species-specific variation and gene expression

With the expansion and deletion of gene families, our study reveals ground-breaking new insights into the molecular mechanisms that underpin the adaptive potential of parasites—a concept that might help us understand the processes of host selection and host switching. However, the study also has conceptual and technological limitations.

First, we detected no species-specific differences in stress gene copy numbers. 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. We suggest that our current approach might not be able to detect species-specific differences in terms of copy numbers as the LCA approach inherently neglects any species-specific duplication events. This approach was taken to avoid mistaking gene variants of populations in each pooled DNA samples as paralogs. Future studies might address this problem through variant calling pipelines [e.g. Kofler et al. (2011)] or by optimising techniques to sequence genomes from individual specimens. The latter approach has recently been successful with monogenean flatworms for the first time (Geraerts et al., 2022a) and recent advances in sequencing technology for medical research means that whole genomes are now sequenced even from single cells (Ogbeide et al., 2022), albeit limited to model organisms (Alfieri et al., 2022). Another issue lies in potentially highly divergent sequences that the exon bait capture approach might fail to detect (see discussion for CYP above). Beyond gene copy numbers, we also found that the evolutionary relationship of the gene orthologs presented some variation compared to the evolutionary history of the species (Fig. 6.5), but this variation might present an artefact of inferring evolutionary histories from small datasets (81–937 amino acids) in contrast to multi-gene phylogenies (93,000 and 26,000 amino acids). Furthermore, we only covered 9/144 described species of *Cichlidogyrus*. Nevertheless, species of

Cichlidogyrus constitute a so far unique study system for host-parasite interactions that combines opportunities to investigate divergent host repertoires (Cruz-Laufer et al., 2022), biological invasions (Jiménez-García et al., 2001; Šimková et al., 2019; Jorissen et al., 2020; Geraerts et al., 2022a), co-infections (Gobbin et al., 2021), host switching (Birgi & Euzet, 1983; Birgi & Lambert, 1986), and divergent speciation rates (Vanhove et al., 2015; Kmentová et al., 2016). Therefore, we suggest that whole-genome sequencing should be applied to a broader selection of species in the future to test differences in stress response pathways through a higher taxonomic resolution.

Second, gene models reveal no information on whether they are expressed or even inducible. In the present study, we reported multiple gene orthologs in flatworms and monogeneans for the first time. Yet detected orthologs of all three targeted gene groups might not necessarily be expressed. Moreover, some genes are only expressed under certain conditions (i.e. inducible genes) that might have not been met by methods employed for the transcriptome we used as a reference (Caña-Bozada et al., 2022), which might explain the absence of many HSP70 and GPx transcripts (Table 6.2). Lastly, even if certain genes are found in transcriptomes, environmental stress might not necessarily lead to their up-regulation (see *hsp* in Antarctic animals: Clark et al., 2008; Bilyk et al., 2018). Therefore, the next step for testing stress responses in metazoan parasites is obtaining genome and transcriptome assemblies that are representative of the functional diversity within parasite lineages. Future studies should also aim to quantify expression of stress response genes under different environmental stressors using experimental studies. These steps could provide key information on the functionality of the gene models assembled for the two flatworm lineages in this study.

Concluding remarks

Stress responses are key for the survival of organisms, yet their role in parasite evolution and adaptive potential remains poorly understood. A strong bias in parasite genomics towards few health-relevant human pathogens results in a lack of DNA or RNA sequence data in closely related and functionally diverse parasite lineages. The present study addresses this bias by analysing the stress response gene presence and copy number variation in 11 genomes of two genera of

monogenean flatworms, *Cichlidogyrus* and *Kapentagyrus*. Even though these data only represent a fraction of the known species from these lineages, we detected several cases of copy number differences, suggesting that alterations in stress response pathways are indeed an important 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.

Acknowledgements

We appreciate the help of the people involved in the field work and sampling procedure including Fidel Muterezi Bukinga, Clément Kalombo Kabalika, Stephan Koblmüller, Holger Zimmermann, and several fishermen. We thank Christoph Hahn for providing access to the annotated genome of *C. casuarinus*. We thank Natascha Steffanie (Hasselt University) for technical support in the laboratory. We thank Leona Milec (Hasselt University, Nord University) and Geert-Jan Bex (Hasselt University) for assistance with the use of the high-performance computing cluster. 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. Molecular work was funded through research grant 1513419 N of the Research Foundation—Flanders (FWO-Vlaanderen). AJCL (BOF19OWB02), MPMV (BOF20TT06), MG (BOF17NI02), and NK (BOF21PD01) are funded by the Special Research Fund of Hasselt University. Parts of the resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation – Flanders (FWO) and the Flemish Government.

Author contributions

AJCL conceptualised the study under supervision of NK and MPMV. AJCL, MG, MB, HB, CFBB, ARBM, ACM, GKK, SM, 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 amplification. AJCL performed all statistical analyses and produced all graphs with input from NK and LB. AJCL and NK wrote the manuscript with input from MPMV, MG, KS, LB, MB, HB, CFBB, ARBM, ACM, TH, GKK, and SM.

Data accessibility

BioProject IDs and GenBank accession numbers of the sequencing data supporting this manuscript will be assigned upon publication of the peer-reviewed version.

References

Abou-El-Naga, I. F., 2020. Heat shock protein 70 (Hsp70) in *Schistosoma mansoni* and its role in decreased adult worm sensitivity to praziquantel. Parasitology 147: 634–642.

Aguileta, G., S. Marthey, H. Chiapello, M. H. Lebrun, F. Rodolphe, E. Fournier, A. Gendrault-Jacquemard, & T. Giraud, 2008. Assessing the performance of single-copy genes for recovering robust phylogenies. Systematic Biology 57: 613–627.

Aguoru, N. A., R. S. Kirk, & A. J. Walker, 2022. Molecular insights into the heat shock proteins of the human parasitic blood fluke *Schistosoma mansoni*. Parasites and Vectors 15: 365.

Alfieri, J. M., G. Wang, M. M. Jonika, C. A. Gill, H. Blackmon, & G. N. Athrey, 2022. A primer for single-cell sequencing in non-model organisms. Genes 13: 380.

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

Anreiter, I., & M. B. Sokolowski, 2019. The *foraging* gene and its behavioral effects: pleiotropy and plasticity. Annual Review of Genetics 53: 373–392.

Assane, I. M., K. D. Prada-Mejia, S. U. Gallani, N. F. Weiser, G. M. R. Valladão, & F. Pilarski, 2022. *Enterogyrus* spp. (Monogenea: Ancyrocephalinae) and *Aeromonas jandaei* co-infection associated with high mortality following transport stress in cultured Nile tilapia. Transboundary and Emerging Diseases 69: e276–e287.

Baeza, J. A., & M. T. González, 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. Marine Genomics 58: 100848.

Bateman, A., M. J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. G. Ebenezer, J. Fan, P. Garmiri, L. J. da Costa Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasaamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D. L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Totoo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. J. Bridge, L. Aimo, G. Argoud-Puy, A. H. Auchincloss, K. B. Axelsen, P. Bansal, D. Baratin, T. M. Batista Neto, M. C. Blatter,

J. T. Bolleman, E. Boutet, L. Breuza, B. C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuche, E. de Castro, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C. J. A. Sigrist, K. Sonesson, S. Sundaram, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, & J. Zhang, 2023. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research 51: D523–D531.

Bijlsma, R., & V. Loeschcke, 2005. Environmental stress, adaptation and evolution: an overview. Journal of Evolutionary Biology 18: 744–749.

Bilyk, K. T., L. Vargas-Chacoff, & C. H. C. Cheng, 2018. Evolution in chronic cold: varied loss of cellular response to heat in Antarctic notothenioid fish. BMC Evolutionary Biology 18: 143.

Birgi, E., & L. Euzet, 1983. Monogènes parasites des poissons des eaux douces du Cameroun. Présence des genres *Cichlidogyrus* et *Dactylogyrus* chez *Aphyosemion* (Cyprinodontidae). Bulletin de la Société Zoologique de France 108: 101–106.

Birgi, E., & A. Lambert, 1986. Présence chez un Nandidae (Téléostéen), *Polycentropsis abbreviata* Boulenger, 1901, du genre *Cichlidogyrus* (Monogenea, Monopisthocotylea, Ancyrocephalidae). Annales de Parasitologie Humaine et Comparée 61: 521–528.

Bolger, A. M., M. Lohse, & B. Usadel, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Borchel, A., A. Z. Komisarczuk, A. Rebl, T. Goldammer, & F. Nilsen, 2018. Systematic identification and characterization of stress-inducible heat shock proteins (HSPs) in the salmon louse (*Lepeophtheirus salmonis*). Cell Stress and Chaperones 23: 127–139.

Brooks, D. R., E. P. Hoberg, & W. A. Boeger, 2019. The Stockholm Paradigm: climate change and emerging disease. University of Chicago Press, Chicago, IL, USA.

Buchfink, B., K. Reuter, & H. G. Drost, 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods 18: 366–368.

Burress, E. D., 2014. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. Hydrobiologia 748: 7–27.

Busby, A. T., N. Ayllón, K. M. Kocan, E. F. Blouin, G. de La Fuente, R. C. Galindo, M. Villar, & J. de la Fuente, 2012. Expression of heat shock proteins and subolesin affects stress responses, *Anaplasma phagocytophilum* infection and questing behaviour in the tick, *Ixodes scapularis*. Medical and Veterinary Entomology 26: 92–102.

Calderwood, S. K., E. A. Repasky, L. Neckers, & L. E. Hightower, 2019. The IXth CSSI international symposium on heat shock proteins in biology and medicine: stress responses in health and disease: Alexandria Old Town, Alexandria, Virginia, November 10–13, 2018. Cell Stress and Chaperones 24: 1–6.

Calla, B., K. Noble, R. M. Johnson, K. K. O. Walden, M. A. Schuler, H. M. Robertson, & M. R. Berenbaum, 2017. Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. Molecular Ecology 26: 6021–6035.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, & T. L. Madden, 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Caña-Bozada, V., R. Llera-Herrera, E. J. Fajer-Ávila, & F. N. Morales-Serna, 2021. Mitochondrial genome of *Scutogyrus longicornis* (Monogenea: Dactylogyridea), a parasite of Nile tilapia *Oreochromis niloticus*. Parasitology International 81: 102281.

Caña-Bozada, V., F. N. Morales-Serna, E. J. Fajer-Ávila, & R. Llera-Herrera, 2022. De novo transcriptome assembly and identification of G-Protein-Coupled-Receptors (GPCRs) in two species of monogenean parasites of fish. Parasite 29: 51.

Caña-Bozada, V., M. W. Robinson, D. I. Hernández-Mena, & F. N. Morales-Serna, 2023. Exploring evolutionary relationships within Neodermata using putative orthologous groups of proteins, with emphasis on peptidases. Tropical Medicine and Infectious Disease 8: 59.

Carlson, C. J., K. R. Burgio, E. R. Dougherty, A. J. Phillips, V. M. Bueno, C. F. Clements, G. Castaldo, T. A. Dallas, C. A. Cizauskas, G. S. Cumming, J. Doña, N. C. Harris, R. Jovani, S. Mironov, O. C. Muellerklein, H. C. Proctor, & W. M. Getz, 2017. Parasite biodiversity faces extinction and redistribution in a changing climate. Science Advances 3: e1602422.

Chapman, L. J., 2021. Respiratory ecology of cichlids In Abate, M. E., & D. L. G. Noakes (eds), The behavior, ecology and evolution of cichlid fishes. Springer Nature, Dordrecht, The Netherlands: 683–714.

Chardonnet, F., C. Capdevielle-Dulac, B. Chouquet, N. Joly, M. Harry, B. Le Ru, J. F. Silvain, & L. Kaiser, 2014. Food searching behaviour of a Lepidoptera pest species is modulated by the *foraging* gene polymorphism. Journal of Experimental Biology 217: 3465–3473.

Chen, H. Y., Y. S. Cheng, D. S. Grabner, S. H. Chang, & H. H. Shih, 2014. Effect of different temperatures on the expression of the newly characterized heat shock protein 90 (Hsp90) in L3 of *Anisakis* spp. isolated from *Scomber australasicus*. Veterinary Parasitology 205: 540–550.

Chiumiento, L., & F. Bruschi, 2009. Enzymatic antioxidant systems in helminth parasites. Parasitology Research 105: 593–603.

Choi, Y. J., S. Fontenla, P. U. Fischer, T. H. Le, A. Costábile, D. Blair, P. J. Brindley, J. F. Tort, M. M. Cabada, & M. Mitreva, 2020. Adaptive radiation of the flukes of the family Fasciolidae inferred from genome-wide comparisons of key species. Molecular Biology and Evolution 37: 84–99.

Clark, M. S., K. P. P. Fraser, & L. S. Peck, 2008. Lack of an HSP70 heat shock response in two Antarctic marine invertebrates. Polar Biology 31: 1059–1065.

Coleman, J. J., S. D. Rounsley, M. Rodriguez-Carres, A. Kuo, C. C. Wasmann, J. Grimwood, J. Schmutz, M. Taga, G. J. White, S. Zhou, D. C. Schwartz, M. Freitag, L. J. Ma, E. G. J. Danchin, B. Henrissat, P. M. Coutinho, D. R. Nelson, D. Straney, C. A. Napoli, B. M. Barker, M. Gribskov, M. Rep, S. Kroken, I. Molnár, C. Rensing, J. C. Kennell, J. Zamora, M. L. Farman, E. U. Selker, A. Salamov, H. Shapiro, J. Pangilinan, E. Lindquist, C. Lamers, I. V. Grigoriev, D. M. Geiser, S. F. Covert, E. Temporini, & H. D. Vanetten, 2009. The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. PLoS Genetics 5: e1000618.

Cruz-Laufer, A. J., T. Artois, S. Koblmüller, A. Pariselle, K. Smeets, M. Van Steenberge, & M. P. M. Vanhove, 2022. Explosive networking: the role of adaptive host radiations and ecological opportunity in a species-rich host–parasite assembly. Ecology Letters 25: 1795–1812.

Cruz-Laufer, A. J., A. Pariselle, M. W. P. Jorissen, F. Muterezi Bukinga, A. Al Assadi, M. Van Steenberge, S. Koblmüller, C. Sturmbauer, K. Smeets, T. Huyse, T. Artois, & M. P. M. Vanhove, 2022. Somewhere I belong: phylogeny and morphological evolution in a species-rich lineage of ectoparasitic flatworms infecting cichlid fishes. Cladistics 38: 465–512.

D'Bastiani, E., K. M. Campiaõ, W. A. Boeger, & S. B. L. Araújo, 2020. The role of ecological opportunity in shaping host-parasite networks. Parasitology 147: 1452–1460.

De Keyzer, E. L. R., Z. De Corte, M. Van Steenberge, J. A. M. Raeymaekers, F. C. F. Calboli, N. Kmentová, T. N'Sibula Mulimbwa, M. Virgilio, C. Vangestel, P. M. Mulungula, F. A. M. Volckaert, & M. P. M. Vanhove, 2019. First genomic study on Lake Tanganyika sprat *Stolothrissa tanganicae*: a lack of population structure calls for integrated management of this important fisheries target species. BMC Evolutionary Biology 19: 6.

Drini, S., A. Criscuolo, P. Lechat, H. Imamura, T. Skalický, N. Rachidi, J. Lukeš, J. C. Dujardin, & G. F. Späth, 2016. Species- and strain-specific adaptation of the HSP70 super family in pathogenic trypanosomatids. Genome Biology and Evolution 8: 1980–1995.

Durak, R., M. Molon, T. Durak, & G. Chrzanowski, 2018. The enzymatic markers of the adaptation of *Cinara tujafilina* to changing the host plant. Ethology Ecology and Evolution 30: 416–429.

Dylus, D., Y. Nevers, A. M. Altenhoff, A. Gürtler, C. Dessimoz, & N. M. Glover, 2020. How to build phylogenetic species trees with OMA. F1000Research 9: 511.

Ebbs, E. T., E. S. Loker, L. Bu, S. A. Locke, V. V. Tkach, R. Devkota, V. R. Flores, H. A. Pinto, & S. V. Brant, 2022. Phylogenomics and diversification of the Schistosomatidae based on targeted sequence capture of ultra-conserved elements. Pathogens 11: 769.

Eberle, J., D. Ahrens, C. Mayer, O. Niehuis, & B. Misof, 2020. A plea for standardized nuclear markers in metazoan DNA taxonomy. Trends in Ecology and Evolution 35: 336–345.

Gallot, A., E. Desouhant, V. Lhuillier, D. Lepetit, A. El Filali, L. Mouton, C. Vieira-Heddi, & I. Amat, 2022. The *for* gene as one of the drivers of foraging variations in a parasitic wasp. Molecular Ecology 32: 1760–1776.

Geraerts, M., T. Huyse, M. Barson, H. Bassirou, C. F. Bilong Bilong, A. R. Bitja-Nyom, A. Chocha Manda, A. J. Cruz-Laufer, C. Kalombo Kabalika, G. Kapepula Kasembele, F. Muterezi Bukinga, S. Njom, T. Artois, & M. P. M. Vanhove, 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.

Geraerts, M., T. Huyse, M. P. M. Vanhove, & T. Artois, 2022b. A new species of *Cichlidogyrus* Paperna, 1960 (Platyhelminthes: Monogenea: Dactylogyridae) infecting tilapias in Lake Kariba (Zimbabwe), with a discussion on its phylogenetic position. Systematics and Biodiversity 20: 2143594.

Geraerts, M., F. Muterezi Bukinga, M. P. M. Vanhove, A. Pariselle, A. Chocha Manda, E. J. Vreven, T. Huyse, & T. Artois, 2020. Six new species of *Cichlidogyrus* Paperna, 1960 (Platyhelminthes: Monogenea) from the gills of cichlids (Teleostei: Cichliformes) from the Lomami River Basin (DRC: Middle Congo). Parasites & Vectors 13: 187.

Gobbin, T. P., M. P. M. Vanhove, O. Seehausen, & M. E. Maan, 2021. Microhabitat distributions and species interactions of ectoparasites on the gills of cichlid fish in Lake Victoria, Tanzania. International Journal for Parasitology 51: 201–214.

Gouin, A., A. Bretaudeau, K. Nam, S. Gimenez, J. M. Aury, B. Duvic, F. Hilliou, N. Durand, N. Montagné, I. Darboux, S. Kuwar, T. Chertemps, D. Siaussat, A. Bretschneider, Y. Moné, S. J. Ahn, S. Hänniger, A. S. G. Grenet, D. Neunemann, F. Maumus, I. Luyten, K. Labadie, W. Xu, F. Koutroumpa, J. M. Escoubas, A. Llopis, M. Maïbèche-Coisne, F. Salasc, A. Tomar, A. R. Anderson, S. A. Khan, P. Dumas, M. Orsucci, J. Guy, C. Belser, A. Alberti, B. Noel, A. Couloux, J. Mercier, S. Nidelet, E. Dubois, N. Y. Liu, I. Boulogne, O. Mirabeau, G. Le Goff, K. Gordon, J. Oakeshott, F. L. Consoli, A. N. Volkoff, H. W. Fescemyer, J. H. Marden, D. S. Luthe, S. Herrero, D. G. Heckel, P. Wincker, G. J. Kergoat, J. Amselem, H. Quesneville, A. T. Groot, E. Jacquin-Joly, N. Nègre, C. Lemaitre, F. Legeai, E. D'Alençon, & P. Fournier, 2017. Two genomes of

highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. Scientific Reports 7: 11816.

Guerin, M. N., D. J. Weinstein, & J. R. Bracht, 2019. Stress adapted Mollusca and Nematoda exhibit convergently expanded Hsp70 and AIG1 gene families. Journal of Molecular Evolution 87: 289–297.

Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, & O. Gascuel, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59: 307–321.

Hahn, C., B. Fromm, & L. Bachmann, 2014. Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparastic neodermata. Genome Biology and Evolution 6: 1105–1117.

Harris, P. D., L. Bachmann, & T. A. Bakke, 2010. The parasites and pathogens of the Atlantic salmon: lessons from *Gyrodactylus salaris* Atlantic Salmon Ecology. Wiley-Blackwell, Oxford, UK: 221–252.

Hector, T. E., A.-L. M. Gehman, & K. C. King, 2023. Infection burdens and virulence under heat stress: ecological and evolutionary considerations. Philosophical Transactions of the Royal Society B: Biological Sciences 378: 20220018.

Hillis, D. M., T. A. Heath, & K. St. John, 2005. Analysis and visualization of tree space. Systematic Biology 54: 471–482.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, & S. Le Vinh, 2018. UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35: 518–522.

Hotez, P. J., S. Aksoy, P. J. Brindley, & S. Kamhawi, 2020. What constitutes a neglected tropical disease?. PLoS Neglected Tropical Diseases 14: e0008001.

Huyse, T., N. A. M. Boon, F. Van den Broeck, M. Mbow, A. Chaturvedi, L. Meurs, F. A. M. Volckaert, & K. Polman, 2018. Evolutionary epidemiology of schistosomiasis: linking parasite genetics with disease phenotype in humans. International Journal for Parasitology 48: 107–115.

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

Iriarte, A., P. Arbildi, S. La-Rocca, H. Musto, & V. Fernández, 2012. Identification of novel glutathione transferases in *Echinococcus granulosus*. An evolutionary perspective. Acta Tropica 123: 208–216.

Ismail, T. N. S. T., N. F. A. Kassim, A. A. Rahman, K. Yahya, & C. E. Webb, 2018. Day biting habits of mosquitoes associated with mangrove forests in Kedah, Malaysia. Tropical Medicine and Infectious Disease 3: 77.

Jiménez-García, M. I., V. M. Vidal-Martínez, & S. Lopez-Jiménez, 2001. Monogeneans in introduced and native cichlids in México: evidence for transfer. Journal of Parasitology 87: 907.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, & N. J. Wickett, 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. Applications in Plant Sciences 4: 1600016.

Jombart, T., M. Kendall, J. Almagro-Garcia, & C. Colijn, 2017. treespace: Statistical exploration of landscapes of phylogenetic trees. Molecular Ecology Resources 17: 1385–1392.

Jordan, A., B. Taborsky, & M. Taborsky, 2021. Cichlids as a model system for studying social behaviour and evolution In Abata, M. E., & D. L. G. Noakes (eds), The behavior, ecology and evolution of cichlid fishes. Springer Nature, Dordrecht, The Netherlands: 587–635.

Jorissen, M. W. P., T. Huyse, A. Pariselle, S. Wamuini Lunkayilakio, F. Muterezi Bukinga, A. Chocha Manda, G. Kapepula Kasembele, E. J. Vreven, J. Snoeks, E. Decru, T. Artois, & M. P. M. Vanhove, 2020. Historical museum collections help detect parasite species jumps after tilapia introductions in the Congo Basin. Biological Invasions 11: 1123.

Jorissen, M. W. P., A. Pariselle, T. Huyse, E. J. Vreven, J. Snoeks, F. A. M. Volckaert, A. Chocha Manda, G. Kapepula Kasembele, T. Artois, & M. P. M. Vanhove, 2018. Diversity and host specificity of monogenean gill parasites (Platyhelminthes) of cichlid fishes in the Bangweulu-Mweru ecoregion. Journal of Helminthology 92: 417–437.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, & L. S. Jermiin, 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature methods 14: 587–589.

Katoh, K., & D. M. Standley, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30: 772–780.

Keating, J. A., D. Bhattacharya, S. S. C. Rund, S. Hoover, R. Dasgupta, S. J. Lee, G. E. Duffield, & R. Striker, 2013. Mosquito protein kinase G phosphorylates flavivirus NS5 and alters flight behavior in *Aedes aegypti* and *Anopheles gambiae*. Vector-Borne and Zoonotic Diseases 13: 590–600.

Kmentová, N., A. J. Cruz-Laufer, A. Pariselle, K. Smeets, T. Artois, & M. P. M. Vanhove, 2022. Dactylogyridae 2022: a meta-analysis of phylogenetic studies and generic diagnoses of parasitic flatworms using published genetic and morphological data. International Journal for Parasitology 52: 427–457.

Kmentová, N., M. Gelnar, M. Mendlová, M. Van Steenberge, S. Koblmüller, & M. P. M. Vanhove, 2016. Reduced host-specificity in a parasite infecting non-littoral Lake Tanganyika cichlids evidenced by intraspecific morphological and genetic diversity. Scientific Reports 6: 39605.

Kmentová, N., C. Hahn, S. Koblmüller, H. Zimmermann, J. Vorel, T. Artois, M. Gelnar, & M. P. M. M. Vanhove, 2021. Contrasting host-parasite population structure: morphology and mitogenomics of a parasitic flatworm on pelagic deepwater cichlid fishes from Lake Tanganyika. Biology 10: 797.

Kmentová, N., S. Koblmüller, M. Van Steenberge, J. A. M. Raeymaekers, T. Artois, E. L. R. De Keyzer, L. Milec, F. Muterezi Bukinga, T. Mulimbwa N'sibula, P. Masilya Mulungula, G. Ntakimazi, F. A. M. Volckaert, M. Gelnar, & M. P. M. Vanhove, 2020. Weak population structure and recent demographic expansion of the monogenean parasite *Kapentagyrus* spp. infecting clupeid fishes of Lake Tanganyika, East Africa. International Journal for Parasitology 50: 471–486.

Kmentová, N., K. Thys, C. Hahn, J. Vorel, S. Koblmüller, M. Van Steenberge, A. Chocha Manda, L. Makasa, T. Mulimbwa N'sibula, P. Masilya Mulungula, & M. P. M. Vanhove, 2023. Comparative population mitogenomics of fish parasites reveals contrasting geographic pattern in the pelagic zone of Lake Tanganyika [PREPRINT]. Authorea DOI: 10.22541/au.167274604.41657501/v1.

Kofler, R., R. V. Pandey, & C. Schlötterer, 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27: 3435–3436.

Konczal, M., K. J. Przesmycka, R. S. Mohammed, K. P. Phillips, F. Camara, S. Chmielewski, C. Hahn, R. Guigo, J. Cable, & J. Radwan, 2020. Gene duplications, divergence and recombination shape adaptive evolution of the fish ectoparasite *Gyrodactylus bullatarudis*. Molecular Ecology 29: 1494–1507.

Kroetz, S. M., J. Srinivasan, J. Yaghoobian, P. W. Sternberg, & R. L. Hong, 2012. The cGMP signaling pathway affects feeding behavior in the necromenic nematode *Pristionchus pacificus*. PLoS ONE 7: e34464.

Layton, K. K. S., J. I. Carvajal, & N. G. Wilson, 2020. Mimicry and mitonuclear discordance in nudibranchs: new insights from exon capture phylogenomics. Ecology and Evolution 10: 11966–11982.

Lin, R., M. Yang, & B. Yao, 2022. The phylogenetic and evolutionary analyses of detoxification gene families in *Aphidinae*[sic] species. PLoS ONE 17: e0263462.

Liu, C., Y. Zhang, Y. Ren, H. Wang, S. Li, F. Jiang, L. Yin, X. Qiao, G. Zhang, W. Qian, B. Liu, & W. Fan, 2018. The genome of the golden apple snail *Pomacea canaliculata* provides insight into stress tolerance and invasive adaptation. GigaScience 7: giy101.

Liu, D., X. Dang, W. Song, L. Xi, Q. Wang, S. Zhang, Y. Miao, G. Li, & J. Jiang, 2019. Molecular characterization and expression patterns of *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae) heat shock protein genes and their response to host stress. Archives of Insect Biochemistry and Physiology 100: e21536.

Lucas, C., R. Kornfein, M. Chakaborty-Chatterjee, J. Schonfeld, N. Geva, M. B. Sokolowski, & A. Ayali, 2010. The locust *foraging* gene. Archives of Insect Biochemistry and Physiology 74: 52–66.

Łukasik, I., S. Goławska, & H. Sytykiewicz, 2022. Differences in oxidative stress markers and antioxidant enzyme activities in black bean aphid morphs (*Aphis fabae* Scop.) fed on the primary host *Viburnum opulus* L. Antioxidants 11: 2476.

Łukasik, I., S. Goławska, A. Wójcicka, & A. Goławski, 2011. Effect of host plants on antioxidant system of pea aphid *Acyrthosiphon pisum*. Bulletin of Insectology 64: 153–158.

Lu, Y., J. H. Ran, D. M. Guo, Z. Y. Yang, & X. Q. Wang, 2014. Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. PLoS ONE 9: e107679.

Makino, T., & M. Kawata, 2019. Invasive invertebrates associated with highly duplicated gene content. Molecular Ecology 28: 1652–1663.

Marcogliese, D. J., 2008. The impact of climate change on the parasites and aquatic diseases of animals. Revue Scientifique et Technique (International Office of Epizootics) 27: 467–484.

Markov, G. V., P. Baskaran, & R. J. Sommer, 2015. The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. Journal of Molecular Evolution 80: 18–36.

Marliére, N. P., M. G. Lorenzo, L. E. Martínez Villegas, & A. A. Guarneri, 2020. Co-existing locomotory activity and gene expression profiles in a kissing-bug vector of Chagas disease. Journal of Insect Physiology 122: 104021.

Martínez-González, J. de J., A. Guevara-Flores, & I. P. Del Arenal Mena, 2022. Evolutionary adaptations of parasitic flatworms to different oxygen tensions. Antioxidants 11: 1102.

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

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, R. Lanfear, & E. Teeling, 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37: 1530–1534.

Mittapalli, O., J. J. Neal, & R. H. Shukle, 2007. Antioxidant defense response in a galling insect. Proceedings of the National Academy of Sciences 104: 1889–1894.

Moons, T., N. Kmentová, A. Pariselle, T. Artois, W. Bert, M. P. M. Vanhove, & A. J. Cruz-Laufer, 2023. 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 30: 25.

Morley, N. J., & J. W. Lewis, 2014. Temperature stress and parasitism of endothermic hosts under climate change. Trends in Parasitology 30: 221–227.

Nelson, D. R., 2018. Cytochrome P450 diversity in the tree of life. Biochimica et Biophysica Acta - Proteins and Proteomics 1866: 141–154.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, & B. Q. Minh, 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32: 268–274.

Ogbeide, S., F. Giannese, L. Mincarelli, & I. C. Macaulay, 2022. Into the multiverse: advances in single-cell multiomic profiling. Trends in Genetics 38: 831–843.

Pakharukova, M. Y., V. A. Vavilin, B. Sripa, T. Laha, P. J. Brindley, & V. A. Mordvinov, 2015. Functional analysis of the unique cytochrome P450 of the liver fluke *Opisthorchis felineus*. PLoS Neglected Tropical Diseases 9: e0004258.

Pan, L. N., F. Z. Wang, X. Y. Zhang, Y. N. Zhao, G. P. Zhu, & M. Li, 2018. Identification and characterization of heat shock proteins in a parasitic wasp *Chouioia cuneae* (Hymenoptera: Eulophidae). Entomological Research 48: 145–155.

Paradis, E., & K. Schliep, 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526–528.

Pariselle, A., S. Morand, M. R. Deveney, & L. Pouyaud, 2003. Parasite species richness of closely related hosts: historical scenario and "genetic" hypothesis In Combes, C., J. Jourdane, A. Ducreux-Modat, & J.-R. Pages (eds), Taxonomie, écologie et évolution des métazoaires parasites: livre-hommage à Louis Euzet. Presses Universitaires de Perpignan, Perpignan, France: 147–166.

Pazdzior, R., Z. Yang, M. S. Mesbahuddin, J. Yee, A. van der Est, & S. Rafferty, 2015. Low reduction potential cytochrome b_5 isotypes of *Giardia intestinalis*. Experimental Parasitology 157: 197–201.

Pease, J. B., D. C. Haak, M. W. Hahn, & L. C. Moyle, 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biology 14: e1002379.

Pereira, M. H., R. M. M. Paim, C. Lahondère, & C. R. Lazzari, 2017. Heat shock proteins and blood-feeding in arthropods In Asea, A. A. A., & Kaur P. (eds), Heat shock proteins in veterinary medicine and sciences. Springer, Cham, Switzerland: 349–359.

Philippe, H., N. Lartillot, & H. Brinkmann, 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Molecular Biology and Evolution 22: 1246–1253.

Pirozhkova, D., & A. Katokhin, 2020. *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. Parasitology Research 119: 2695–2702.

Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus, & A. Korobeynikov, 2020. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70: e102.

Rahmouni, C., M. P. M. Vanhove, S. Koblmüller, & A. Šimková, 2022. Molecular phylogeny and speciation patterns in host-specific monogeneans (*Cichlidogyrus*,

Dactylogyridae) parasitizing cichlid fishes (Cichliformes, Cichlidae) in Lake Tanganyika. International Journal for Parasitology 52: 359–375.

Raizen, D. M., J. E. Zimmerman, M. H. Maycock, U. D. Ta, Y. J. You, M. V. Sundaram, & A. I. Pack, 2008. Lethargus is a *Caenorhabditis elegans* sleep-like state. Nature 451: 569–572.

R Core Team, 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing, Vienna, Austria, Accessed January 8, 2022. https://www.r-project.org/

Rosa, B. A., Y. J. Choi, S. N. McNulty, H. Jung, J. Martin, T. Agatsuma, H. Sugiyama, T. H. Le, P. N. Doanh, W. Maleewong, D. Blair, P. J. Brindley, P. U. Fischer, & M. Mitreva, 2020. Comparative genomics and transcriptomics of 4 *Paragonimus* species provide insights into lung fluke parasitism and pathogenesis. GigaScience 9: gia0073.

Roy, S., A. K. Prasad, & S. Roy, 2018. First report on heat shock protein expression in *Helopeltis theivora* Waterhouse (Hemiptera: Miridae) in response to different hosts. Journal of the Kansas Entomological Society 91: 238–241.

Sabadin, G. A., M. A. Xavier, & I. Da Silva Vaz, 2019. Control of redox homeostasis in tick blood feeding. Acta Scientiae Veterinariae 47: 1678.

Santi, A. M. M., J. M. Ribeiro, J. L. Reis-Cunha, G. de A. Burle-Caldas, I. F. M. Santos, P. A. Silva, D. de M. Resende, D. C. Bartholomeu, S. M. R. Teixeira, & S. M. F. Murta, 2022. Disruption of multiple copies of the Prostaglandin F2alpha synthase gene affects oxidative stress response and infectivity in *Trypanosoma cruzi*. PLoS Neglected Tropical Diseases 16: e0010845.

Saveria, T., A. Halbach, R. Erdmann, R. Volkmer-Engert, C. Landgraf, H. Rottensteiner, & M. Parsons, 2007. Conservation of PEX19-binding motifs required for protein targeting to mammalian peroxisomal and trypanosome glycosomal membranes. Eukaryotic Cell 6: 1439–1449.

Shaheen, S., K. F. Barrett, S. Subramanian, S. L. M. Arnold, J. A. Laureanti, P. J. Myler, W. C. Van Voorhis, & G. W. Buchko, 2020. Solution structure for an *Encephalitozoon cuniculi* adrenodoxin-like protein in the oxidized state. Protein Science 29: 809–817.

Sies, H., C. Berndt, & D. P. Jones, 2017. Oxidative stress. Annual Review of Biochemistry 86: 25.1-25.34.

Šimková, A., E. Řehulková, J. R. Rasoloariniaina, M. W. P. Jorissen, T. Scholz, A. Faltýnková, Š. Mašová, & M. P. M. Vanhove, 2019. Transmission of parasites from introduced tilapias: a new threat to endemic Malagasy ichthyofauna. Biological Invasions 21: 803–819.

Skern-Mauritzen, R., K. Malde, C. Eichner, M. Dondrup, T. Furmanek, F. Besnier, A. Z. Komisarczuk, M. Nuhn, S. Dalvin, R. B. Edvardsen, S. Klages, B. Huettel, K. Stueber, S. Grotmol, E. Karlsbakk, P. Kersey, J. S. Leong, K. A. Glover, R. Reinhardt, S. Lien,

I. Jonassen, B. F. Koop, & F. Nilsen, 2021. The salmon louse genome: copepod features and parasitic adaptations. Genomics 113: 3666–3680.

Slater, G. S. C., & E. Birney, 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31.

Sokolowski, M. B., 1980. Foraging strategies of *Drosophila melanogaster*: a chromosomal analysis. Behavior Genetics 10: 291–302.

Stanley, T. R., K. S. Kim Guisbert, S. M. Perez, M. Oneka, I. Kernin, N. R. Higgins, A. Lobo, M. M. Subasi, D. J. Carroll, R. G. Turingan, & E. Guisbert, 2022. Stress response gene family expansions correlate with invasive potential in teleost fish. Journal of Experimental Biology 225: jeb243263.

Struk, A. A., J. Mugon, A. Huston, A. A. Scholer, G. Stadler, E. Tory Higgins, M. B. Sokolowski, & J. Danckert, 2019. Self-regulation and the foraging gene (PRKG1) in humans. Proceedings of the National Academy of Sciences of the United States of America 116: 4434–4439.

Swenson, K. M., & N. El-Mabrouk, 2012. Gene trees and species trees: irreconcilable differences. BMC Bioinformatics 13(Suppl 19): S15.

Talavera, G., & J. Castresana, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56: 564–577.

Tapia, D. H., A. X. Silva, G. I. Ballesteros, C. C. Figueroa, H. M. Niemeyer, & C. C. Ramírez, 2015. Differences in learning and memory of host plant features between specialist and generalist phytophagous insects. Animal Behaviour 106: 1–10.

Tian, R., Y. Geng, Y. Yang, I. Seim, & G. Yang, 2021. Oxidative stress drives divergent evolution of the glutathione peroxidase (GPX) gene family in mammals. Integrative Zoology 16: 696–711.

Törönen, P., & L. Holm, 2022. PANNZER—A practical tool for protein function prediction. Protein Science 31: 118–128.

Tsai, I. J., M. Zarowiecki, N. Holroyd, A. Garciarrubio, A. Sanchez-Flores, K. L. Brooks, A. Tracey, R. J. Bobes, G. Fragoso, E. Sciutto, M. Aslett, H. Beasley, H. M. Bennett, J. Cai, F. Camicia, R. Clark, M. Cucher, N. De Silva, T. A. Day, P. Deplazes, K. Estrada, C. Fernández, P. W. H. Holland, J. Hou, S. Hu, T. Huckvale, S. S. Hung, L. Kamenetzky, J. A. Keane, F. Kiss, U. Koziol, O. Lambert, K. Liu, X. Luo, Y. Luo, N. MacChiaroli, S. Nichol, J. Paps, J. Parkinson, N. Pouchkina-Stantcheva, N. Riddiford, M. Rosenzvit, G. Salinas, J. D. Wasmuth, M. Zamanian, Y. Zheng, X. Cai, X. Soberon, P. D. Olson, J. P. Laclette, K. Brehm, M. Berriman, E. Morett, T. Portillo, M. V. Jose, J. C. Carrero, C. Larralde, J. Morales-Montor, J. Limon-Lason, M. A. Cevallos, V. Gonzalez, A. Ochoa-Leyva, A. Landa, L. Jimenez, & V. Valdes, 2013. The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496: 57–63.

Vanhove, M. P. M., A. G. Briscoe, M. W. P. Jorissen, D. T. J. Littlewood, & T. Huyse, 2018. The first next-generation sequencing approach to the mitochondrial phylogeny of African monogenean parasites (Platyhelminthes: Gyrodactylidae and Dactylogyridae). BMC Genomics 19: 520.

Vanhove, M. P. M., P. I. Hablützel, A. Pariselle, A. Šimková, T. Huyse, & J. A. M. Raeymaekers, 2016. Cichlids: a host of opportunities for evolutionary parasitology. Trends in Parasitology 32: 820–832.

Vanhove, M. P. M., R. Hermans, T. Artois, & N. Kmentová, 2021. 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 11: 3578.

Vanhove, M. P. M., A. Pariselle, M. Van Steenberge, J. A. M. Raeymaekers, P. I. Hablützel, C. Gillardin, B. Hellemans, F. C. Breman, S. Koblmüller, C. Sturmbauer, J. Snoeks, F. A. M. Volckaert, & T. Huyse, 2015. Hidden biodiversity in an ancient lake: phylogenetic congruence between Lake Tanganyika tropheine cichlids and their monogenean flatworm parasites. Scientific Reports 5: 13669.

Vanhove, M. P. M., B. Tessens, C. Schoelinck, U. Jondelius, D. T. J. Littlewood, T. Artois, & T. Huyse, 2013. Problematic barcoding in flatworms: a case-study on monogeneans and rhabdocoels (Platyhelminthes). ZooKeys 365: 355–379.

Vorel, J., N. Kmentová, C. Hahn, P. Bureš, & M. Kašný, 2023. An insight into the functional genomics and species classication of *Eudiplozoon nipponicum* (Monogenea, Diplozoidae), a haematophagous parasite of the common carp *Cyprinus carpio*. BMC Genomics 24: 363.

Wagner, C. E., 2021. Ecological opportunity, genetic variation, and the origins of African cichlid radiations In Abate, M. E., & D. L. G. Noakes (eds), The behavior, ecology and evolution of cichlid fishes. Springer Nature, Dordrecht, The Netherlands: 79–105.

Wang, B., X. Liang, M. L. Gleason, R. Zhang, & G. Sun, 2018. Comparative genomics of *Botryosphaeria dothidea* and *B. kuwatsukai*, causal agents of apple ring rot, reveals both species expansion of pathogenicityrelated genes and variations in virulence gene content during speciation. IMA Fungus 9: 243–257.

Waterhouse, R. M., M. Seppey, F. A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, & E. M. Zdobnov, 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution 35: 543–548.

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis. Springer, New York, USA.

Wilson, A. B., G. G. Teugels, & A. Meyer, 2008. Marine incursion: the freshwater herring of Lake Tanganyika are the product of a marine invasion into West Africa. PLoS ONE 3: e1979.

Wisedpanichkij, R., R. Grams, W. Chaijaroenkul, & K. Na-Bangchang, 2011. Confutation of the existence of sequence-conserved cytochrome P450 enzymes in *Plasmodium falciparum*. Acta Tropica 119: 19–22.

Wu, M., J. L. Kostyun, M. W. Hahn, & L. C. Moyle, 2018. Dissecting the basis of novel trait evolution in a radiation with widespread phylogenetic discordance. Molecular Ecology 27: 3301–3316.

Xu, L., J. Yang, M. Xu, D. Shan, Z. Wu, & D. Yuan, 2020a. Speciation and adaptive evolution reshape antioxidant enzymatic system diversity across the phylum Nematoda. BMC Biology 18: 181.

Xu, Z., M. Ji, C. Li, X. Du, W. Hu, D. P. McManus, & H. You, 2020b. A biological and immunological characterization of *Schistosoma japonicum* heat shock proteins 40 and 90α. International Journal of Molecular Sciences 21: 4034.

Yoo, W. G., D. W. Kim, J. W. Ju, P. Y. Cho, T. I. Kim, S. H. Cho, S. H. Choi, H. S. Park, T. S. Kim, & S. J. Hong, 2011. Developmental transcriptomic features of the carcinogenic liver fluke, *Clonorchis sinensis*. PLoS Neglected Tropical Diseases 5: e1208.

Yuan, H., C. Atta, L. Tornabene, & C. Li, 2019. Assexon: Assembling Exon using gene capture data. Evolutionary Bioinformatics 15: 1–13.

Yuan, X., W.-W. Zhou, Y. Zhou, S. Liu, F. Lu, M.-F. Yang, J. Cheng, G. M. Gurr, & Z.-R. Zhu, 2014. Composition and expression of heat shock proteins in an invasive pest, the rice water weevil (Coleoptera: Curculionidae). Florida Entomologist 97: 611–619.

Yu, G., T. T.-Y. Lam, H. Zhu, & Y. Guan, 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. Molecular Biology and Evolution 35: 3041–3043.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, & T. T.-Y. Lam, 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8: 28–36.

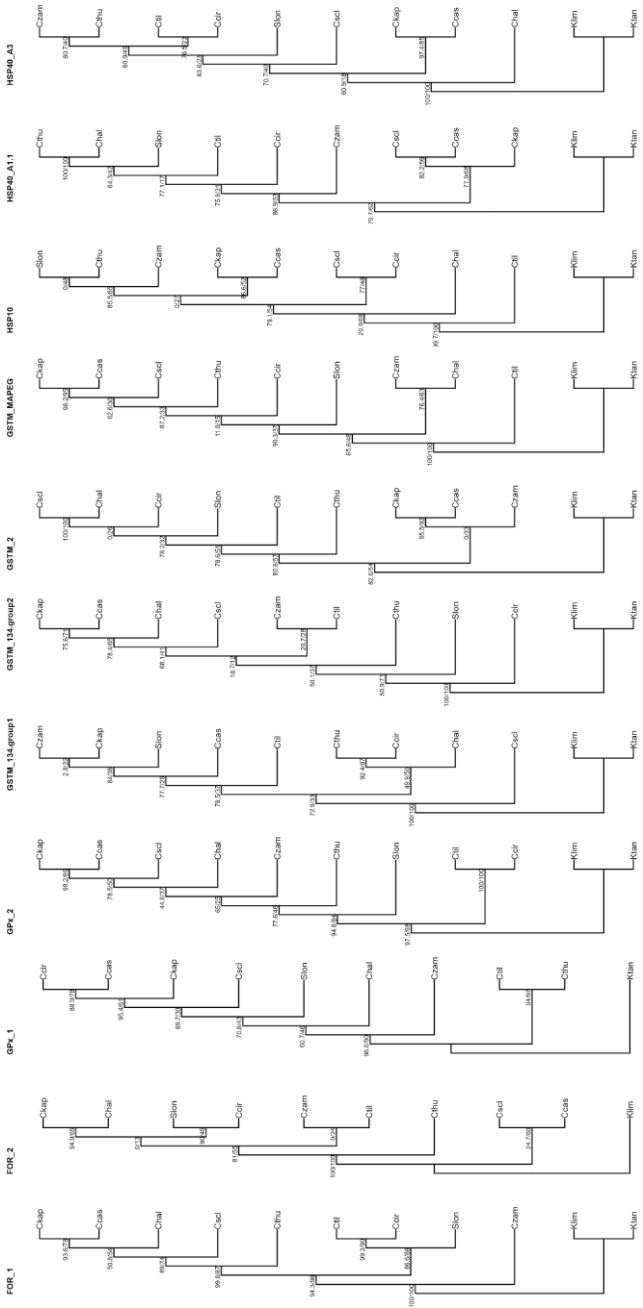
Žárský, V., & J. Tachezy, 2015. Evolutionary loss of peroxisomes - not limited to parasites. Biology Direct 10: 74.

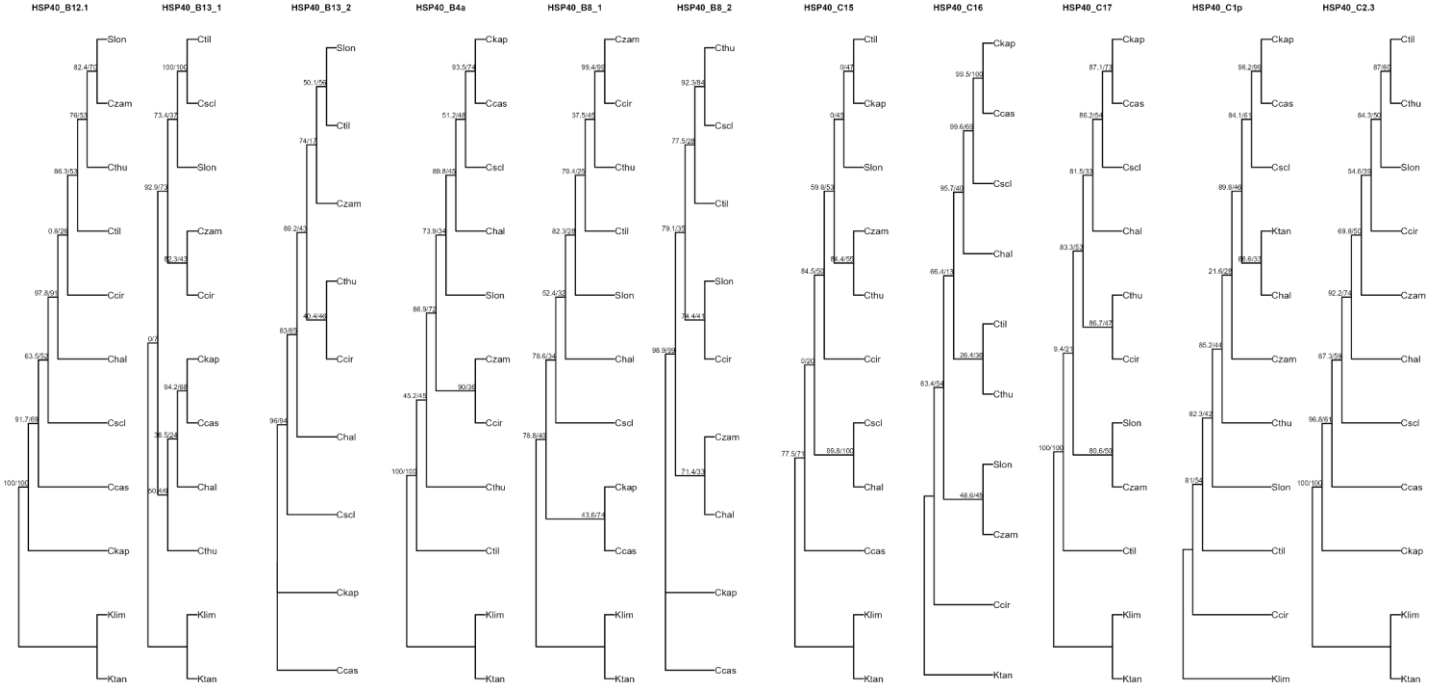
Zhang, W., H. Yu, Y. Lv, K. E. Bushley, J. D. Wickham, S. Gao, S. Hu, L. Zhao, & J. Sun, 2020. Gene family expansion of pinewood nematode to detoxify its host defence chemicals. Molecular Ecology 29: 940–955.

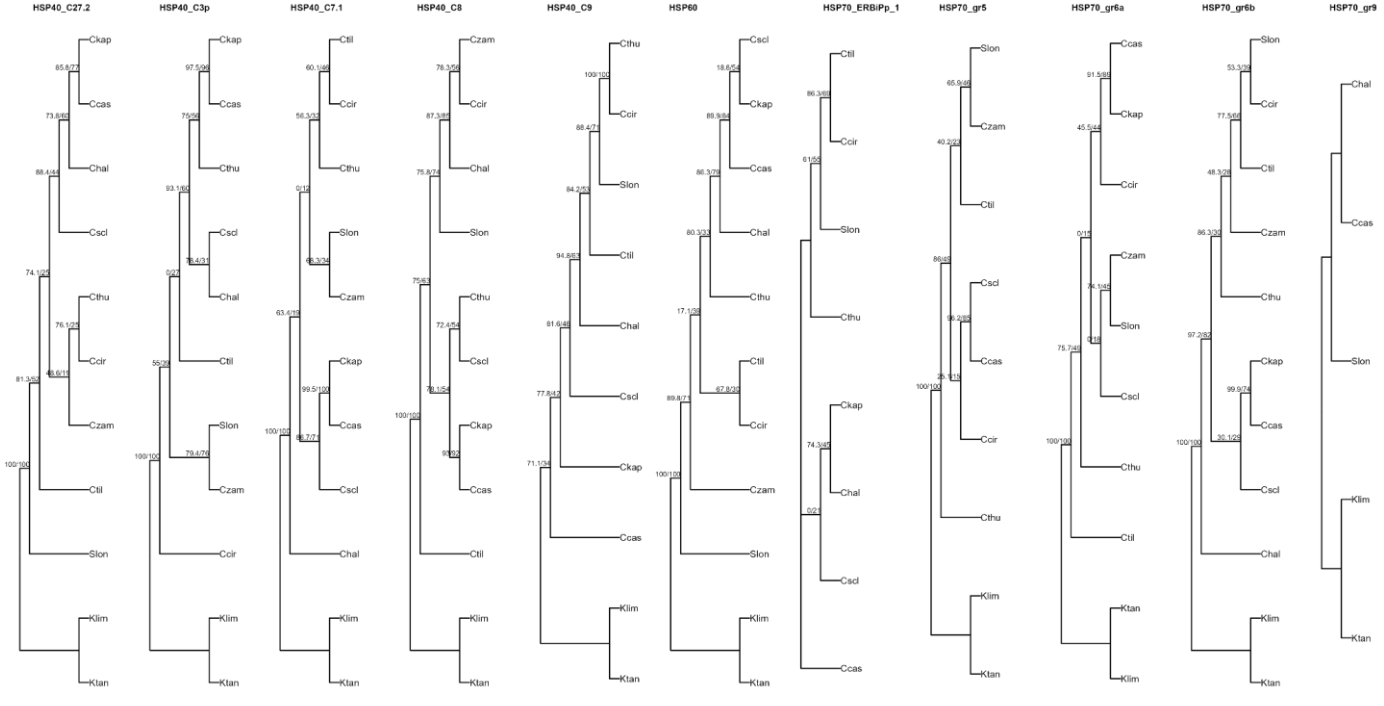
Ziniel, P. D., B. Karumudi, A. H. Barnard, E. M. S. Fisher, G. R. J. Thatcher, L. M. Podust, & D. L. Williams, 2015. The *Schistosoma mansoni* cytochrome P450 (CYP3050A1) is essential for worm survival and egg development. PLoS Neglected Tropical Diseases 9: e0004279.

Appendices

Appendix S7.1. Maximum likelihood cladogram of putative protein sequences of stress genes of species of *Cichlidogyrus*. For species names, see Fig. 6.1. For abbreviation of protein names, see Table 6.2.




HSP40_B12.1
Slon
Czam
Cthu
Ctil
Ccir
Chal
Cscl
Ccas
Ckap
Klim
Ktan
HSP40_B13_1
Ctil
Cscl
Slon
Czam
Ccir
Ckap
Ccas
Chal
Cthu
Klim
Ktan
HSP40_B13_2
Slon
Ctil
Czam
Cthu
Ccir
Chal
Cscl
Ckap
Ccas
HSP40_B4a
Ckap
Ccas
Cscl
Chal
Slon
Czam
Ccir
Cthu
Ctil
Klim
Ktan
HSP40_B8_1
Czam
Ccir
Cthu
Ctil
Slon
Chal
Cscl
Ckap
Ccas
Klim
Ktan
HSP40_B8_2
Cthu
Cscl
Ctil
Slon
Ccir
Czam
Chal
Ckap
Ccas
HSP40_C15
Ctil
Ckap
Slon
Czam
Cthu
Ccir
Cscl
Chal
Ccas
Klim
Ktan
HSP40_C16
Ckap
Ccas
Cscl
Chal
Ctil
Cthu
Slon
Czam
Ccir
Ktan
HSP40_C17
Ckap
Ccas
Cscl
Chal
Cthu
Ccir
Slon
Czam
Ctil
Klim
Ktan
HSP40_C1p
Ckap
Ccas
Cscl
Ktan
Chal
Czam
Cthu
Slon
Ctil
Ccir
Klim
HSP40_C2.3
Ctil
Cthu
Slon
Ccir
Czam
Chal
Cscl
Ccas
Ckap
Klim
Ktan


HSP40_C27.2
Ckap
Ccas
Chal
Cscl
Cthu
Ccir
Czam
Ctil
Slon
Klim
Ktan
HSP40_C3p
Ckap
Ccas
Cthu
Cscl
Chal
Ctil
Slon
Czam
Ccir
Klim
Ktan
HSP40_C7.1
Ctil
Ccir
Cthu
Slon
Czam
Ckap
Ccas
Cscl
Chal
Klim
Ktan
HSP40_C8
Czam
Ccir
Chal
Slon
Cthu
Cscl
Ckap
Ccas
Ctil
Klim
Ktan
HSP40_C9
Cthu
Ccir
Slon
Ctil
Chal
Cscl
Ckap
Ccas
Klim
Ktan
HSP60
Cscl
Ckap
Ccas
Chal
Cthu
Ctil
Ccir
Czam
Slon
Klim
Ktan
HSP70_ERBiPp_1
Ctil
Ccir
Slon
Cthu
Ckap
Chal
Cscl
Ccas
HSP70_gr5
Slon
Czam
Ctil
Cscl
Ccas
Ccir
Cthu
Klim
Ktan
HSP70_gr6a
Ccas
Ckap
Ccir
Czam
Slon
Cscl
Cthu
Ctil
Ktan
Klim
HSP70_gr6b
Slon
Ccir
Ctil
Czam
Cthu
Ckap
Ccas
Cscl
Chal
Klim
Ktan
HSP70_gr9
Chal
Ccas
Slon
Klim
Ktan