Somewhere I Belong: Phylogenetic Comparative Methods and Machine Learning to Investigate the Evolution of a Species-Rich Lineage of Parasites.

Armando J. Cruz-Laufer[1], Antoine Pariselle[2,3], Michiel W. P. Jorissen[1,4], Fidel Muterezi Bukinga[5], Anwar Al Assadi[6], Maarten Van Steenberge[7,8], Stephan Koblmüller[9], Christian Sturmbauer[9], Karen Smeets[1], Tine Huyse[4,7], Tom Artois[1], Maarten P. M. Vanhove[1,7,10]

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

2 ISEM, Université de Montpellier, CNRS, IRD, Montpellier, France.

3 Faculty of Sciences, Laboratory "Biodiversity, Ecology and Genome", Research Centre "Plant and Microbial Biotechnology, Biodiversity and Environment", Mohammed V University, Rabat, Morocco.

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

5 Section de Parasitologie, Département de Biologie, Centre de Recherche en Hydrobiologie, Uvira, Democratic Republic of the Congo.

6 Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Nobelstraße 12, 70569 Stuttgart, Germany.

7 Laboratory of Biodiversity and Evolutionary Genomics, KU Leuven, Charles Deberiotstraat 32, B-3000, Leuven, Belgium.

*8 Operational Directorate Taxonomy and Phylogeny, Royal Belgian Institute of Natural Sciences, Vautierstraat 29, B-1000 Brussels, Belgium. *

*9 Institute of Biology, University of Graz, Universitätsplatz 2, 8010, Graz, Austria. *

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

10 Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, CZ-611 37, Brno, Czech Republic.

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

Abstract

Metazoan parasites encompass a significant portion of the global biodiversity. Their relevance for environmental and human health calls for a better understanding as parasite macroevolution remains mostly understudied. Yet limited molecular, phenotypic, and ecological data have so far discouraged complex analyses of evolutionary mechanisms and encouraged the use of data discretisation and body-size correction. In this case study, we aim to highlight the limitations of these methods and propose new methods optimised for small datasets. We apply multivariate phylogenetic comparative methods (PCMs) and statistical classification using support vector machines (SVMs) to a data-deficient host-parasite system. We use continuous morphometric and host range data currently widely inferred from a species-rich lineage of parasites (*Cichlidogyrus* incl. *Scutogyrus* - Platyhelminthes: Monogenea, Dactylogyridae) infecting cichlid fishes. For PCMs, we modelled the attachment organ and host range evolution using the data of 135 species and an updated multi-marker (28S and 18S rDNA, ITS1, COI mtDNA) phylogenetic reconstruction of 58/137 described species. Through a cluster analysis, SVM-based classification, and taxonomic literature survey, we infered the systematic informativeness of discretised and continuous characters. We demonstrate that an update to character coding and size-correction techniques is required as some techniques mask phylogenetic signals but remain useful for characterising species

groups of *Cichlidogyrus*. Regarding the attachment organ evolution, PCMs suggest a pattern associated with genetic drift. Yet host and environmental parameters might put this structure under stabilising selection as indicated by a limited morphological variation. This contradiction, the absence of a phylogenetic signal and multicollinearity in most measurements, a moderate 73% accordance rate of taxonomic approach and SVMs, and a low phylogenetic informativeness of reproductive organ data suggest an overall limited systematic value of the measurements included in most species characterisations. We conclude that PCMs and SVM-based approaches are suitable tools to investigate the character evolution of data-deficient taxa.

KEY WORDS

Cichlidogyrus; co-divergence; Dactylogyridae; host switching; Monogenea; phylogenetic signal, *Scutogyrus*; support vector machines.

ACKNOWLEDGEMENTS

We thank J.-F. Agnèse (Institut de Recherche pour le Développement & Université de Montpellier) for taking responsibility for part of the molecular work, and to F. A. M. Volckaert (KU Leuven) and J. Snoeks (Royal Museum for Central Africa & KU Leuven) for their guidance since the early stages of this research. J. Bamps, A. F. Grégoir, L. Makasa, J. K. Zimba (Department of Fisheries), C. Katongo (University of Zambia), T. Veall, O. R. Mangwangwa (Rift Valley Tropicals), V. Nshombo Muderhwa, T. Mulimbwa N'sibula, D. Muzumani Risasi, J. Mbirize Ndalozibwa, V. Lumami Kapepula (Centre de Recherche en Hydrobiologie-Uvira), the Schreyen-Brichard family (Fishes of Burundi), F. Willems (Kasanka Trust), S. Dessein (Botanic Garden Meise), A. Chocha Manda, G. Kapepula Kasembele, E. Abwe, B. K. Manda, C. Mukweze Mulelenu, M. Kasongo Ilunga Kayaba and C. Kalombo Kabalika (Université de Lubumbashi), M. Collet and P. N'Lemvo (Institut

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM Congolais pour la Conservation de la Nature), D. Kufulu-ne-Kongo (École Muilu Kiawanga), L. Matondo Mbela (Université Kongo), S. Wamuini Lunkayilakio, P. Nguizani Bimbundi, B. Boki Fukiakanda, P. Ntiama Nsiku (Institut Supérieur Pédagogique de Mbanza-Ngungu), P Nzialu Mahinga (Institut National pour l'Etude et la Recherche Agronomiques—Mvuazi/Institut Supérieur d'études agronomiques—Mvuazi) and M Katumbi Chapwe are thanked for administrative, field and lab support, making this study possible. Furthermore, we thank A. Avenant-Oldewage, M. Geraerts, P. C. Igeh, N. Kmentová, M. Mendlová, and C. Rahmouni for providing raw species measurements used for their respective publications. Field work was supported by travel grants V.4.096.10.N.01, K.2.032.08.N.01 and K220314N from the Research Foundation—Flanders (FWO-Vlaanderen) (to MPMV), two travel grants from the King Leopold III Fund for Nature Conservation and Exploration (to MPMV and MVS), FWO-Vlaanderen Research Programme G.0553.10, the University Development Cooperation of the Flemish Interuniversity Council (VLIR-UOS) South Initiative ZRDC2014MP084, the OCA type II project S1_RDC_TILAPIA and the Mbisa Congo project (2013–2018), the latter two being framework agreement projects of the RMCA with the Belgian Development Cooperation. When data collection for this study commenced, MVS and MPMV were PhD fellows, and TH a post-doctoral fellow, of FWO-Vlaanderen. Part of the research leading to 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. MWPJ was supported by the Belgian Federal Science Policy Office (BRAIN-be Pioneer Project (BR/132/PI/TILAPIA), and a BOF Reserve Fellowship from Hasselt University. AJCL is funded by Hasselt University (BOF19OWB02), and MPMV receives support from the Special Research Fund of Hasselt University (BOF20TT06).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The morphological data underlying this article are available in MorphoBank at www.morphobank.org, at https://dx.doi.org/XXXXXXXX. The DNA sequence data are available in the GenBank Nucleotide Database at https://www.ncbi.nlm.nih.gov/genbank, and can be accessed with the accession numbers XXXXXX–XXXXXX. Phylogenetic trees and data matrices are available in TreeBase at https://treebase.org, and can be accessed with accession number XXXXXX.

Parasite Speciation: Together Yet Apart

Parasitic organisms encompass a substantial part of metazoan biodiversity (Poulin 2014) and are known to shape species communities (Gómez and Nichols 2013) ranging from the tropics (Dunne et al. 2013) to the Arctic (Amundsen et al. 2009). Metazoan parasites are also known to impact human health and food security (Thompson 2013; Webster et al. 2016). Consequently, macroevolutionary and speciation processes of these organisms have received increasing attention in recent years, especially in the context of human-wildlife spillover events (Thompson 2013; Webster et al. 2016). Yet concepts for speciation processes first characterised in free-living organisms, are not easily transferable to parasitic organisms. For instance, sympatric speciation is defined as species formation within the same geographical area and allopatric speciation as species formation through geographic isolation. Yet parasites whose hosts share a habitat do not necessarily live in sympatry. Differences in host ecology and behaviour can create physical (and temporal) barriers for the parasites that might mimic spatial isolation. These barriers could result in allopatric species formation even if the infected host species live in sympatry (McCoy 2003). Classic examples for allopatric speciation events in parasites include co-divergence of host and parasite lineages as well as isolation of peripheral populations through host switching (Huyse et al. 2005). Recent studies (McCoy 2004; Huyse et al. 2005) suggest that for parasite lineages with sympatric hosts, species formation might occur as a result of both sympatric and allopatric speciation processes. In many cases, allopatric speciation processes might play a larger role than previously thought (McCoy 2004).

Rich in Species, Poor in Data: A Case Study for a Host-Parasite System

Species networks of host and parasites have far-reaching potential as eco-evolutionary model systems in speciation research (see Toju et al. 2017). Host-parasite interactions are

usually well- preserved in natural history collections (Harmon et al. 2019; Jorissen et al. 2020) because often parasites remain attached to the body or inside the body of the collected specimens. One potential model system could be the parasite interactions of cichlid fishes (Cichlidiformes: Cichlidae). African cichlids are possibly one of the best known study systems for adaptive radiation research (Salzburger 2018) with approximately 2000 (described and undescribed) species reported from Eastern Africa alone (Turner et al. 2001; Salzburger et al. 2014). Notably, a group of gill parasites (Fig. 1a) infecting these fishes, the monogenean flatworms belonging to *Cichlidogyrus* sensu Paperna, 1960 (incl. *Scutogyrus* Pariselle & Euzet, 1995, together termed *Cichlidogyrus* in the following) (Fig. 1), rivals (Cruz-Laufer et al. 2020) and likely exceeds (see Poulin 2014) its hosts in terms of species numbers. Currently, 137 species have been described on 126 fish species with a majority of the species likely remaining undiscovered (Cruz-Laufer et al. 2020). Because of this diversity, the cichlid-*Cichlidogyrus* species network has been proposed as model system for macroevolutionary research in host-parasite systems (Pariselle et al. 2003; Vanhove et al. 2016). Recent advances in immunological (Zhi et al. 2018), pathological (Igeh and Avenant-Oldewage 2020), genomic (Vanhove et al. 2018; Caña-Bozada et al. 2021), and microscopy research (Fannes et al. 2015) have brought the vision of a cichlid-*Cichlidogyrus* model system closer to reality (see Cruz-Laufer et al. (2020) for a detailed review of model system qualities). These advances are not least rooted in a long and productive tradition of international cichlid research (see Van Steenberge et al. 2011; Ronco et al. 2020).

Previous evolutionary studies have attributed the species-richness of *Cichlidogyrus* to co-divergence (Pariselle et al. 2003; Vanhove et al. 2015), host switching (Pariselle et al. 2003; Messu Mandeng et al. 2015), and within host speciation (Mendlová et al. 2012; Vanhove et al. 2015) similar to most parasites (Huyse et al. 2005). Morphological characters of the hard parts have formed an essential part of species descriptions of monogenean

flatworms. Therefore, evolutionary studies have paid considerable attention to the evolution of these structures (e.g. Šimková et al. 2002; Mendlová et al. 2012). For species of *Cichlidogyrus*, these characters are mostly derived from the sclerotised parts of the attachment and reproductive organs (Fig. 1b, c), the latter including the male copulatory organ (MCO) and the vagina. The attachment organ is the parasite's main physical connection to the host. Intuitively, the evolution of the sclerotised parts is, therefore, heavily influenced by host and environmental parameters (Kaltz and Shykoff 1998). However, based on the similar morphologies observed in related species, some studies have suggested that phylogenetic constraints, i.e. ancestry of the parasite species, are the sole determinants of the morphology of the attachment organ in *Cichlidogyrus* (Vignon et al. 2011). This claim has been questioned more recently. At least one species appears to have changed its attachment organ morphology compared to its close relatives as a response to a host switch towards non-cichlid fishes (Messu Mandeng et al. 2015). This change suggests that host parameters can play a role at least for recent speciation events. But despite this discovery, no study has investigated the macroevolution of the morphological characters of the attachment and reproductive organs for almost a decade (see Mendlová et al. 2012).

The absence of recent studies on the evolution of species of *Cichlidogyrus* might be rooted in the data deficiency of the cichlid-*Cichlidogyrus* system (Cruz-Laufer et al. 2020). In fact, out of the 137 species of *Cichlidogyrus* that are currently described, DNA sequences of only up to 18 species were included in the most recent phylogenetic study (Messu Mandeng et al. 2015). Even if regularly published data such as morphometric measurements are provided, many data beyond DNA sequences are rarely uploaded to centralised databases, which inhibits data accesibility (Cruz-Laufer et al. 2020). Raw data of measurements are often not provided with few exceptions e.g. Kmentová et al. (2016b). As a consequence of

this data scarcity, some studies have proposed discretising the attachment organ measurements or host range data to model the evolution of these organisms. Vignon et al. (2011) and Mendlová et al. (2012) grouped species by the the relative size of the anchors, bars, and hooks of the attachment organ (for an overview of the terminology of these structures, see Figure 1c). Similarly, Mendlová and Šimková (2014) proposed an index of specificity (IS) to group species as species, genus, and tribe-specific, or generalist. Yet discretisation has been known to cause information loss as the variability of the raw continuous data is largely ignored (Altman and Royston 2006; Goloboff et al. 2006; Parins-Fukuchi 2018). Other studies have used size correction to investigate the evolution of species of *Cichlidogyrus*. Size correction is widely employed as a method to separate shape from size information in continuous measurements (McCoy et al. 2006). As monogenean flatworms are variable in body size due to lack of a skeleton, Pariselle and Euzet (2009) proposed the length of the second hook pair in the attachment organ (U2). These structurues maintain their embryonal size in later stages of development (Llewellyn 1963). Yet little evidence supports a correlation of U2 and the body size of these organisms as studies that have investigated allometric effects in monogeneans have focused on the anchors of the attachment organ (Rodríguez-González et al. 2016, 2017). These obstacles emphasise the need for new approaches and more extensive and accessible datasets to gain insight into the evolution of *Cichlidogyrus* and the phylogenetic information of morphological and ecological characters.

Phylogenetic Comparative Methods and Statistical Classification: A New Approach to Parasite Taxonomy

We propose two new quantitative approaches to characterise evolutionary processes and species groups in *Cichlidogyrus*: phylogenetic comparative methods and statistical classification through machine learning algorithms. Both methods can shed light on different aspects of morphological and ecological characters in an evolutionary context. Phylogenetic

comparative methods (PCMs) are used to model the evolution of morphological or ecological characters. Phenotypic or ecological characters can be fitted to a tree using a range of different evolutionary models. These models reflect evolutionary processes such as genetic drift, stabilising natural selection, and adaptive radiation (see Pennell et al. 2014; Clavel et al. 2019). To the knowledge of the authors, no study has applied this multivariate approach to parasitic organisms to date. Only univariate PCMs, i.e. for single parameters, have been used for other monogeneans in combination with geomorphometric analyses such as for mugilid-infecting species belonging to *Ligophorus* (Rodríguez-González et al. 2017). Most recently, new software packages offer multivariate approaches (Goolsby et al. 2017; Clavel et al. 2019) to account for potential interactions between characters.

The exponential increase of computational power in recent years has also amplified the use of machine learning for image recognition (Kalafi et al. 2016; Wäldchen and Mäder 2018) and for an assessment of the predictive power of morphometric measurements (Zischke et al. 2016; Fang et al. 2018). We applied the latter goal to species of *Cichlidogyrus*. Amongst the existing methods, we chose supervised machine learning with support vector machines (SVMs) as this approach has been applied in other taxonomic studies using classification approaches (e.g. Zischke et al. 2016; Fang et al. 2018). Machine learning algorithms can be used to classify organisms in different hypothesised groups. In supervised machine learning approaches, a dataset is provided to train the learning algorithm to classify individuals in distinct groups. This training phase is followed by a validation phase where predictions generated by the algorithm are confirmed against a set of samples with a known affiliation. The prediction accuracy can then provide an insight into the predictive power of the model. In the last step, the validated algorithm can be applied to a test dataset of samples with unknown group affiliation.

The present study aims to investigate the evolutionary processes that lie at the source of the species-richness of *Cichlidogyrus* by applying PCMs and SVMs to morphological characters of the attachment and reproductive organs as well as host ranges data. With these approaches, we want to answer questions on the systematic informativeness of these characters and on the patterns in the character evolution. These questions are:

- Which continuous morphometric measurements that are currently widely used, are systematically informative?
- Do discretisation and body-size correction lead to information loss in morphometric and host range data?
- Can PCMs detect which evolutionary mechanisms, e.g. genetic drift, stabilising selection, or adaptive radiation, mediate the character evolution of the attachment and reproductive organs?
- Based on morphometric and host range data, how accurately can we classify species of *Cichlidogyrus* into species groups using morphological taxonomy and machine learning?

MATERIALS AND METHODS

Sampling

Fish specimens belonging to Cichlidae Bonaparte, 1835 (Cichliformes), the nothibranchiid genus *Aphyosemion* Myers, 1924 (Cyprinodontiformes: Nothibranchiidae), and *Polycentropsis abbreviata* Boulenger, 1901 (Ovalentaria, *incertae sedis*: Polycentridae) were collected during various expeditions across Africa between 2008 and 2019 and from few specimens from aquaculture facilities in France (Appendix 1). As samples collected during these expeditions have been included in previous studies, sampling details can be found in the respective publications (Vanhove et al. 2011, 2013; Muterezi Bukinga et al. 2012; Pariselle et al. 2015a, 2015b; Jorissen et al. 2018a, 2018b, 2020). We dissected the

gills of the fish and stored the samples in 96% ethanol to preserve the DNA of the parasites attached to the gills. Then, we screened the gills for parasitic infections of flatworm species belonging to *Cichlidogyrus* and *Scutogyrus*, removed the parasites from the gills using dissection needles under a stereomicroscope, mounted them temporarily in water for species-level identification at a magnification of 1000x (100x magnification with oil immersion and 10x ocular) and stored the parasite specimens individually in 96% ethanol for DNA extraction. As mostly entire specimens were used for DNA extraction, we refer to the above-mentioned articles for type and voucher specimens from the same host individuals and deposited in curated collections (see Appendix 1).

DNA extraction, amplification, and sequencing

For the DNA extraction of recent samples, we used the Nucleospin Kit (Macherey-Nagel, USA) following manufacturer guidelines but with 60 µl rather than 100 µl of elution buffer added in the final step. For some samples processed at an earlier stage, we suspended DNA samples through homogenisation with no added extraction steps following the procedures described by Marchiori et al. (2015). We amplified and sequenced three partial nuclear ribosomal genes including a fragment of the large subunit ribosomal DNA (28S rDNA) with the primers C1 (5’-ACCCGCTGAATTTAAGCAT-3’) and D2 (5’-TGGTCCGTGTTTCAAGAC-3’) (Hassouna et al. 1984), a fragment of the small subunit ribosomal DNA (18S rDNA) and the internal transcribed spacer 1 (ITS1) with the primers S1 (5′-ATTCCGATAACGAACGAGACT-3′) (Matejusová et al. 2001) and IR8 (5′-GCAGCTGCGTTCTTCATCGA-3′) (Šimková et al. 2003) and a fragment of the mitochondrial gene coding for the cytochrome oxidase *c* subunit 1 protein (COI mtDNA) with the primers ASmit1 (5’-TTTTTTGGGCATCCTGAGGTTTAT-3’) (Littlewood et al. 1997), Cox1_Schisto_3 (5’-TCTTTRGATCATAAGCG-3’) (Lockyer et al. 2003), and

ASmit2 (5'-TAAAGAAAGAACATA ATGAAAATG-3') (Littlewood et al. 1997). Reaction protocols followed the procedures of Messu Mandeng et al. (2015) for 28S rDNA, Mendlová et al. (2012) for 18S rDNA, and Vanhove et al. (2015) for COI mtDNA. We purified PCR products were with a NucleoFast 96 PCR kit (Macherey-Nagel, USA) or with a GFX PCR DNA kit and Gel Band Purification kit (GE Healthcare, USA) following manufacturer guidelines. Bidirectional Sanger sequencing was conducted according to the Big Dye Terminator v3.1 sequencing protocol (Applied Biosystems, USA) at a 1:8 dilution with an ABI PRISM 3130 Avant Genetic Analyser automated sequencer (Applied Biosystems, USA) and the primers included in the PCR protocols.

Phylogenetic Analyses

We assembled a four-locus concatenated multiple alignment from the sequences generated during this study and sequences available on GenBank (see Appendix 2). The alignment includes partial ITS1, 18S rDNA, 28S rDNA, and COI mtDNA sequence data. We aligned the sequences of each locus using the L-INS-i algorithm in MAFFT v.7.409 (Katoh and Standley 2013) as recommended for ribosomal DNA by the MAFFT manual, and removed poorly aligned positions and divergent regions using the options for less stringent parameters in Gblocks v0.91b (Talavera and Castresana 2007). We partitioned the DNA sequence data by gene and, for the COI mtDNA, by codon position. For each partition, we selected the substitution models according to the Bayesian information criterion (BIC) as rendered by ModelFinder in IQ-tree (Kalyaanamoorthy et al. 2017) using partition merging (Chernomor et al. 2016) (Table 1). We estimated tree topologies using Bayesian inference (BI), maximum likelihood (ML), and maximum parsimony (MP) methods on the individual loci and on the concatenated data set using MrBayes v3.2.6 (Ronquist and Huelsenbeck 2003) on the CIPRES Science Gateway online server (Miller et al. 2010), IQ-Tree v1.6.12 (Nguyen et al. 2015), and MPBoot v1.1.0 (Hoang et al. 2018b) respectively. For BI analyses,

we selected only models implemented by MrBayes (Table 1). *Cichlidogyrus berrebii* Pariselle & Euzet, 1994, *C. kothiasi* Pariselle & Euzet, 1994, and *C. pouyaudi* Pariselle & Euzet, 1994, parasites of tylochromine cichlids, were used to reroot the phylogenetic trees (after the phylogenetic inference) due to the well-documented early diverging phylogenetic position of these species (Mendlová et al. 2012; Messu Mandeng et al. 2015). For BI analyses, we used two parallel runs and four chains of Metropolis-coupled Markov chain Monte Carlo iterations. We ran 100 million generations with a burn-in fraction of 0.25 and sampled the trees every 1000th generation. We checked convergence criteria by assessing the average standard deviation of split frequencies (< 0.01 in all datasets) and the effective sample size (> 200) using Tracer v1.7 (Rambaut et al. 2018). For ML analyses, we estimated branch support values using both ultrafast bootstrap approximation (Hoang et al. 2018a) and Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRT) (Guindon et al. 2010) with 1000 replicates following the recommendations of the IQ-TREE manual. MP cladogram and branch support values were estimated using a fast bootstrap approximation with 1000 replicates and the default settings in MPBoot with four character states for DNA sequence data and a subtree pruning and regrafting (SPR) radius of 6 (see Hoang et al. 2018b). The performance of MPBoot was reported to be similar to established MP softwares such as PAUP and TNT (Hoang et al. 2018b). We confirmed this observation as consensus tree topologies resulting from an additional MP analysis in PAUP* v4.0a (Swofford 2003) were similar. For PAUP*, we used unordered, equally weighted and non-additive characters and a traditional heuristic search using 500 random sequence addition (RAS) replicates and tree bisection and reconnection (TBR) branch swapping, and treating gaps treated as missing data. We considered nodes with a BI posterior probability (PP) ≥ 0.95, an ultrafast bootstrap

values for ML and MP ≥ 95, and SH-aLRT statistic ≥ 80 as well-supported (Hoang et al. 2018a, 2018b).

To assure congruence of the four DNA markers, we analysed the congruence of the BI, ML, and MP single-locus tree topologies using Congruence Among Distance Matrices (CADM) tests (Legendre and Lapointe 2004; Campbell et al. 2011). For these tests, we included only the nuclear markers (ITS1, 18S, 28S rDNA) because of the low number of specimens (nine) covering all four markers. We also applied this test to test for congruence amongst the BI, ML, and MP concatenated tree topologies. We used the 'ape' package v5.3 (Paradis and Schliep 2019) in R v4.0.0 (R Core Team 2019) to calculate phylogenetic pair-wise distance matrices and to conduct the CADM test.

Morphological and Ecological Data Collection

We assembled morphometric host range data for all species included in the phylogenetic analyses (Table 2). Raw measurements were taken from AP's personal morphometric database and from the raw data of previous publications kindly provided by the authors of the cited studies (Pariselle and Euzet 2003; Vanhove et al. 2011; Gillardin et al. 2012; Muterezi Bukinga et al. 2012; Pariselle et al. 2013; Řehulková et al. 2013; Messu Mandeng et al. 2015; Van Steenberge et al. 2015; Igeh et al. 2017; Rahmouni et al. 2017, 2018; Kmentová et al. 2016a, 2016b, 2016c; Geraerts et al. 2020). These raw data were deposited at MorphoBank (www.morphobank.org, project XXXXX). We calculated the standard errors for each measurement from the raw data. However, for species where no or only partial raw data were available, we used the mean values provided in the literature (Douëllou 1993; Pariselle and Euzet 2003; Řehulková et al. 2013; Rahmouni et al. 2017; Jorissen et al. 2018a, 2018b) without standard errors. If no mean was reported, we inferred the measures from drawings provided in previous publications (Dossou 1982; Dossou and

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Birgi 1984; Birgi and Lambert 1986; Muterezi Bukinga et al. 2012). Measurements of *Cichlidogyrus bifurcatus* Paperna, 1960, *C. dionchus* Paperna. 1968, and *C. longipenis* Paperna & Thurston, 1969 were not included as no mean values were provided in respective publications (Paperna 1979, 1968, 1960) and drawings were often incomplete with some structures, e.g. the second hook, entirely missing. For the auxiliary plate (AuP) of the male copulatory organ, we calculated the surface area from the plate length and width assuming an ellipsoid shape. We used this derived measurement including the relative error resulting of the sum of relative errors of the original measurements in all subsequent analyses. In addition, we classified all species in proposed categories for the haptor morphology and host range, which include the configuration of the hooks (Vignon et al. 2011), the similarity of the anchors, and the shape of the ventral bar (Mendlová et al. 2012), and the index of specificity (IS) (Mendlová and Šimková 2014). We inferred IS data from the host-parasite interactions listed by Cruz-Laufer et al. (2020). For the classification of cichlid species into tribes, we followed the approach of Dunz and Schliewen (2013).

Cluster Analysis

To minimise model complexity and avoid overfitting statistical models, we reduced the number of continuous variables used for downstream analyses, i.e. phylogenetic comparative methods and statistical classification. Therefore, we identified clusters of multicollinear measurements and excluded all but one measurements from each cluster and all other measurements without a detectable phylogenetic signal. Multicollinearity amongst the variables can lead to unstable and unreliable models (Mundry 2014). Therefore, we used the Pearson pairwise correlation matrix to detect clustered variables (Dormann et al. 2013). If the absolute values of the Pearson correlation coefficients (|r|) exceeded 0.7 (Dormann et al. 2013), we considered the diagnosis positive.

To detect phylogenetic signals in the clusters, we modelled the character evolution of the proxy variables on 100 random samples of tree topologies drawn from the post–burn in dataset of the BI analysis. We excluded all but one specimen per species from the phylogenetic trees to avoid pseudoreplication in the phylogenetic comparative analyses. Specimen included in the subset trees, were chosen at random and are highlighted in Table 2. We used the 'fitContinuous' function in the R package 'geiger' v2.0 (Pennell et al. 2014). To account for multiple evolutionary processes, we fitted different models of character evolution implemented in the R package including the Brownian motion, Ornstein-Uhlenbeck, and Early Burst (detecting both decelerating and accelerating divergence rates) models (see below for model specifications). We then compared the model fits to a null model fit (H_0) by using a tree transformation with Pagel's $\lambda = 0$ (Pagel 1999) on the BI consensus tree topology, which assumes absolute phylogenetic independence of the measurements. To assess model fits, we used the sample size–corrected Akaike information criterion (AICc). We inferred significant differences from H_0 directly through the precentage of AICc values larger than or equal to the AICc of H_0. For each cluster, we selected the measurements with strongest phylogenetic signal as proxy variable for multivariate phylogenetic signal detection. Measurements that did not form part of a cluster yet showed a detectable phylogenetic signal, were also included as proxy variables. The strength of the phylogenetic signal was assessed as mean difference of AICc values of the different models and H_0.

Character Evolution of the Attachment and Reproductive Organs

Using the proxy variables resulting from the cluster analysis, we tested the possible effect of genetic drift (H_1) stabilising selection (H_2) and adaptive radiation with a decelerating (H_3) and accelerating divergence of character (H_4) on the evolution of the attachment and reproductive organs. To detect these effects, we fitted a range of multivariate models of continuous character evolution to the 100 randomly sampled and subset trees used for the

cluster analysis. Similar to the cluster analysis, we compared the performance of these models to the null model (H_0). The tested models include:

- the Brownian motion model (BM) (Felsenstein 1973), which is used to model a 'random walk', i.e. genetic drift (H_1),
- the Ornstein-Uhlenbeck model (OU) (Butler and King 2004) modelling the character evolution under stabilising selection (H_2),
- the Early-Burst (EB) (Harmon et al. 2010) or accelerate-decelarate model (ACDC) (Blomberg et al. 2003), which is used to model an early rapid evolution of characters followed by a slow-down such as might occur during adaptive radiation events (H_3),
- the Late-Burst (LB), a variation of the ACDC model with an accelerating divergence of characters (Blomberg et al. 2003) (H_4),
- and, the null model (H_0), which assumes absolute phylogenetic independence by transforming the tree transformation with Pagel's $\lambda = 0$ (Pagel 1999) resulting in a star-like phylogeny.

We fitted all multivariate models using the 'fit_t_pl' function (Clavel et al. 2019) in the R package 'RPANDA' v1.8 (Morlon et al. 2016) as recommended for independently evolving measurements with a natural orientation. In addition, we expected the cluster analysis to minimise interdependence of the measurement variables. Therefore, we used a restricted maximum likelihood (REML) estimation with a leave-one-out cross-validation of the penalized log-likelihood (PL-LOOCV), an estimated within-species variation, an archetypal ridge estimator, and a diagonal unequal variance target matrix (Clavel et al. 2019). To run the LB model with 'fit_t_pl', we set the upper bound of the EB search algorithm to 10. Similar to the univariate analyses with the R package 'geiger', we compared all model fits to a null model fit (H_0). As traditional information criteria such as AICs are not applicable to penalised

likelihoods (Clavel et al. 2019), we assessed the model fit with the generalised information criterion (GIC) (Konishi 1996) using the 'GIC' function in the R package 'RPANDA'. We inferred significance levels directly from the percentage of GIC values large than or equal to the the GIC of H_0. Model performance was summarised as mean difference (ΔGIC) of GIC values of H_1–H_4 and H_0.

To assess the effects of data discretisation, we modelled the discrete character evolution of the categorical data under a univariate Pagel's λ model (Pagel 1999) using the 'fitDiscrete' function in the R package 'geiger' v2.0 (Pennell et al. 2014). As before, we applied this function to all 100 randomly selected BI subset trees to account for phylogenetic uncertainty.

Taxonomic and Statistical Classification

We used a statistical classification approach to infer the discriminative power of the measurements and species characterisations currently provided by taxonomic literature. Specifically, we classified species not included in the phylogenetic analysis in the species groups inferred from the tree topology in Figure 1. We used a taxonomic literature survey to validate the results of the statistical classification. Hence, we summarised key features and the host range (by species and tribe of cichlids) of each species group. We assigned species of *Cichlidogyrus* and *Scutogyrus* not included in the phylogenetic analyses to these clades based on the morphology while taking host and geographical ranges into consideration. Natural and invasive host ranges and geographical ranges by country were taken into account as we expected the host environment to be a relevant factor in allopatric speciation processes of these parasites as proposed by Huyse et al. (2005) and McCoy (2003). We used the host-parasite list and literature databases provided by Cruz-Laufer et al. (2020) to summarise the host ranges.

For statistical classification, we employed a supervised learning algorithm using support vector machines (SVMs). We carried out the classification analysis through a C-classification with a radial basis kernel function as implemented in the 'svm' function of the R package 'e1071' (Meyer et al. 2020) and including only the proxy variables from the cluster analysis to minimise the data loss caused through empty values and to avoid overfitting. The *C* and γ parameters of the kernel function were optimised through a grid-search with exponentially growing sequences as suggested by Hsu et al. (2003) using a tenfold cross-validation procedure as implemented in the 'tune' function of the R package 'e1071' (Meyer et al. 2020). To assess the performance of the algorithms, we computed accuracy estimates using the 'confusionMatrix' function in R package 'caret' (Kuhn 2020) as well as through comparison with the results of the taxonomic classification approach.

Graphing

We plotted the graphs and phylogenetic trees using the R packages 'ggplot2' (Wickham 2016) and 'ggtree' (Yu et al. 2017, 2018). Based on the best fitting model, we mapped ancestral character states of the continuous measurements on the BI consensus tree with unsupported nodes collapsed to polytomies (Fig. 1) and sequences of the same species dropped from the topology as for the phylogenetic comparative analysis (Table 2). To calculate the node states, we used generalised least squares as implemented in the 'ancestral' function in R package 'RPANDA' (Clavel et al. 2019). To map of discrete characters on the consensus tree, we estimated character states through stochastic mapping using an equal-rates model as implemented by the 'make.simmap' function in the R package 'phytools' v0.7-47 (Revell 2012).

RESULTS

Phylogenetic Analyses

We generated 60 sequences for 33 species and 64 specimens of *Cichlidogyrus* and *Scutogyrus* including 47 28S, 14 combined 18S and ITS1 rDNA, and five COI mtDNA sequences. For 21 of these species, these sequences are the first to be published representing a 57% increase in taxon coverage to 42% of all currently described species, 45% of described species of *Cichlidogyrus* and 71% of *Scutogyrus*. The alignments include 93 sequences/820 base pairs for 28S rDNA, 54 sequences/481 base pairs for 18S rDNA, 57 sequences/383 base pairs for ITS rDNA, and 19 sequences/489 base pairs for CO1. The combined data set includes sequences of103 specimens belonging to 58 species (see Appendix 1), and has a length of 2173 base pairs following the removal of poorly aligned positions and divergent regions. We deposited sequences generated for this study in GenBank (XXXXXX – XXXXXX). All sequences are listed in Appendix 2 including the respective GenBank accession numbers, the host species, and the sampling locations. The concatenated alignments and BI, ML, MP tree topologies can be accessed at Treebase (www.treebase.org, accession number: XXXXX). Substitution models used for the different partitions are provided in Table 1. All COI codons were merged into a single partition. As topological congruence between single-locus 18S, 28S, and ITS rDNA trees was high for BI (Kendall's $W = 0.74$, $\chi^2 = 2506$, $p < 0.01$), ML (Kendall's $W = 0.75$, $\chi^2 = 2523$, $p < 0.01$), and MP (Kendall's $W = 0.65$, $\chi^2 = 2117$, $p < 0.01$), we proceeded with the concatenated analysis for all four genes. Because of the topological congruence of the BI, ML, and MP concatenated trees (Kendall's $W = 0.71$, $\chi^2 = 11044$, $p < 0.01$), we only displayed the BI phylogram but report support values of BI, ML, and MP analyses (Figure 2).

We found 10 well-supported clades within *Cichlidogyrus* (including *Scutogyrus*) based on the node support values inferred from BI, ML, and MP analyses (Fig. 2). The clades and the their respective node support values (PP/UFBoot/SH-aLRT/MPBoot) are: The

'Hemi+' clade (1.00/100/100/100), *Scutogyrus* (1.00/100/99/100) without *S. vanhovei* Pariselle, Bitja Nyom & Bilong Bilong, 2013), the *'Cop+'* clade (1.00/100/100/*), the *'Oreo1+'* clade (1/99/98/*), the *'EAR'* clade (1.00/100/100/*), the *'CPO+3'* clade (1.00/100/100/100), the *'Bulb'* clade (1/100/100/*), the *'Oreo2+'* clade (1.00/97/94/95), the *'Halli'* clade (1.00/100/100/100), and the basal *'Tylo'* clade (1.00/100/100/100). Beyond the well-supported monophyly of all species excluding the *Tylo* species (out)group (1/100/100/100), the relationship between the other species groups remains unresolved with only partial support for macroclades containing *Oreo1+*, *Cop+*, *Scutogyrus*, and *Hemi+* (0.99/*/*/*) or *Cop+*, *Scutogyrus*, and *Hemi+* (*/*/87/*). Both macroclades would also include *C. tilapiae* Paperna, 1960, which forms no well-supported cluster with the other clades. However, we refrained from creating another monospecific clade similar to the '*Halli*' clade. *Cichlidogyrus halli* (Douëllou 1993; Jorissen et al. 2018a) and *C. tilapiae* (Pouyaud et al. 2006) have both been hypothesised to be species complexes. However, this study includes only a small number of sequence and morphometric data for *C. tilapiae*. Furthermore, *Scutogyrus vanhovei* appears to be basal to other species of *Scutogyrus* with moderate support (0.99/*/90/*).

Cluster Analysis

We assembled morphometric and host specificity data for all 134 species of *Cichlidogyrus* and *Scutogyrus* including the 58 species of the phylogenetic tree (Figure 2) All discrete and continuous morphometric and host specificity data are available at TreeBase alongside the DNA sequence alignments and tree topologies (www.treebase.org, accession number: XXXXX). Raw morphometric data are publicly available at MorphoBank (www.morphobank.org, project XXXXX).

Based on a Pearson pairwise correlation matrix (Appendix 3), we detected five clusters of multicollinear measurements. Furthermore, we detected a phylogenetic signal in 17 measurements resulting in a selection of nine proxy variables from the attachment organ (for information on the clustered variables and mean and standard deviation of ΔAICc, see Table 2). No measurement of the reproductive organs produced a significant phylogenetic signal. Measurements from the fifth cluster (E) produced no detectable phylogenetic signal (Table 2).

Character Evolution of the Attachment and Reproductive organs

The null model H_0 was outperformed by every other model (Fig. 3) in the multivariate phylogenetic comparative analysis of the nine proxy variables. This result confirms that the proxy variables express a significant phylogenetic signal. Brownian Motion (BM), Ornstein-Uhlenbeck (OU), and Early-Burst (EB) models all performed equally well (Fig. 3). The Late Burst (LB) performed slightly better. However, the difference to the other models was not significant. Out of the discrete characters, only the hook configuration (HC) and ventral bar shape (VBS) produced strong phylogenetic signals (Fig. 4). A character map (Fig. 5a) illustrates the phylogenetic pattern of the HC. We detected a relatively weak phylogenetic signal for the anchor similarity (AS) and no signal for the index of specificity (IS). A character map of the IS (Fig. 5b) confirms the lack of a phylogenetic pattern.

Taxonomic Literature Survey and Statistical Classification

Out of the 10 species groups inferred from the phylogenetic analyses (Fig. 2), we found shared morphological characteristics in eight. We were able to affiliate all but 12 species without available DNA sequences to these clades (Appendix 5). Some morphological structures were particularly relevant for characterising species groups of *Cichlidogyrus*. Character states of these structures were frequently shared and defining features of the

proposed species groups (Appendix 4; also Fig. 6, Fig. 7, and Fig. 8). Based on the number of species groups defined by a structure, the most distinctive structure was the male copulatory organ (MCO) as phenotypes were often specific to a species group or subgroup (Fig. 6). Consequently, the MCO was included in the characterisation of eight species groups (Fig. 6). The size and shape of the first hook pair (U1) was used to characterise four species groups (Fig. 6; Fig. 7a, c) and one subgroup of the morphologically diverse *EAR* group. The length of the auricles of the dorsal bar (DB h) was used for two species groups (Fig. 6; Fig. 7a, d) and one subgroup of the *EAR* group. The size of the hook pairs 3–7 was used for two groups (Fig. 6; Fig. 7b, c, d). For detailed morphological characterisations of the species groups and a discussion on the host and geographical ranges, see below and the overview of the morphological features highlighted in Figure 6 (full species names are cited in Appendix 4 and reported alongside key features and host names; for abbreviations of morphometric measurements, see Table 2).

Accordance of the proposed group affiliation of the species with the SVM-based statistical classification approach (Appendix 5) was overall 71% (CI: 67%, 75%) when using the nine proxy variables of the attachment organ with 715 observations in the training data and 609 observations in the test data (not including observations of species without proposed species group). Model tuning provided the best result for $C = 2$ and $\gamma = 0.5$. The accordance rate of taxonomic and SVM-based approach varied substantially between the different species groups (Table 3) ranging from 100% for species classified in the species group of *Scutogyrus* to 25% and 0% respectively for species classified in the *Tylo* and *Bulb* species groups. Species of the *Oreo2+* and the *EAR* species groups shared no apparent features in their attachment and reproductive organs with species of the same group apart from the

characteristics shared by all species of the genus. Yet the accordance of the classification approaches was above average for the *EAR* group with 88%.

Characterisation of Species Groups

1. *Hemi+: Species infecting hemichromine cichlids and non-cichlid fishes*. Dorsal anchor with elongated narrow inner root and rounded outer root. Ventral anchor slightly larger with short and robust inner root. Dorsal bar slightly arched with short auricles. Ventral bar arched with small membranous extensions. First hook pair (U1) large and well-developed (Fig. 6; Fig. 7c) except for *C. amieti*. Hook pairs 3–7 medium-sized to large with bases more developed than for second hook pair. MCO consists of thin and long penis that frequently forms loops (*C. amieti, C.* cf. *bychowskii, C. dracolemma, C. inconsultans, C. kmentovae, C. nandidae*) (Fig. 3) or spirals with one (*C. calycinus, C. teugelsi*) to multiple turns (*C. euzeti, C. longicirrus, C. polyenso, C sanseoi*) and re-joins the accessory piece in its distal portion, and an accessory piece of two distinct portions. These portions can be shaped like a V or a simple spiral with an expanded knee-like bend (*C. amieti, C. calycinus, C.* cf. *bychowskii, C. dageti, C. dionchus, C. dracolemma, C. falcifer, C. kmentovae, C. nandidae, C.teugelsi*) (Fig. 3) (see Dossou and Birgi 1984; Pariselle and Euzet 2004), or a large spiral followed by a non-spiralled distal portion (*C. euzeti, C longicirrus, C. polyenso, C. sanseoi*). In the presence of the knee-like bend a heel is visible. The sclerotised vagina is long, sinuous (except for *C. dageti, C. falcifer, C. kmentovae*) or spiralled (*C. longicirrus, C. sanseoi*). Species of the *Cop+* and *Tylo* groups share some characteristics with the species of the *Hemi+* group but have either a substantially shorter penis or do not display the characteristic knee-like bend in the accessory piece. Previous studies have grouped species of the *Hemi+* species group together before based on phylogenetic (Mendlová et al. 2012; Řehulková et al. 2013) and morphological (e.g. Dossou and Birgi 1984; Pariselle and Euzet 2004; Jorissen et al. 2018a) analyses.

2. *Scutogyrus: Species infecting oreochromine and pelmatolapiine cichlids*. This group has been described as separate genus *Scutogyrus* including a detailed morphological characterisation (Pariselle and Euzet 1995a). Concerning the haptor morphology, species of *Scutogyrus* displayed well-developed hook pairs 3–7, elongated auricles at the dorsal bar (Fig. 6; Fig. 7d, d) and a ventral bar that supports a large plate (Fig. 6).

3. *Cop+: Species infecting coptodonine cichlids*. Anchor pair dissimilar in shape. Dorsal anchor with V-shaped indentation between the roots and an elongated inner root at least four times longer (Fig. 2B) than outer root. Ventral anchor with shallower indentation, and relatively short roots including a considerably shorter inner root. Dorsal bar thick with large and broad auricles. Ventral bar V-shaped and truncated at the ends and thinner towards the mid-portion, displays membranous extensions. First hook pair (U1) well-developed (Fig. 6; Fig. 7b). Hook pairs 3–7 (U3–7) small. MCO consists of short tubular penis with a marked heel and an accessory piece with a sickle-like terminal hook connected to the basal bulb of the Pe. No sclerotised vagina has been observed for species belonging to this group except for *C. digitatus* in the original description (Dossou 1982). However, none of the more recent reports of species of the *Cop+* clade (Pariselle and Euzet 1996, 2003) remark on its presence in the original description. Previous studies have grouped species of the *Hemi+* clade, e.g. *C. dionchus* (Pariselle and Euzet 2003), and the *Tylo* clade, e.g. *C. kothiasi* (Pariselle et al. 2013), with the species of the *Cop+* clade because of the similar morphology of the hook pairs, i.e. large U1 and small U3–7. However, none of these species displays an accessory piece with a sickle-like terminal hook. The species of the *Cop+* clade have previously been grouped together (Pariselle and Euzet 2003; Jorissen et al. 2018b) with the omission of *C. berminensis*.

4. *Oreo1+: Species infecting oreochromine and coptodonine cichlids*. Anchor pairs similar in shape and size but dorsal anchor somewhat smaller. Anchors with short and broadened outer root and long inner root. Dorsal bar arched with two broad and large auricles. Ventral bar with thin mid-portion, and thickened towards the ends includes triangular membranous extensions. Hooks small with small shaft. MCO consists of long tubular (*C. cirratus, C. giostrai, C. mbirizei, C. mvogoi, C. ornatus*) (Fig. 6) or short and thickened (*C. acerbus, C. lagoonaris, C. nageus, C. njinei, C. slembroucki*) penis sometimes with bulbous extension at the base of the tube (*C. njinei, C. ornatus, C. giostrai*) and complicated roughly C-shaped (Fig. 6) accessory piece often with finger or hook-shaped outgrowths and marked heel. Proximal end attached to Pe bulb and distal end holds the distal end or a mid-portion of the penis. Vagina simple (*C. acerbus, C. lagoonaris, C. slembroucki*), sinuous (*C. giostrai, C. njinei, C. ornatus*), or coiled (*C. cirratus, C. mbirizei, C. mvogoi*) tube and only slightly sclerotised.

5. *EAR: Species infecting cichlids from the East African Radiation*. The *EAR* species group comprises a large heterogeneous group of species that infect cichlids belonging to East African lineages, i.e. the East African Radiation (Schedel et al. 2019). The *EAR* clade consists of multiple potential subgroups based on morphological features (Fig. 6) and node support values from the phylogenetic analyses. Frequently, the character states of these subgroups represent a spectrum rather than fixed categories. The subgroups of the *EAR* group include species with a long heel infecting non-haplochromine cichlids such as bathybatine, ectodine, boulengerochromine, and possibly cyphotilapiine cichlids (*Heel*), species infecting cyprichromine and trematocarine and potentially eretmodine cichlids (*CT*), species infecting tropheine cichlids (*Troph*), and species sampled from Lake Victoria (*Victoria*) (see Figure 6 except for *Victoria* as no DNA sequence data are currently available for this group). The morphology of the attachment and reproductive organs vary substantially between and within

the subgroups (Fig. 6). Hence, we observed no shared morphological features. However, some species subgroups display the following shared morphological features.

- *Heel*. Dorsal anchor unlike ventral with elongated inner root. Dorsal bar with short auricles. Well-developed first hook pair. MCO with short penis sometimes with spirally screw thread-like thickened wall (*C. casuarinus*, *C. centesimus*, *C. nshomboi*), characteristic elongated heel of one third to twice the length of the penis, and reduced filiform or sometimes missing (only *C. centesimus*) accessory piece (Fig. 6). The species of the *Heel* subgroup have previously been grouped together (Muterezi Bukinga et al. 2012), most recently by Rahmouni et al. (2018).
- *CT*. Dorsal and ventral anchors similar in shape but can vary in size (see *C. brunnensis*) with deep to shallow (*C. brunnensis*) indentation. Dorsal bar with short auricles. Hooks generally short. MCO consists of penis and accessory piece and sometimes a marked heel (*C. brunnensis*). Penis medium-length, broadened, and mostly straight (*C. brunnensis*, *C. evikae*, *C. jeanloujustinei*, *C. milangelnari*, *C. sturmbaueri*) with thickened wall. Accessory piece with bifurcate distal end (*C. brunnensis*), or in two-parts (*C. evikae*, *C. jeanloujustinei*, *C. milangelnari*). Similarities of some species of this subgroup with species infecting cyphotilapiine and ectodine cichlids with elongated auricles (*C. adkoningsi*, *C. discophonum*, *C. glacicremoratus*, *C. koblmuelleri*, *C. makasai*, *C. vandekerkhovei*) or elongated marginal hooks (*C. rectangulus*, *C. sturmbaueri*) have previously been discussed (Rahmouni et al. 2017, 2018) but the phylogenetic analysis provided no support for a monophyletic group including species of the *CT* subgroup and *C. vandekerkhovei* (Fig. 2).
- *Troph*. These species share few similarities and were placed in a subgroup due to the host species, which all belong to the tropheine radiation. Species without available DNA

sequences sharing morphological features with species in this group, were also added to this subgroup. Dorsal anchor unlike ventral with elongated inner root (except for *C. antoineparisellei*) and deeper indentation between the roots. However, indentation generally shallow, sometimes with fused roots (see parasites of *Interochromis loocki* including *C. antoineparisellei*, *C. buescheri*, *C. schreyenbrichardorum*, *C. vealli* on the ventral and *C. antoineparisellei* and *C. buescheri* on the dorsal anchor) (Fig. 6). Dorsal bar with well-developed but not elongated auricles. Ventral bar frequently with membranous extensions. Hooks generally small but can be slightly more developed (e.g. *C. irenae, C. masilyai, C. salzburgeri*). MCO consists of penis and accessory piece (AP), both variably shaped; the heel fis inger-like if present. Penis thin and arched (*C. antoineparisellei*, *C. buescheri*) or looped (*C. schreyenbrichardorum*) or slightly broadened (*C. vealli*) (Fig. 6) for parasite of *Interochromis loocki*. Penis broadened to wide and short (reminiscent of *C. halli*) for other species. AP simple, elongated with few remarkable structures, sometimes with thickened (*C. gistelincki*) or forked (*C. antoineparisellei, C. buescheri, C. masilyai, C. salzburgeri*), or hook-shaped (*C. raeymaekersi, C. salzburgeri*) distal end.

- *Victoria*. No recent reports of *C. haplochromii* and *C. longipenis* exist but both species have been reported from Lake Victoria (Pariselle and Euzet 2009). Anchors similar in shape and size. Dorsal bar with short auricles. Hooks slightly developed with presence of small shaft. MCO simple with tubular penis, no heel, and simple accessory piece.
- *Other species*. The remaining species of the *EAR* species group shared no evident features with any of the proposed subgroups. *Cichlidogyrus attenboroughi* appears to be the sister taxon to the *Heel* group with moderate support (0.97/*/*) but shares no notable characteristics with these species. Several species infecting cyphotilapiine and ectodine cichlids display a broadened (*C. glacicremoratus*, *C. koblmuelleri*) or straight (*C.*

adkoningsi, C. rectangulus, C. sturmbaueri) penis with an accessory piece in two-parts (*C. glacicremoratus, C. koblmuelleri*) or with a bifurcated distal end (*C. makasai, C. rectangulus, C. sturmbaueri, C. vandekerkhovei*) similar to species of the *CT* group. Yet many of these species also display elongated auricles associated with the dorsal bar (*C. adkoningsi, C. discophonum, C. glacicremoratus, C. koblmuelleri, C. makasai*) similar to *C. vandekerkhovei*, which appears to be unrelated to species of the *CT* group (Fig. 2) but also displays a bifurcated distal end. According to the phylogenetic analyses (Fig. 2), *C. consobrini* and *C. gillardinae* form a monophyletic group. However, we found few shared characteristics except for a simple tubular penis with a small heel and a simple accessory piece. These characteristics might place these two species close to the *Victoria* subgroup. Furthermore, both subgroups infect haplochromine hosts similar to species from the *Troph* subgroup.

6. *CPO+3: Species infecting coptodonine, pelmatolapiine, oreochromine, tilapiine, heterotilapiine, and gobiocichline cichlids*. Dorsal anchor with inner root longer than outer root and V-shaped indentation between the roots but inner root of *C. arthracanthus* considerably longer than of other species, and roots of *C. cubitus* and *C. louipaysani* fused. Ventral anchor similar shape to dorsal anchor but slightly larger. However, roots of ventral anchor of *C. tiberianus* fused. Dorsal bar slightly arched with large medium-sized to large auricles. Ventral bar always with membranous extensions. First hook pair (U1) small and hook pairs 3–7 (U3–7) very long, i.e. more than triple the size of first pair (Fig. 6; Fig. 7c) but U1 of *C. arthracanthus* large and U3–7 of *C. cubitus* and *C. louipaysani* short (Fig. 6; outliers in Fig. 7c). MCO consists of long, arched or sometimes spiralled (*C. arthracanthus*), tubular penis with a well-marked irregularly shaped heel, and a massive, roughly S-shaped accessory piece (Fig. 6) that is frequently connected to the heel. The AP has an extension or

thickening at the first turn in the proximal half and frequently displays a folded back (*C. paganoi, C. vexus*), straight and pointy (*C. guirali*), or hook-like (*C. bilongi, C. douellouae, C. ergensi, C. gallus, C. legendrei, C. microscutus*) distal end, or sometimes additional terminations resulting in a furcate ending with two (*C. aegypticus, C. agnesi, C. anthemocolpos, C. bouvii, C. cubitus, C. flexicolpos, C. lemoallei, C. louipaysani, C. ouedraogoi, C. testificatus, C. thurstonae, C. tiberianus*) or three (*C. bonhommei, C. hemi, C. kouassii*) digitations. However, the first turn is never V-shaped or knee-like as in species of the *Hemi+* clade and the hook-shaped termination is never sickle-like such as in species of the *Cop+* clade. Several species display an auxiliary plate (AuP) close to the distal end of the MCO including *C. aegypticus, C. agnesi, C. bilongi, C. gallus, C. guirali* (two pieces), *C. microscutus, C. paganoi,* and *C. thurstonae*. A sclerotised vagina is mostly present but has not been reported for *C. arthracanthus* (Paperna 1960; Ergens 1981). Other species with more developed hook pairs 3–7 include *Cichlidogyrus bulbophallus, C. flagellum, C. lobus,* and *C. ranula,* some species of the *Hemi+* clade, and all species of the *Heel* subgroup of the *EAR* clade. *Cichlidogyrus bulbophallus*, *C. calycinus*, and species of the *Hemi+* clade and the *Heel* subgroup have been likened to *C. arthracanthus* because of the additional well-developed first hook pair (U1) (Geraerts et al. 2020). However, all of these species differ from the confirmed clade members as follows.

- The dorsal and ventral anchors differ in shape and size but are similar for species of the *CPO+3* clade.
- The penis is never long, arched, and tubular but short (*Heel*), bulbous (*C. bulbophallus, C, flagellum, C. ranula*), or draws a loop (Fig. 6) or long spiral (*Hemi+*).
- The accessory piece is not complex and S-shaped with an extension in the proximal half but mostly reduced (*Heel*), simple (*C. bulbophallus, C, flagellum, C. ranula*), or has a sharp knee-like bend (*Hemi+*).

Previous phylogenetic studies (Mendlová et al. 2012; Messu Mandeng et al. 2015) have placed *C. cubitus* alongside species of the *CPO+3* clade with long marginal hooks ("*tiberianus*" group, see Pouyaud et al. 2006) but morphological similarities were not further discussed.

7. *Bulb: Species from Southern Africa with a bulbous penis*. Dorsal and ventral anchors dissimilar. Dorsal anchors with deep indentation between short outer root and a long inner root (Fig. 6). Dorsal bar short and stout with auricles medium-sized to large (average > 20 µm for *C. zambezensis*). First hook pair (U1) generally well-developed (Fig. 6) albeit somewhat less in *C. zambezensis* (Fig. 7a). Hook pairs 3–7 small to slightly developed. Male copulatory organ consists of a broad penis with a swollen, bulbous middle portion, a small heel (Fig. 6), and a sharp tube-like (*C. philander*) or filiform (*C. maeander*, *C. papernastrema*, *C. pseudozambezensis*, *C. zambezensis*) termination, and a simple accessory piece of similar length, which is articulated to the base of the penis bulb (Price et al. 1969; Douëllou 1993; Jorissen et al. 2018b). *Cichlidogyrus amphoratus* Pariselle & Euzet, 1996, *C. bulbophallus* Geraerts & Muterezi Bukinga in Geraerts *et al*., 2020, *C. flagellum* Geraerts & Muterezi Bukinga in Geraerts *et al*., 2020, *C. giostrai* Pariselle, Bilong Bilong, & Euzet, 2003, *C. karibae* Douëllou, 1993, *C. lagoonaris* Paperna, 1969, *C. njinei* Pariselle, Bilong Bilong, & Euzet 2003, *C. ornatus* Pariselle & Euzet, 1996, *C. sanjeani* Pariselle & Euzet, 1997 also present a swollen portion in the penis and a simple accessory piece but differ from the species of the group as follows.

- *C. amphoratus* has similar dorsal and ventral anchors with no well-developed inner root of the doral anchor (Fig. 6)

- The first hooks (U1) of *C. amphoratus* (Fig. 6), *C. flagellum, C. giostrai, C. karibae, C. lagoonaris, C. njinei, C. ornatus, C. ranula,* and *C. sanjeani* are considerably shorter and less developed.
- The hooks U3–U7 of *C. bulbophallus* and *C. ranula* are considerably longer and resemble the marginal hooks of the species of the *CPO+3* species group (VI).
- The tubular distal portion of the penis of *C. amphoratus* (Fig. 6), *C. bulbophallus, C. flagellum, C. giostrai, C. njinei* and *C. ornatus* exceed the length of the swollen portion of the penis considerably.
- *C. amphoratus, C. giostrai, C. lagoonaris, C. njinei, C. ornatus,* and *C. sanjeani* have only been reported from Western Africa not Southern Africa (see Pariselle and Euzet 2009).
- *C. amphoratus* and *C. njinei* belong to the *Cop1+* and *Oreo2+* species groups respectively.

8. *Oreo2+: Two species infecting oreochromine and coptodonine cichlids*. Both species display simple features such as similar anchor pairs and small hook pairs. Most other features differ considerably between the species. Dorsal and ventral anchors are similar in size and shape, roots variable in length, e.g. long inner root in *C. amphoratus* (see drawing Pariselle and Euzet 1996) or barely to no distinct roots in *C. sclerosus* (Douëllou 1993). Dorsal bar X-shaped with two auricles of variable shape and length. Ventral bar with membranous extensions (Mendlová et al. 2012) variable in shape and size. Hooks small with developed but small shaft except for the second pair (Douëllou 1993; Pariselle and Euzet 1996). MCO consists of arched, tubular penis with large heel ('subspherical' in *C. amphoratus* or shaped as a 'serrated plate' in *C. sclerosus*), and associated relatively simple accessory piece (Douëllou 1993; Pariselle and Euzet 1996). Sclerotised vagina present. *Cichlidogyrus amphoratus* presents a swollen middle portion of the penis reminiscent of species of the *Bulb*

clade (Pariselle and Euzet 1996). However, in the latter group, the dorsal anchors have long inner roots and the first hook pair is well developed (Fig. 6; Fig. 7a). This species group has previously been reported but was not further discussed (see Messu Mandeng et al. 2015).

9. *Halli*: *The species complex of Cichlidogyrus halli*. For detailed characterisation see original species description by Price and Kirk (1967).

10. *Tylo*: *Species infecting tylochromine cichlids*. Dorsal and ventral anchors dissimilar. Roots of anchors frequently fused forming fenestrae (windows), e.g. for *Cichlidogyrus berrebii, C. chrysopiformis, C. dijetoi, C. kothiasi, C. pouyaudi* (Fig. 6), *C. sergemorandi* (Pariselle and Euzet 1994; Pariselle et al. 2014; Rahmouni et al. 2018). Dorsal anchors with short outer root and a long inner root, median portion of blade hollow in *C. berrebii, C. pouyaudi* (Pariselle and Euzet 1994), and *C. sigmocirrus* (Pariselle et al. 2014). Dorsal bar with two generally short auricles fused to the bar (Fig. 6; Fig. 7a) is one of the most distinctive features of species of *Cichlidogyrus* infecting tylochromine cichlids (Pariselle et al. 2014; Rahmouni et al. 2018). Hook pairs generally small but developed first pair in *C. bixlerzavalai, C. chrysopiformis, C. dijetoi, C. kothiasi*, and *C. muzumanii* (Fig. 7a). Male copulatory organ consists of a coiled tubular penis, a heel, and a flat ribbon-like (*C. berrebii, C. kothiasi, C. pouyaudi, C. sigmocirrus*) (Pariselle and Euzet 1994; Pariselle et al. 2014), drape-like (*C. chrysopiformis, C. dijetoi, C. mulimbwai, C. omari, C. sergemorandi*) (Muterezi Bukinga et al. 2012; Pariselle et al. 2014; Jorissen et al. 2018a; Rahmouni et al. 2018), or reduced (*C. bixlerzavalai, C. muzumanii*) (Muterezi Bukinga et al. 2012; Jorissen et al. 2018a) accessory piece. Unlike other species of *Cichlidogyrus*, the penis and accessory piece of the MCO are separated at the base but might be connected through a filament in *C. chrysopiformis* and *C. sigmocirrus* (Pariselle et al., 2014). Species of the *Tylo* group have

regularly been grouped together since the descriptions by Pariselle and Euzet (1994) with most recent additions provided by Jorissen et al. (2018a) and Rahmouni et al. (2018).

Other species. For 12 species, we no found morphological similarities that could place them near any of the species groups characterised above. However, the statistical classification suggested some hypothesis for several of these species (Appendix 5). All ten observations of *Cichlidogyrus fontanai* and *C. rognoni* were classified in *Oreo1+* group. Nine out of ten observations of *C. levequei* were grouped with the *Oreo2+* group. All these species infect cichlids related to the host species typical to the respective group, i.e. oreochromine cichlids. Therefore, a close relationship to these species seems possible but requires further studies and samples. *Cichlidogyrus albareti* and *C. arfii* are the only monogenean parasites specific to gobiocichline, i.e. *"Tilapia" brevimanus*, and pelmatochromine cichlids, i.e. *Pelmatochromis buettikoferi* (Pariselle and Euzet 1995b, 1998), respectively. The SVMs classified all eight observations of *C. albareti* in the *EAR* group. All ten observations of *Cichlidogyrus arfii*, a parasite of *Pelmatochromis buttikoferi*, were classified in the *Hemi+* clade. This result might relate to the close relationship of pelmatochromine and hemichromine cichlids (Schwarzer et al. 2009) and mirror the presence of species from another dactylogyridean lineage of parasites (*Onchobdella* Paperna, 1968) on host species of both tribes. *Cichlidogyrus bulbophallus*, *C. karibae*, *C. ranula*, *Cichlidogyrus lobus*, and *C. flagellum* share a bulbous portion of the penis and a similar geographical area (Geraerts et al. 2020) with species of the *Bulb* group and might be included if expanded to species with a long flagellate distal end of the penis. However, the classification provided by the SVMs suggest the *EAR* clade for all 14 observations of *C. bulbophallus* and *C. flagellum* unlike *C. ranula* (two observations) which was classified in the *CPO+3* group. No or few observations could be provided for *C. karibae* and *C. lobus* (Appendix 5). Classifications for *C. sanjeani* and *C. tilapiae* through SVMs revealed no coherent patterns as no more than 50%

of observations were placed in a single group. Furthermore, no predictions were made for *C. bifurcatus* due to missing morphometric measurements in the literature as some measurements were not standardised until after the release of respective publications (Paperna 1960, 1979).

DISCUSSION

In the present study, we investigated the evolution of the attachment and reproductive organs of the gill parasite species belonging to *Cichlidogyrus* (incl. the nested genus *Scutogyrus*). Species of *Cichlidogyrus* form a morphologically diverse lineage of parasitic flatworms and a proposed model system for parasite macroevolution. For the purpose of this study, we employed new approaches to parasite taxonomy including multivariate phylogenetic comparative methods (PCMs) and a statistical classification approach using support vector machines (SVMs). The study is the most extensive reconstruction of intra-generic evolutionary relationships of the monogenean genera *Cichlidogyrus* to date. We were able to identify and characterise 10 different species groups amongst the 137 species that are currently known (Fig. 2; Fig. 6). For some of these groups at least some member species had been hypothesised to be closely related to each other before, e.g. species within the *Tylo*, *Hemi+*, *CPO+3*, and '*Scutogyrus*' groups. However, this study is the first to summarise all these groups and to analyse the evolution of the haptor and reproductive morphology using sequence data of 56 species and morphometric data of 137 species. We were able to suggest and test a classification for all but 12 species of *Cichlidogyrus* and *Scutogyrus* in the proposed clades through a synthesis of morphological, molecular phylogenetic, and host range data. However, due to missing data, e.g. a limited coverage of the four gene regions and a limited taxon coverage with the majority of species undiscovered (Cruz-Laufer et al.

2020), the relationship between the proposed species groups remains unresolved with only moderate support for a few macroclades (Fig. 2).

Systematic Informativeness of Morphometric Measurements and Host Range Data

To identify and classify species of *Cichlidogyrus* (incl. *Scutogyrus*), publications have applied a specific set of measurements of the sclerotised attachment and reproductive organs. For the present study, we indentified the most systematically informative of the 29 measurements most frequently provided in in the literatures (Table 2). A cluster analysis revealed that a majority of the measurements (20) lack a detectable phylogenetic signal or present multicollinear relationships with each other (Table 2). Therefore, we used only a fraction of these measurement (9/29) for downstream analyses. This low systematic informativeness is also reflected by the performance of a classification analysis using support vector machines (SVMs). The trained SVM algorithm was not able to discern species groups based on the most informative proxy variables. Consequently, the accordance rate of the taxonomic literature survey and SVM-based approaches in this study never exceeded 88% for any of the proposed species groups. One cause for this moderate performance might be the underrepresentation of reproductive organ measurements in the training dataset. None of the measurements expressed a significant phylogenetic signal. This result contradicts a taxonomic literature survey, which showed that the phenotype of the male copulatory organ was key for characterising groups of related species present. In contrast to the SVM algorithm, this taxonomic approach allowed us to consider shape information of the reproductive organs available in taxonomic literature. Therefore, we could affiliate all but 12 species with the proposed species groups. This relative success highlights phylogenetic information of the reproductive organ structures similar to other dactylogyrid monogeneans, e.g. species of *Characidotrema* Paperna & Thurston, 1968 (Řehulková et al. 2019). For *Cichlidogyrus*, Rahmouni et al. (2017) noted that the heel of the male copulatory organ was

shown to be similarly shaped in possibly related species from Lake Tanganyika. Earlier studies have explained apparent lack of a phylogenetic signal in reproductive organ measurements by a fast evolutionary rate in the reproductive organs (Pouyaud et al. 2006). Conversely, we suggest that this observation might be linked to a limited taxon coverage in the study. Instead, the lack of a detectable phylogenetic could simply reflect a lack of informative measurements as the number of reproductive organ measurements is low compared to the attachment organ (5 vs. 24). The measurements that are currently employed, fail to incorporate the systematic information offered by attachment and reproductive organs.

The low predictive power of morphometrics has encouraged the use of discretisation for measurements of the sclerotised parts of the attachment organs (Vignon et al. 2011; Mendlová et al. 2012) as well as host range data (Mendlová and Šimková 2014) in previous evolutionary studies. However, discretisation of characters can introduce a researcher bias and produce misleading results. For instance, Mendlová et al. (2012) proposed that similar anchor shapes were an ancestral state of the binary character ‘shape of anchors’ (‘anchor similarity’ in the present study) based on a character map. Yet the phylogenetic informativeness of the ‘shape of anchors’ character was never investigated. In fact, the cluster analysis of the present study shows that dorsal anchor measurements and the respective ventral anchor measurements are highly multicollinear (Table 2; Appendix 3) with the exception of the inner root (DAd and VAd). Furthermore, the ventral bar shape was also introduced as variable for character mapping but no ancestral state could be inferred (Mendlová et al. 2012). Yet we found a strong phylogenetic signal (Table 2) likely because of the unique character state of species of *Scutogyrus*. These inconsistent results between continuous characters and their discretised counterparts underline that interpretations based on these discretised characters should be treated with caution. For instance, the hook

configuration character states introduced by Vignon et al. (2011) are based on the gaps detected between the hook measurements of different species (see Pariselle and Euzet 2009). New species discoveries in the last decade demonstrate that these hook character states might not be as clear-cut (see Rahmouni et al. 2017; Geraerts et al. 2020). Nonetheless, we detected a strong phylogenetic signal indicating that while an update to the coding of this character might be required, some character states might still prove useful for characterising species groups of *Cichlidogyrus* (and *Scutogyrus*), e.g. the monophyletic *CPO+3* group species with well-developed hook pairs 3–7. We propose that to use discretised morphological characters in the future, researchers should redefine character states through quantitative character coding techniques such as proposed by Garcia-Cruz and Sosa (2006). The resulting character states would be more independent from researcher bias, which could increase the repeatability of a study.

A discrete variable proposed to reflect host specificity has also led to doubtful conclusions. Mendlová and Šimková (2014) used the index of specificity (IS) to summarise the phylogenetic specificity of species of *Cichlidogyrus* (and *Scutogyrus*) into four categories including species, genus, and tribe-specific species as well as generalists. Their study concluded that host specificity is related to parental care of the cichlid hosts. However, this observation was made due to an inherent sampling bias towards hosts performing mouthbrooding (Cruz-Laufer et al. 2020), a specific type of parental care. Mendlová and Šimková (2014) also observed a correlation of the IS with the host phylogeny. However, this observation might also reflect a study bias in cichlid parasitology towards large economically relevant host species (Cruz-Laufer et al. 2020). Because of their importance to fisheries, these fish species unlike others have been introduced to various new habitats. These introductions might also increase the chances of co-introduction for their parasites and create more opportunities for host range expansions (e.g. Jorissen et al. 2020). To avoid these biases,

future studies should treat host specificity as a complex parameter encompassing structural, phylogenetic, and geographical aspects (Poulin et al. 2011). We suggest that these parameters could be included, e.g., in a genus-wide comparative study in the future.

In addition to discretisation, body size correction has been used to minimise the variability of metrics of species of *Cichlidogyrus* and to account for size-related effects. For instance, size correction through the second hook pair (U2) has been applied to the measurements of the sclerotised structures of attachment organs (see Pariselle and Euzet 2009; Messu Mandeng et al. 2015). However, the connection between the size of the organism and the size of U2 remains unverified. Therefore, using U2 to correct for size-related effects, might distort rather than filter the phylogenetic signal detected in other characters. We detected no phylogenetic signal in U2 (Table 2). This observation could indicate a possible independence of U2 from the phylogeny of the parasites but fails to provide statistical evidence for a size-dependent (isometric) scaling. Therefore, we recommend further studies to investigate the use of U2 and other measurements for size corrections.

In general, size of morphological structures should first be treated as continuous, unmodified characters before further processing to avoid losing systematic information. Discretisation and size correction should only be applied with sufficient statistical evidence to back up character coding as well as size correction techniques. We suggest increasing sampling and sequencing efforts to address data deficiency.

Attachment Organ Evolution: Allopatric Speciation and Stabilising Selection

To infer evolutionary patterns from the attachment organ morphology, we applied a penalised likelihood framework for multivariate phylogenetic comparative methods (PCMs)

to the character data. The evolutionary models representing different evolutionary hypotheses performed almost equally well (Fig. 2). Only the null model (H_0) was outperformed by every other model, which was expected as we removed all non-informative measurements from the dataset following the cluster analysis. We found no signficantly increased performance of the Ornstein-Uhlenbeck (OU), Early Burst (EB), and Late Burst (LB) models compared to the Brownian Motion (BM) model. Only the LB model performed slightly but significantly better than the other models (Fig. 3). As the OU, EB, and LB models present generalisations of the Brownian Motion (BM) model (Butler and King 2004; Harmon et al. 2010), we conclude that the character evolution of the attachment organs measurements follows a pattern associated with genetic drift. This pattern can also be observed in the attachment organ morphology as related species frequently share characteristic features (Fig. 3) similar to the related monogeneans belonging to *Ligophorus* (Rodríguez-González et al. 2017). The morphology of the attachment organs appears to be determined by phylogenetic constraints, i.e. the relatedness of the parasite species, as suggest by a previous study (Vignon et al. 2011).

In contrast, the taxonomic literature survey paints a more nuanced picture of the attachment organ morphology. Unlike the models implemented through PCMs, we detected patterns that would be expected under stabilising selection. For instance, the number of different phenotypes appears relatively restricted. The same morphological characteristics can be found in multiple species groups (Fig. 5a; Fig. 6) such as a well-developed outer root of the dorsal anchor (e.g. species of the *Bulb* and *Cop+* groups), large first hooks (e.g. species of *Tylo*, *EAR – Heel*, and *Hemi+* groups), and long marginal hooks (e.g. species of the *EAR – Heel*, *CPO+3* groups). Character shifts that have occured within species groups appear to be adaptations to a diverging host range or environment (see Messu Mandeng et al. 2015). Host-induced shifts of the attachment organ morphology have also been observed amongst other dactylogyrid monogeneans, e.g. species of *Kapentagyrus* (Kmentová et al. 2018) and

Thaparocleidus (Šimková et al. 2013), and parasite lineages, e.g. cymothoid isopods (Baillie et al. 2019). For *Cichlidogyrus*, host species infected by parasites from the same species group frequently belong to the same tribe or related tribes of cichlids (Fig. 2; Appendix 3). Consequently, the phylogenetic congruence amongst some clades of *Cichlidogyrus* and their hosts is strong (Vanhove et al. 2015). Examples for host-induced character shifts amongst species of *Cichlidogyrus* can be found across the genus:

- *Cichlidogyrus amieti* has undergone a recent character shift by losing the well-developed first hook pair that is characteristic for the other species of the *Hemi+* clade as a result of its host switch (Messu Mandeng et al. 2015).
- *Cichlidogyrus sclerosus*, a generalist infecting fishes belonging to various cichlid tribes (see Appendix 4), shares almost no attachment organ features with its sister species *C. amphoratus*, a specialist infecting only coptodonine cichlids (Fig. 1), beyond those shared by all species of the genus, e.g. auricles associated with the dorsal bar of the attachment organ (see Pariselle and Euzet 2009).
- The diverse morphological features of species of the *EAR* group might relate to the diversity of the hosts. The parasites, which we grouped into subgroups of the *EAR* clade, might have adapted to these different hosts resulting in divergent attachment and reproductive organ morphologies.

As for many parasites (Huyse et al. 2005), host-related and environmental effects might explain patterns associated with stabilising selection in species of *Cichlidogyrus*. Notably, the host ranges of the parasites offer clues for evolutionary processes frequently associated with host-induced effects such as co-divergence and host switching. For instance, co-divergence might be reflected in the host (Appendix 4) and geographical specificity of species of *Cichlidogyrus* (Vanhove et al. 2013) (e.g. *EAR* group species are exclusive to

Eastern Africa and *Bulb* group species to Southern Africa). Furthemore, a small yet increasing number of natural and invasion-induced host switches are reported in the literature from cichlids to non-cichlids, e.g. in the *Hemi+* clade (Birgi and Euzet 1983; Birgi and Lambert 1986), and from African cichlids to cichlids and non-cichlids in the Americas, the Levant, and Madagascar (*C. arthracanthus*, *C. sclerosus*, *C. tiberianus*, *C. tilapiae*, *C. thurstonae*, and *Scutogyrus longicornis*) (Jiménez-Garcia et al. 2001; Šimková et al. 2019; Paperna 1960). More host switches might be discovered in the future if more (non-cichlid) host species are studied.

Despite these host-induced character shifts, the evolutionary models employed for multivariate PCMs were not able to detect patterns of stabilising selection in the attachment organ measurements. We suggest that this observation might be a result of incomplete data, which have also affected similar studies in the past. The datasets used in previous studies were limited to single structures, i.e. the anchors (Rodríguez-González et al. 2017), or to the species described at the time, which comprised less phenotypical diversity than is currently known to exist (Vignon et al. 2011). The phylogenetic position of some species with features diverging from their respective species groups, e.g. *C. amieti* from other *Hemi+* group species, were not known. Therefore, the similarities of the attachment organ morphology in related species might reflect the host specificity of the parasites rather than only their relatedness. While the present analysis includes *C. amieti*, DNA sequence data of less than half of all described species are available at GenBank and many species are likely to be discovered in the future (Cruz-Laufer et al. 2020). This data deficiency likely affects the performance of the models implemented through PCMs. A more complete taxon coverage of morphological, molecular, and ecological data is needed to provide more insight into the evolutionary patterns of the genus. Furthermore, a closer look into the evolution of single species groups might reveal different patterns than for a genus-wide study. For instance, the

high variability in the attachment and reproductive organ morphology of the EAR group might indicate a late burst in the divergence of characters. A slightly better performance of the LB model in the phylogenetic comparative analysis provides reason to investigate the evolution of this species group further.

Reproductive Organ Evolution

Unlike the attachment organ, the reproductive organs show no character changes correlated to the host relatedness. Reproductive organ characters are evolutionarily stable within the different clades as evidenced by the role of the male copulatory organ in the morphological characterisation of the species groups. The character evolution of the reproductive organs, unlike the attachment organ, has no apparent connection to the host range and seems to be driven mostly by phylogenetic constraints. Differences within species groups seem to reflect the phylogenetic positions, e.g. the short penis of *C. falcifer* (*Hemi+*) and simple accessory piece of *C. falcifer* and *C. cubitus* (*CPO+3*) appear to be the ancestral state of the respective clade. Previous studies on the reproductive organs of monogenean flatworms have suggested that the complex surface of the reproductive organ might correlate with environmental factors, i.e. to avoid separation during copulation through the water flow (Kearn and Whittington 2015). This hypothesis would suggest that the sclerotised structures evolve under selective pressure. However, while the screw thread-like thickening of the penis wall for some species (*C. casuarinus, C. centesimus, C. nshomboi*) (Fig. 3, *Heel* subgroup) might present such an example, we see no general trend to confirm this hypothesis. Neither generalist species such as *C. halli, C. tilapiae,* and *C. sclerosus* nor species infecting potentially more motile hosts living in fast-flowing waters, e.g. *C. consobrini, C. maeander,* and *C. ranula* infecting species of *Orthochromis,* display more complex MCOs than their congeners. Consequently, the diversification of the reproductive organs might be explained

by reproductive isolation such as previously suggested for other dactylogyrid monogeneans (Šimková et al. 2002). A quantitative approach to measure reproductive organ shape could provide a more detailed answer to this question by capturing the shape variation currently only recorded as qualitative information in taxonomic literature (see below for proposals).

Classification, Prediction, and Research prospects

We demonstrated that learning algorithms can be a practical tool to quantify the predictive power of morphometric measurements. We were able to emphasise the existing yet limited predictive power of the measurements currently employed by most taxonomists. Notably, some species groups of *Cichlidogyrus* such as the '*Scutogyrus*' and *EAR* groups were classified with high confidence. However, to improve the accuracy of these algorithms overall, datasets with a more complete taxon coverage and number of observations are needed. Results of taxonomic literature survey and statistical classification approaches have highlighted a discrepancy between the systematic informativeness of the attachment and reproductive organ morphology in taxonomic literature, and the low number of phylogenetically informative measurements. Future studies must address this discrepancy by increasing the number of systematically informative metrics.

Several possible solutions have been proposed to increase the number of metrics. First, some studies included novel measures such as the lengths of the shaft of the hooks, the curved length of the accessory piece, or the width of the penis tube and the bulbous portion of the penis (e.g. Geraerts et al. 2020) to describe species of *Cichlidogyrus* in more detail. Second, geomorphometric analyses have been employed for the anchors of some species of *Cichlidogyrus* (Rahmouni et al. 2020) as well as other monogeneans (Vignon and Sasal 2010; Rodríguez-González et al. 2017; Kmentová et al. 2020) to infer the most systematically informative structures. Third, 3D imaging techniques as previously developed

for monogenean sclerotised structures (e.g. García-Vásquez et al. 2012) have been proposed to improve characterisations of species of *Cichlidogyrus* (Cruz-Laufer et al. 2020) using methods employed for shape analyses (Klingenberg 2010; Dryden and Mardia 2016). Key characters that could be inferred using this approach are for example elongation (long vs round), torsion (straight vs coiled), and complexity (few vs many extensions). However, all of these solutions remain currently unsuitable for comparative analyses. Data resulting from the proposed additional measures, geomorphometric measurements, and 3D imaging techniques remain limited to few morphological structures, species, and studies. Future research should increase the taxon coverage for proposed methods and improve quantification methods to address these limitations. Future datasets should encompass an increasing number of DNA sequences and morphometric measurements from all known species. Furthermore, raw data should regularly be published in supplements or data repositories (e.g. Kmentová et al. 2016b) to improve the accessibility of morphometric data. This approach could enhance information flow and academic capacity for research into species of monogenean parasites in the future (Cruz-Laufer et al. 2020).

CONCLUSION

Due to the species-richness and the model system status of their cichlid hosts, the parasite lineage including species of *Cichlidogyrus* and *Scutogyrus* could be an ideal macroevolutionary model system for speciation in parasitic organisms in the future. However, the limited availability of phenotypic and genotypic data for the 137 species currently described has hampered investigations into the evolutionary history of the lineage. This study provides an example of how an increased availability of morphometric and sequence data can promote computational approaches to answer basic questions of evolutionary research in parasites. The phylogenetically informative measurements and

characterisation of multiple clades within *Cichlidogyrus* have highlighted key features and measurements required to describe new species as well as the shortcomings of the measurements that are currently used. These features of the attachment organ used to characterise the proposed species groups remain key descriptors alongside the features of the reproductive organs. However, these characters only serve as predictors in a multivariate context. This conclusion presents a stark contrast to previous classification systems, which used to predict species groups based on single discrete morphological characters.

Using multivariate phylogenetic comparative methods, we were able to model evolutionary mechanisms of the attachment organ of species of *Cichlidogyrus* that were suggested by previous studies. No support for alternative models suggests that the variation of the morphology of the attachment organ across the genus follows a pattern associated with genetic drift. However, morphological patterns in different species groups indicate a more nuanced pictures with the attachment organ morphology being constrained by host parameters. These constraints do not apply to the reproductive organs that appears to have a higher evolutionary stability. Other mechanisms might play a role in the evolution of this parasite lineage, e.g. within-host speciation as suggested by other studies. Therefore, more morphometric and DNA sequence data will be needed to provide a more detailed analysis of the evolution of the genus. We encourage researchers to publish morphometric raw data more regularly to improve the data accessibility for the scientific community, and to sequence systematically informative DNA regions of the remaining described or undiscovered species.

REFERENCES

Altman DG, Royston P. 2006. The cost of dichotomising continuous variables. BMJ. 332:1080. https://doi.org/10.1136/bmj.332.7549.1080.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Amundsen P-A, Lafferty KD, Knudsen R, Primicerio R, Klemetsen A, Kuris AM. 2009. Food web topology and parasites in the pelagic zone of a subarctic lake. J Anim Ecol. 78:563–572. https://doi.org/10.1111/j.1365-2656.2008.01518.x.

Baillie C, Welicky RL, Hadfield KA, Smit NJ, Mariani S, Beck RMD. 2019. Hooked on you: shape of attachment structures in cymothoid isopods reflects parasitic strategy. BMC Evol Biol. 19:207. https://doi.org/10.1186/s12862-019-1533-x.

Birgi E, Euzet L. 1983. Monogènes parasites des poissons des eaux douces du Cameroun: Présence des genres *Cichlidogyrus* et *Dactylogyrus* chez *Aphyosemion* (Cyprinodontidae). Bull Soc Zool Fr. 108:101–106.

Birgi E, Lambert A. 1986. Présence chez un Nandidae (Téléostéen), *Polycentropsis abbreviata* Boulenger, 1901, du genre *Cichlidogyrus* (Monogenea, Monopisthocotylea, Ancyrocephalidae). Ann Parasitol Hum Comp. 61:521–528. https://doi.org/10.1051/parasite/1986615521.

Bitja-Nyom AR, Agnèse J-F, Pariselle A, Bilong-Bilong CF, Gilles A, Snoeks J. 2021. A systematic revision of the five-spotted *Hemichromis* complex (Cichliformes: Cichlidae) from West Africa and Lower Guinea, with the description of a new species from Cameroon. Hydrobiologia. https://doi.org/10.1007/s10750-020-04506-5.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution. 57:717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x.

Butler MA, King AA. 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. Am Nat. 164:683–695. https://doi.org/10.1086/426002.

Campbell V, Legendre P, Lapointe F-J. 2011. The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. BMC Evol Biol. 11:64. https://doi.org/10.1186/1471-2148-11-64.

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

Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65:997–1008. https://doi.org/10.1093/sysbio/syw037.

Clavel J, Aristide L, Morlon H. 2019. A penalized likelihood framework for high-dimensional phylogenetic comparative methods and an application to New-World Monkeys brain evolution. Syst Biol. 68:93–116. https://doi.org/10.1093/sysbio/syy045.

Cruz-Laufer AJ, Artois T, Smeets K, Pariselle A, Vanhove MPM. 2020. The cichlid–*Cichlidogyrus* network: a blueprint for a model system of parasite evolution. Hydrobiologia. 365:1001. https://doi.org/10.1007/s10750-020-04426-4.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography. 36:27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Dossou C. 1982. Parasites de poissons d'eau douce du Bénin III: espèces nouvelles du genre *Cichlidogyrus* (Monogenea) parasites de Cichlidae. Bull Inst Fond Afr Noire Sér A Sci Nat. 44:295–322.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Dossou C, Birgi E. 1984. Monogènes parasites d'*Hemichromis fasciatus* Peters, 1857 (Teleostei, Cichlidae). Ann Sci Nat Zool. 6:101–109.

Douëllou L. 1993. Monogeneans of the genus *Cichlidogyrus* Paperna, 1960 (Dactylogyridae: Ancyrocephalinae) from cichlid fishes of Lake Kariba (Zimbabwe) with descriptions of five new species. Syst Parasitol. 25:159–186. https://doi.org/10.1007/BF00007007.

Dryden IL, Mardia KV. 2016. Statistical shape analysis: with applications in R. Second edition. Chichester, UK: John Wiley & Sons.

Dunne JA, Lafferty KD, Dobson AP, Hechinger RF, Kuris AM, Martinez ND, McLaughlin JP, Mouritsen KN, Poulin R, Reise K, et al. 2013. Parasites affect food web structure primarily through increased diversity and complexity. PLoS Biol. 11:e1001579. https://doi.org/10.1371/journal.pbio.1001579.

Dunz AR, Schliewen UK. 2013. Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as "*Tilapia*". Mol Phylogenet Evol. 68:64–80. https://doi.org/10.1016/j.ympev.2013.03.015.

Ergens R. 1981. Nine species of the genus *Cichlidogyrus* Paperna, 1960 (Monogenea: Ancyrocephalinae) from Egyptian fishes. Folia Parasitol. 28:205–214.

Fang Z, Fan J, Chen X, Chen Y. 2018. Beak identification of four dominant octopus species in the East China Sea based on traditional measurements and geometric morphometrics. Fish Sci. 84:975–985. https://doi.org/10.1007/s12562-018-1235-0.

Fannes W, Vanhove MPM, Huyse T. 2017. Redescription of *Cichlidogyrus tiberianus* Paperna, 1960 and *C. dossoui* Douëllou, 1993 (Monogenea: Ancyrocephalidae), with

special reference to the male copulatory organ. Syst Parasitol. 94:133–144. https://doi.org/10.1007/s11230-016-9685-1.

Fannes W, Vanhove MPM, Huyse T, Paladini G. 2015. A scanning electron microscope technique for studying the sclerites of *Cichlidogyrus*. Parasitol Res. 114:2031–2034. https://doi.org/10.1007/s00436-015-4446-7.

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am J Hum Genet. 25:471–492.

Firmat C, Alibert P, Mutin G, Losseau M, Pariselle A, Sasal P. 2016. A case of complete loss of gill parasites in the invasive cichlid *Oreochromis mossambicus*. Parasitol Res. 115:3657–3661. https://doi.org/10.1007/s00436-016-5168-1.

Garcia-Cruz J, Sosa V. 2006. Coding quantitative character data for phylogenetic analysis: a comparison of five methods. Syst Bot. 31:302–309. https://doi.org/10.1600/036364406777585739.

García-Vásquez A, Shinn AP, Bron JE. 2012. Development of a light microscopy stain for the sclerites of *Gyrodactylus* von Nordmann, 1832 (Monogenea) and related genera. Parasitol Res. 110:1639–1648. https://doi.org/10.1007/s00436-011-2675-y.

Geraerts M, Muterezi Bukinga F, Vanhove MPM, Pariselle A, Chocha Manda A, Vreven EJ, Huyse T, Artois T. 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). Parasit Vectors. 13:1–20. https://doi.org/10.1186/s13071-020-3927-4.

Gillardin C, Vanhove MPM, Pariselle A, Huyse T, Volckaert FAM. 2012. Ancyrocephalidae (Monogenea) of Lake Tanganyika: II: description of the first *Cichlidogyrus* spp. parasites

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM from Tropheini fish hosts (Teleostei, Cichlidae). Parasitol Res. 110:305–313. https://doi.org/10.1007/s00436-011-2490-5.

Goloboff PA, Mattoni CI, Quinteros AS. 2006. Continuous characters analyzed as such. Cladistics. 22:589–601. https://doi.org/10.1111/j.1096-0031.2006.00122.x.

Gómez A, Nichols E. 2013. Neglected wild life: parasitic biodiversity as a conservation target. Int J Parasitol Parasit Wildl. 2:222–227. https://doi.org/10.1016/j.ijppaw.2013.07.002.

Goolsby EW, Bruggeman J, Ané C. 2017. Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within species variation. Methods Ecol Evol. 8:22–27. https://doi.org/10.1111/2041-210X.12612.

Goren M. 2014. *Tristramella sacra*. The IUCN Red List of Threatened Species 2014:e.T61372A19010617. https://doi.org/10.2305/IUCN.UK.2014-1.RLTS.T61372A19010617.en.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59:307–321. https://doi.org/10.1093/sysbio/syq010.

Harmon A, Littlewood DTJ, Wood CL. 2019. Parasites lost: using natural history collections to track disease change across deep time. Front Ecol Environ. 17:157–166. https://doi.org/10.1002/fee.2017.

Harmon LJ, Losos JB, Jonathan Davies T, Gillespie RG, Gittleman JL, Bryan Jennings W, Kozak KH, McPeek MA, Moreno-Roark F, Near TJ, et al. 2010. Early bursts of body size

and shape evolution are rare in comparative data. Evolution. 64:2385–2396. https://doi.org/10.1111/j.1558-5646.2010.01025.x.

Hassouna N, Michot B, Bachellerie JP. 1984. The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. Nucleic Acids Res. 12:3563–3583. https://doi.org/10.1093/nar/12.8.3563.

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

Hoang DT, Le Vinh S, Flouri T, Stamatakis A, Haeseler A von, Minh BQ. 2018b. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol Biol. 18:11. https://doi.org/10.1186/s12862-018-1131-3.

Hsu C-W, Chang C-C, Lin C-J. 2003. A practical guide to support vector classification: Technical Report. Taipei, Taiwan: Department of Computer Science and Information Engineering, National Taiwan University. [cited 2016 May 19]. Available from: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

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

Igeh PC, Avenant-Oldewage A. 2020. Pathological effects of *Cichlidogyrus philander* Douëllou, 1993 (Monogenea, Ancyrocephalidae) on the gills of Pseudocrenilabrus philander (Weber, 1897) (Cichlidae). J Fish Dis. 43:177–184. https://doi.org/10.1111/jfd.13121.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Igeh PC, Dos Santos QM, Avenant-Oldewage A. 2017. Redescription of *Cichlidogyrus philander* (Monogenea, Ancyrocephalidae) using scanning electron microscopy (SEM) and molecular analysis. Parasite. 24:49. https://doi.org/10.1051/parasite/2017046.

Jiménez-Garcia MI, Vidal-Martínez VM, Lopez-Jiménez S. 2001. Monogeneans in introduced and native cichlids in México: Evidence for transfer. J Parasitol. 87:907. https://doi.org/10.2307/3285151.

Jorissen MWP, Huyse T, Pariselle A, Wamuini Lunkayilakio S, Muterezi Bukinga F, Chocha Manda A, Kapepula Kasembele G, Vreven EJ, Snoeks J, Decru E, et al. 2020. Historical museum collections help detect parasite species jumps after tilapia introductions in the Congo Basin. Biol Invasions. 11:1123. https://doi.org/10.1007/s10530-020-02288-4.

Jorissen MWP, Pariselle A, Huyse T, Vreven EJ, Snoeks J, Decru E, Kusters T, Lunkayilakio SW, Muterezi Bukinga F, Artois T, et al. 2018a. Six new dactylogyrid species (Platyhelminthes, Monogenea) from the gills of cichlids (Teleostei, Cichliformes) from the Lower Congo Basin. Parasite. 25:64. https://doi.org/10.1051/parasite/2018059.

Jorissen MWP, Pariselle A, Huyse T, Vreven EJ, Snoeks J, Volckaert FAM, Chocha Manda A, Kasembele GK, Artois T, Vanhove MPM. 2018b. Diversity and host specificity of monogenean gill parasites (Platyhelminthes) of cichlid fishes in the Bangweulu-Mweru ecoregion. J Helminthol. 92:417–437. https://doi.org/10.1017/S0022149X17000712.

Jorissen MWP, Vanhove MPM, Pariselle A, Snoeks J, Vreven EJ, Šimková A, Volckaert FAM, Wamuini Lunkayilakio S, Chocha Manda A, Kapepula Kasembele G, et al. in prep. Parasite gene flow along the Congo Basin: a phylogenetic approach.

Kalafi EY, Tan WB, Town C, Dhillon SK. 2016. Automated identification of monogeneans using digital image processing and K-nearest neighbour approaches. BMC Bioinform. 17:511. https://doi.org/10.1186/s12859-016-1376-z.

Kaltz O, Shykoff JA. 1998. Local adaptation in host–parasite systems. Heredity. 81:361–370. https://doi.org/10.1046/j.1365-2540.1998.00435.x.

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

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780. https://doi.org/10.1093/molbev/mst010.

Kearn G, Whittington I. 2015. Sperm transfer in monogenean (platyhelminth) parasites. Acta Parasitol;60:567–600. https://doi.org/10.1515/ap-2015-0082.

Klingenberg CP. 2010. Evolution and development of shape: integrating quantitative approaches. Nat Rev Genet. 11:623–635. https://doi.org/10.1038/nrg2829.

Kmentová N, Gelnar M, Koblmüller S, Vanhove MPM. 2016a. Deep-water parasite diversity in Lake Tanganyika: description of two new monogenean species from benthopelagic cichlid fishes. Parasit Vectors. 9:426. https://doi.org/10.1186/s13071-016-1696-x.

Kmentová N, Gelnar M, Koblmüller S, Vanhove MPM. 2016b. First insights into the diversity of gill monogeneans of '*Gnathochromis*' and *Limnochromis* (Teleostei, Cichlidae) in Burundi: do the parasites mirror host ecology and phylogenetic history? PeerJ. 4:e1629. https://doi.org/10.7717/peerj.1629.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Kmentová N, Gelnar M, Mendlová M, Van Steenberge M, Koblmüller S, Vanhove MPM. 2016c. Reduced host-specificity in a parasite infecting non-littoral Lake Tanganyika cichlids evidenced by intraspecific morphological and genetic diversity. Sci Rep. 6:39605. https://doi.org/10.1038/srep39605.

Kmentová N, Koblmüller S, Van Steenberge M, Artois T, Muterezi Bukinga F, Mulimbwa N'sibula T, Muzumani Risasi D, Masilya Mulungula P, Gelnar M, Vanhove MPM. 2020. Failure to diverge in African Great Lakes: The case of *Dolicirroplectanum lacustre* gen. nov. comb. nov. (Monogenea, Diplectanidae) infecting latid hosts. J Great Lakes Res. 46:1113–1130. https://doi.org/10.1016/j.jglr.2019.09.022.

Kmentová N, Van Steenberge M, Raeymaekers JAM, Koblmüller S, Hablützel PI, Muterezi Bukinga F, N’sibula TM, Mulungula PM, Nzigidahera B, Ntakimazi G, et al. 2018. Monogenean parasites of sardines in Lake Tanganyika: diversity, origin and intraspecific variability. Contrib Zool;87:105–132. https://doi.org/10.1163/18759866-08702004.

Konishi S. 1996. Generalised information criteria in model selection. Biometrika. 83:875–890. https://doi.org/10.1093/biomet/83.4.875.

Kuhn M. 2020. caret: Classification and Regression Training: R package version 6.0-86. [cited 2020 Mar 20]. Available from: https://CRAN.R-project.org/package=caret.

Legendre P, Lapointe F-J. 2004. Assessing congruence among distance matrices: single-malt scotch whiskies revisited. Aust N Z J Stat. 46:615–629. https://doi.org/10.1111/j.1467-842X.2004.00357.x.

Littlewood DTJ, Rohde K, Clough KA. 1997. Parasite speciation within or between host species?—Phylogenetic evidence from site-specific polystome monogeneans. Int J Parasitol. 27:1289–1297. https://doi.org/10.1016/S0020-7519(97)00086-6.

Llewellyn J. 1963. Larvae and larval development of monogeneans. In: Dawes B, editor. Advances in parasitology. Vol. 1. London, New York: Academic Press; p. 287–326 (vol. 1).

Lockyer AE, Olson PD, Ostergaard P, Rollinson D, Johnston DA, Attwood SW, Southgate VR, Horak P, Snyder SD, Le TH, et al. 2003. The phylogeny of the Schistosomatidae based on three genes with emphasis on the interrelationships of *Schistosoma* Weinland, 1858. Parasitology. 126:203–224. https://doi.org/10.1017/s0031182002002792.

Marchiori NC, Pariselle A, Pereira J, Agnèse J-F, Durand J-D, Vanhove MPM. 2015. A comparative study of *Ligophorus uruguayense* and *L. saladensis* (Monogenea: Ancyrocephalidae) from *Mugil liza* (Teleostei: Mugilidae) in southern Brazil. Folia Parasitol. 62. https://doi.org/10.14411/fp.2015.024.

Matejusová I, Gelnar M, McBeath AJA, Collins CM, Cunningham CO. 2001. Molecular markers for gyrodactylids (Gyrodactylidae: Monogenea) from five fish families (Teleostei). Int J Parasitol. 31:738–745. https://doi.org/10.1016/S0020-7519(01)00176-X.

McCoy KD. 2003. Sympatric speciation in parasites – what is sympatry? Trends Parasitol. 19:400–404. https://doi.org/10.1016/S1471-4922(03)00194-6.

McCoy KD. 2004. Response to Le Gac and Giraud: questioning is a good start. Trends Parasitol. 20:208–209. https://doi.org/10.1016/j.pt.2004.02.006.

McCoy MW, Bolker BM, Osenberg CW, Miner BG, Vonesh JR. 2006. Size correction: comparing morphological traits among populations and environments. Oecologia. 148:547–554. https://doi.org/10.1007/s00442-006-0403-6.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Mendlová M, Desdevises Y, Civáňová K, Pariselle A, Šimková A. 2012. Monogeneans of West African cichlid fish: evolution and cophylogenetic interactions. PloS One. 7:e37268. https://doi.org/10.1371/journal.pone.0037268.

Mendlová M, Pariselle A, Vyskočilová M, Šimková A. 2010. Molecular phylogeny of monogeneans parasitizing African freshwater Cichlidae inferred from LSU rDNA sequences. Parasitol Res. 107:1405–1413. https://doi.org/10.1007/s00436-010-2008-6.

Mendlová M, Šimková A. 2014. Evolution of host specificity in monogeneans parasitizing African cichlid fish. Parasit Vectors. 7:69. https://doi.org/10.1186/1756-3305-7-69.

Messu Mandeng FD, Bilong Bilong CF, Pariselle A, Vanhove MPM, Bitja Nyom AR, Agnèse J-F. 2015. A phylogeny of *Cichlidogyrus* spp. (Monogenea, Dactylogyridea) clarifies a host-switch between fish families and reveals an adaptive component to attachment organ morphology of this parasite genus. Parasit Vectors. 8:582. https://doi.org/10.1186/s13071-015-1181-y.

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. 2020. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien: R package version 1.7-4. [cited 2019 Oct 14]. Available from: https://cran.r-project.org/web/packages/e1071/e1071.pdf.

Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of Gateway Computing Environments Workshop (GCE), 14. Nov. 2010, New Orleans, LA. 2010 Gateway Computing Environments Workshop (GCE); 11/14/2010; New Orleans, LA, USA. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers (IEEE).

Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Methods Ecol Evol. 7:589–597. https://doi.org/10.1111/2041-210X.12526.

Mundry R. 2014. Statistical issues and assumptions of phylogenetic generalized least squares. In: Garamszegi LZ, editor. Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice. Berlin/Heidelberg, Germany: Springer; p. 131–153.

Muterezi Bukinga F, Vanhove MPM, Van Steenberge M, Pariselle A. 2012. Ancyrocephalidae (Monogenea) of Lake Tanganyika: III: *Cichlidogyrus* infecting the world's biggest cichlid and the non-endemic tribes Haplochromini, Oreochromini and Tylochromini (Teleostei, Cichlidae). Parasitol Res. 111:2049–2061. https://doi.org/10.1007/s00436-012-3052-1.

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

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884. https://doi.org/10.1038/44766.

Paperna I. 1960. Studies on monogenetic trematodes in Israel: 2. Monogenetic trematodes of cichlids. Bamidgeh. 12:20–33.

Paperna I. 1968. Monogenetic trematodes collected from fresh water fish in Ghana: Second Report. Bamidgeh. 20:88–99.

Paperna I. 1979. Monogenea of inland water fish in Africa. Tervuren, Belgium: Koninlijk Museum voor Midden-Afrika / Musée Royal de l'Afrique Centrale.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Paperna I, Thurston JP. 1969. Monogenetic trematodes collected from cichlid fish in Uganda; including the description of five new species of *Cichlidogyrus*. Rev Zool Bot Afr. 79:15–33.

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

Parins-Fukuchi C. 2018. Use of continuous traits can improve morphological phylogenetics. Syst Biol. 67:328–339. https://doi.org/10.1093/sysbio/syx072.

Pariselle A, Bitja Nyom AR, Bilong Bilong CF. 2013. Checklist of the ancyrocephalids (Monogenea) parasitizing *Tilapia* species in Cameroon, with the description of three new species. Zootaxa. 3599:78–86. https://doi.org/10.11646/zootaxa.3599.1.7.

Pariselle A, Bitja Nyom AR, Bilong Bilong CF. 2014. Four new species of *Cichlidogyrus* (Monogenea, Ancyrocephalidae) from *Sarotherodon mvogoi* and *Tylochromis sudanensis* (Teleostei, Cichlidae) in Cameroon. Zootaxa. 3881:258–266. https://doi.org/10.11646/zootaxa.3881.3.4.

Pariselle A, Euzet L. 1994. Three new species of *Cichlidogyrus* Paperna, 1960 (Monogenea, Ancyrocephalidae) parasitic on *Tylochromis jentinki* (Steindachner, 1895) (Pisces, Cichlidae) in West Africa. Syst Parasitol. 29:229–234. https://doi.org/10.1007/BF00009678.

Pariselle A, Euzet L. 1995a. *Scutogyrus* gen. n. (Monogenea: Ancyrocephalidae) for *Cichlidogyrus longicornis minus* Dossou, 1982, *C. l. longicornis*, and *C. l. gravivaginus* Paperna and Thurston, 1969, with description of three new species parasitic on African cichlids. J Helminthol Soc Wash. 62:157–173.

Pariselle A, Euzet L. 1995b. Trois monogènes nouveaux parasites branchiaux de *Pelmatochromis buettikoferi* (Steindachner, 1895) (Cichlidae) en Guinée. Parasite. 2:203–209. https://doi.org/10.1051/parasite/199502s2203.

Pariselle A, Euzet L. 1996. *Cichlidogyrus* Paperna, 1960 (Monogenea, Ancyrocephalidae): gill parasites from West African Cichlidae of the subgenus *Coptodon* Regan, 1920 (Pisces), with descriptions of six new species. Syst Parasitol. 34:109–124. https://doi.org/10.1007/BF00009685.

Pariselle A, Euzet L. 1998. Five new species of *Cichlidogyrus* (Monogenea: Ancyrocephalidae) from *Tilapia brevimanus*, *T. buttikoferi* and *T. cessiana* from Guinea, Ivory Coast and Sierra Leone (West Africa). Folia Parasitol. 45:275–282.

Pariselle A, Euzet L. 2003. Four new species of *Cichlidogyrus* (Monogenea: Ancyrocephalidae), gill parasites of *Tilapia cabrae* (Teleostei: Cichlidae), with discussion on relative length of haptoral sclerites. Folia Parasitol. 50:195–201. https://doi.org/10.14411/fp.2003.035.

Pariselle A, Euzet L. 2004. Two new species of *Cichlidogyrus* Paperna, 1960 (Monogenea, Ancyrocephalidae) gill parasites on *Hemichromis fasciatus* (Pisces, Cichlidae) in Africa, with remarks on parasite geographical distribution. Parasite. 11:359–364. https://doi.org/10.1051/parasite/2004114359.

Pariselle A, Euzet L. 2009. Systematic revision of dactylogyridean parasites (Monogenea) from cichlid fishes in Africa, the Levant and Madagascar. Zoosystema. 31:849–898. https://doi.org/10.5252/z2009n4a6.

Pariselle A, Morand S, Deveney MR, Pouyaud L. 2003. Parasite species richness of closely related hosts: historical scenario and "genetic" hypothesis. In: Combes C, Jourdane J, Ducreux-Modat A, Pages J-R, editors. Taxonomie, écologie et évolution des métazoaires

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM parasites: Taxonomy, ecology and evolution of metazoan parasites [Livre-hommage à Louis Euzet]. Vol. 2. Perpignan, France: Presses Universitaires de Perpignan; p. 147–166.

Pariselle A, Muterezi Bukinga F, Van Steenberge M, Vanhove MPM. 2015a. Ancyrocephalidae (Monogenea) of Lake Tanganyika: IV: *Cichlidogyrus* parasitizing species of Bathybatini (Teleostei, Cichlidae): reduced host-specificity in the deepwater realm? Hydrobiologia. 748:99–119. https://doi.org/10.1007/s10750-014-1975-5.

Pariselle A, Van Steenberge M, Snoeks J, Volckaert FAM, Huyse T, Vanhove MPM. 2015b. Ancyrocephalidae (Monogenea) of Lake Tanganyika: Does the *Cichlidogyrus* parasite fauna of *Interochromis loocki* (Teleostei, Cichlidae) reflect its host's phylogenetic affinities? Contrib Zool. 84:25–38.

Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 30:2216–2218. https://doi.org/10.1093/bioinformatics/btu181.

Poulin R. 2014. Parasite biodiversity revisited: frontiers and constraints. Int J Parasitol. 44:581–589. https://doi.org/10.1016/j.ijpara.2014.02.003.

Poulin R, Krasnov BR, Mouillot D. 2011. Host specificity in phylogenetic and geographic space. Trends Parasitol. 27:355–361. https://doi.org/10.1016/j.pt.2011.05.003.

Pouyaud L, Desmarais E, Deveney M, Pariselle A. 2006. Phylogenetic relationships among monogenean gill parasites (Dactylogyridea, Ancyrocephalidae) infesting tilapiine hosts (Cichlidae): systematic and evolutionary implications. Mol Phylogenet Evol. 38:241–249. https://doi.org/10.1016/j.ympev.2005.08.013.

Price CE, Kirk RG. 1967. First description of a monogenetic trematode from Malawi. Rev Zool Bot Afr. 76:137–144.

Price CE, Peebles HE, Bamford T. 1969. The monogenean parasites of African fishes. IV. Two new species from South African hosts. Rev Zool Bot Afr. 79:117–124.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing]. Available from: https://www.r-project.org/.

Rahmouni C, Van Steenberge M, Vanhove MPM, Šimková A. 2020. Intraspecific morphological variation in *Cichlidogyrus* (Monogenea) parasitizing two cichlid hosts from Lake Tanganyika exhibiting different dispersal capacities. Hydrobiologia:in press. https://doi.org/10.1007/s10750-020-04429-1.

Rahmouni C, Vanhove MPM, Šimková A. 2017. Underexplored diversity of gill monogeneans in cichlids from Lake Tanganyika: eight new species of *Cichlidogyrus* Paperna, 1960 (Monogenea: Dactylogyridae) from the northern basin of the lake, with remarks on the vagina and the heel of the male copulatory organ. Parasit Vectors. 10:591. https://doi.org/10.1186/s13071-017-2460-6.

Rahmouni C, Vanhove MPM, Šimková A. 2018. Seven new species of *Cichlidogyrus* Paperna, 1960 (Monogenea: Dactylogyridae) parasitizing the gills of Congolese cichlids from northern Lake Tanganyika. PeerJ. 6:e5604. https://doi.org/10.7717/peerj.5604.

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

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Řehulková E, Kičinjaová ML, Mahmoud ZN, Gelnar M, Seifertová M. 2019. Species of *Characidotrema* Paperna & Thurston, 1968 (Monogenea: Dactylogyridae) from fishes of the Alestidae (Characiformes) in Africa: new species, host-parasite associations and first insights into the phylogeny of the genus. Parasit Vectors. 12:366. https://doi.org/10.1186/s13071-019-3580-y.

Řehulková E, Mendlová M, Šimková A. 2013. Two new species of *Cichlidogyrus* (Monogenea: Dactylogyridae) parasitizing the gills of African cichlid fishes (Perciformes) from Senegal: morphometric and molecular characterization. Parasitol Res. 112:1399–1410. https://doi.org/10.1007/s00436-013-3291-9.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Rodríguez-González A, Míguez-Lozano R, Sarabeev V, Balbuena JA. 2016. Evolutionary modularity and morphological integration in the haptoral anchor structures of *Ligophorus* spp. (Monogenea: Dactylogyridae). Parasitol Res. 115:3519–3533. https://doi.org/10.1007/s00436-016-5117-z.

Rodríguez-González A, Sarabeev V, Balbuena JA. 2017. Evolutionary morphology in shape and size of haptoral anchors in 14 *Ligophorus* spp. (Monogenea: Dactylogyridae). PloS One. 12:e0178367. https://doi.org/10.1371/journal.pone.0178367.

Ronco F, Büscher HH, Indermaur A, Salzburger W. 2020. The taxonomic diversity of the cichlid fish fauna of ancient Lake Tanganyika, East Africa. J Great Lakes Res. 46:1067–1078. https://doi.org/10.1016/j.jglr.2019.05.009.

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

Salzburger W. 2018. Understanding explosive diversification through cichlid fish genomics. Nat Rev Genet. 19:705–717. https://doi.org/10.1038/s41576-018-0043-9.

Salzburger W, van Bocxlaer B, Cohen AS. 2014. Ecology and evolution of the African Great Lakes and their faunas. Annu Rev Ecol Evol Syst. 45:519–545. https://doi.org/10.1146/annurev-ecolsys-120213-091804.

Schedel FDB, Musilova Z, Schliewen UK. 2019. East African cichlid lineages (Teleostei: Cichlidae) might be older than their ancient host lakes: new divergence estimates for the east African cichlid radiation. BMC Evol Biol. 19:94. https://doi.org/10.1186/s12862-019-1417-0.

Schwarzer J, Misof B, Tautz D, Schliewen UK. 2009. The root of the East African cichlid radiations. BMC Evol Biol. 9:186. https://doi.org/10.1186/1471-2148-9-186.

Šimková A, Matějusová I, Cunningham CO. 2006. A molecular phylogeny of the Dactylogyridae *sensu* Kritsky & Boeger (1989) (Monogenea) based on the D1-D3 domains of large subunit rDNA. Parasitology. 133:43–53. https://doi.org/10.1017/S0031182006009942.

Šimková A, Ondračková M, Gelnar M, Morand S. 2002. Morphology and coexistence of congeneric ectoparasite species: reinforcement of reproductive isolation? Biol J Linn Soc. 76:125–135. https://doi.org/10.1111/j.1095-8312.2002.tb01719.x.

Šimková A, Plaisance L, Matějusová I, Morand S, Verneau O. 2003. Phylogenetic relationships of the Dactylogyridae Bychowsky, 1933 (Monogenea: Dactylogyridea): the

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM need for the systematic revision of the Ancyrocephalinae Bychowsky, 1937. Syst Parasitol. 54:1–11. https://doi.org/10.1023/A:1022133608662.

Šimková A, Řehulková E, Rasoloariniaina JR, Jorissen MWP, Scholz T, Faltýnková A, Mašová Š, Vanhove MPM. 2019. Transmission of parasites from introduced tilapias: a new threat to endemic Malagasy ichthyofauna. Biol Invasions. 21:803–819. https://doi.org/10.1007/s10530-018-1859-0.

Šimková A, Serbielle C, Pariselle A, Vanhove MPM, Morand S. 2013. Speciation in *Thaparocleidus* (Monogenea: Dactylogyridae) parasitizing Asian pangasiid catfishes. BioMed Res Int. 2013:353956. https://doi.org/10.1155/2013/353956.

Swofford DL. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sunderland, MA, USA: Sinauer Associates.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56:564–577. https://doi.org/10.1080/10635150701472164.

Thieme ML, Abell R, Stiassny MLJ, Skelton P, Lehner B, Teugels GG, Dinerstein E, Toham AK, Burgess N, Olson D. 2005. Freshwater ecoregions of Africa and Madagascar: A conservation assessment. Washington, USA: Island Press.

Thompson RCA. 2013. Parasite zoonoses and wildlife: One Health, spillover and human activity. Int J Parasitol. 43:1079–1088. https://doi.org/10.1016/j.ijpara.2013.06.007.

Thurston JP. 1970. The incindence of Monogenea and parasitic Crustacea on the gills of fish in Uganda. Rev Zool Bot Afr. 82:111–130.

Toju H, Yamamichi M, Guimarães PR, Olesen JM, Mougi A, Yoshida T, Thompson JN. 2017. Species-rich networks and eco-evolutionary synthesis at the metacommunity level. Nat Ecol Evol. 1:24. https://doi.org/10.1038/s41559-016-0024.

Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL. 2001. How many species of cichlid fishes are there in African lakes? Mol Ecol. 10:793–806. https://doi.org/10.1046/j.1365-294x.2001.01200.x.

Van Steenberge M, Pariselle A, Huyse T, Volckaert FAM, Snoeks J, Vanhove MPM. 2015. Morphology, molecules, and monogenean parasites: an example of an integrative approach to cichlid biodiversity. PloS One. 10:e0124474. https://doi.org/10.1371/journal.pone.0124474.

Van Steenberge M, Vanhove MPM, Muzumani Risasi D, Mulimbwa N'Sibula T, Muterezi Bukinga F, Pariselle A, Gillardin C, Vreven E, Raeymaekers JAM, Huyse T, et al. 2011. A recent inventory of the fishes of the north-western and central western coast of Lake Tanganyika (Democratic Republic Congo). Acta Ichthyol Piscat. 41:201–214. https://doi.org/10.3750/AIP2011.41.3.08.

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

Vanhove MPM, Hablützel PI, Pariselle A, Šimková A, Huyse T, Raeymaekers JAM. 2016. Cichlids: a host of opportunities for evolutionary parasitology. Trends Parasitol. 32:820–832. https://doi.org/10.1016/j.pt.2016.07.002.

Vanhove MPM, Pariselle A, Van Steenberge M, Raeymaekers JAM, Hablützel PI, Gillardin C, Hellemans B, Breman FC, Koblmüller S, Sturmbauer C, et al. 2015. Hidden

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM biodiversity in an ancient lake: phylogenetic congruence between Lake Tanganyika tropheine cichlids and their monogenean flatworm parasites. Sci Rep. 5:13669. https://doi.org/10.1038/srep13669.

Vanhove MPM, Van Steenberge M, Dessein S, Volckaert FAM, Snoeks J, Huyse T, Pariselle A. 2013. Biogeographical implications of Zambezian *Cichlidogyrus* species (Platyhelminthes: Monogenea: Ancyrocephalidae) parasitizing Congolian cichlids. Zootaxa. 3608:398–400. https://doi.org/10.11646/zootaxa.3608.5.8.

Vanhove MPM, Volckaert FAM, Pariselle A. 2011. Ancyrocephalidae (Monogenea) of Lake Tanganyika: I: Four new species of *Cichlidogyrus* from *Ophthalmotilapia ventralis* (Teleostei: Cichlidae), the first record of this parasite family in the basin. Zoologia. 28:253–263. https://doi.org/10.1590/S1984-46702011000200016.

Vignon M, Pariselle A, Vanhove MPM. 2011. Modularity in attachment organs of African *Cichlidogyrus* (Platyhelminthes: Monogenea: Ancyrocephalidae) reflects phylogeny rather than host specificity or geographic distribution. Biol J Linn Soc. 102:694–706. https://doi.org/10.1111/j.1095-8312.2010.01607.x.

Vignon M, Sasal P. 2010. The use of geometric morphometrics in understanding shape variability of sclerotized haptoral structures of monogeneans (Platyhelminthes) with insights into biogeographic variability. Parasitol Int. 59:183–191. https://doi.org/10.1016/j.parint.2010.01.006.

Wäldchen J, Mäder P. 2018. Machine learning for image based species identification. Methods Ecol Evol. 9:2216–2225. https://doi.org/10.1111/2041-210X.13075.

Webster JP, Gower CM, Knowles SCL, Molyneux DH, Fenton A. 2016. One health - an ecological and evolutionary framework for tackling Neglected Zoonotic Diseases. Evol Appl. 9:313–333. https://doi.org/10.1111/eva.12341.

Wickham H. 2016. ggplot2: Elegant graphics for data analysis. 2nd ed. Switzerland: Springer.

Wu XY, Zhu XQ, Xie MQ, Li AX. 2006. The radiation of *Haliotrema* (Monogenea: Dactylogyridae: Ancyrocephalinae): molecular evidence and explanation inferred from LSU rDNA sequences. Parasitology. 132:659–668. https://doi.org/10.1017/S003118200500956X.

Wu XY, Zhu XQ, Xie MQ, Li AX. 2007. The evaluation for generic-level monophyly of Ancyrocephalinae (Monogenea, Dactylogyridae) using ribosomal DNA sequence data. Mol Phylogenet Evol. 44:530–544. https://doi.org/10.1016/j.ympev.2007.03.025.

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

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

Zhi T, Xu X, Chen J, Zheng Y, Zhang S, Peng J, Brown CL, Yang T. 2018. Expression of immune-related genes of Nile tilapia *Oreochromis niloticus* after *Gyrodactylus cichlidarum* and *Cichlidogyrus sclerosus* infections demonstrating immunosupression in coinfection. Fish Shellfish Immunol. 80:397–404. https://doi.org/10.1016/j.fsi.2018.05.060.

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

Zischke MT, Litherland L, Tilyard BR, Stratford NJ, Jones EL, Wang Y-G. 2016. Otolith morphology of four mackerel species (*Scomberomorus* spp.) in Australia: Species differentiation and prediction for fisheries monitoring and assessment. Fish Res. 176:39–47. https://doi.org/10.1016/j.fishres.2015.12.003.

TABLES

Table 1. Substitution models of molecular evolution and partitions for Bayesian inference (BI) and maximum likelihood estimation (ML) of phylogeny of species of *Cichlidogyrus* and *Scutogyrus*. Models include the general time reversible model (GTR), the Kimura 1980 model (K80), the transitional model 3 with unequal base frequencies (TIM3e), the Tamura-Nei model (TN), and the three-parameter model 2 (TPM2) plus empirical base frequencies (+ F), a proportion of invariable sites (+ I), a discrete Γ model with four rate categories (Γ4), or a FreeRate model with three categories (+ R3). For model specification see IQ-TREE ModelFinder manual (Kalyaanamoorthy et al. 2017).

Partition	**Bayesian inference (BI)**	**Maximum likelihood estimation (ML)**
28S rDNA	GTR + F + I + Γ4	TIM3e + F + R3
18S rDNA	K80 + I + Γ4	K80 + I + Γ4
ITS rDNA	GTR + F + Γ4	TPM2 + F + Γ4
CO1 mtDNA	GTR + F + I + Γ4	TN + F + I + Γ4

Table 2. Clusters of multicollinear measures (A–E), and presence and strength of phylogenetic signals in continuous morphometric measures of species of *Cichlidogyrus* and *Scutogyrus* used to select phylogenetically informative variables (bold) for multivariate phylogenetic comparative analyses. Clusters were detected through Pearson pairwise correlation matrix (|r| > 0.7). Phylogenetic signals were detected through difference in fit of models of character evolution and a null model (H_0) assuming a total polytomy (Pagel's λ = 0). Model fit was assessed through the mean difference of the sample size–corrected Akaike information criterion (ΔAICc) and the standard deviation of the AICc (SD).

Organ	**Character**	**Cluster**	**Brownian motion (H_1)**			**Ornstein-Uhlenbeck (H_2)**			**Early Burst: (H_3)**		
			Mean	SD	p	Mean	SD	p	Mean	SD	p
Attachment organ	DA a	A	-4.2	2.1		-4.6	1.3	*	-1.1	2.6	
	DA b	A	-5.9	2.6		-6.0	1.5	*	-2.6	3.2	
	VA a	A	-0.5	0.5		1.3	0.8		2.0	0.5	
	VA b	A	-0.1	0.2		1.5	1.0		2.2	0.2	
	DA c	B	-8.0	3.3		-8.3	2.7	**	-4.0	4.3	
	VA c	B	-4.7	2.2		-4.4	1.7	*	-1.5	2.7	
	DA d		-0.4	0.7		0.8	1.6		2.1	0.6	
	VA d		-1.0	0.6		1.2	0.5		0.9	0.9	
	DA e	C	-1.0	0.7		0.1	1.3		1.8	0.8	
	VA e	C	-7.0	2.1		-5.7	1.3	*	-3.0	3.5	
	U 1		-23.0	4.4	*	-22.3	3.1	*	-17.7	7.9	*
	U 2		0.6	0.0		1.2	10.3		2.9	0.0	
	U 3	D	-28.6	7.0	**	-29.8	2.6	*	-24.5	8.7	*
	U 4	D	-28.0	5.9	***	-28.8	2.5	*	-23.3	9.2	

	U 5	D	-23.8	5.1	**	-23.9	2.1	*	-19.3	8.4	
	U 6	D	-24.6	6.6	**	-25.9	2.5	*	-21.9	7.0	*
	U 7	D	-21.9	5.2	**	-22.5	2.1	*	-17.5	7.7	
	DB h		-57.9	3.1	***	-57.0	3.4	*	-55.0	4.1	*
	DB x	E	0.0	0.0		2.2	0.4		2.3	0.0	
	DB y	E	-0.6	1.2		-0.1	2.1		1.8	1.2	
	DB w		-5.9	1.5	**	-5.1	1.7	*	-2.5	2.7	
	VB x		-2.3	1.2		-1.1	0.8	*	0.6	1.5	
	VB w		-6.8	2.0		-5.2	0.8	*	-3.4	3.3	
Reproductive organs	Pe		7.4	2.0		20.8	24.3		10.4	2.2	
	AP		7.4	0.2		8.9	6.2		9.8	0.1	
	He		-3.1	2.5		-2.0	6.6		-0.4	2.8	
	AuP		-6.1	2.8		-5.7	7.5		-1.5	3.7	
	Vg l		3.4	1.4		6.3	9.8		7.4	2.2	
	Vg w		2.8	0.1		12.2	21.8		5.1	0.1	

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. Abbreviations: DA, dorsal anchor; VA, ventral anchor; U, hooks; DB, dorsal bar; VB, ventral bar; Vg, vagina; AuP, surface of auxiliary plate of male copulatory organ; a, anchor total length; b, anchor blade length; c, anchor shaft length; d, anchor guard length; e, anchor point length; 1–7, hook lengths; w, bar width; x, bar length; h, auricle length; y, distance between auricles; AP, accessory piece length; Pe, penis length; He, heel length; l, length; w, width. Terminology and methodology of measurements according to Fannes et al. (2017).

Table 3. Confusion matrix of statistical classification of specimens belonging to species of *Cichlidogyrus* without available DNA sequence data into proposed species groups (taxonomic classification) based on proxy variables of attachment organ morphology (Table

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM

2). Species groups were proposed based on phylogenetic analysis (Fig. 2) and taxonomic literature survey.

Species group	**Statistical classification**										Accordance rate
Taxonomic classification	*Bulb*	*Cop+*	*CPO+3*	*EAR*	*Halli*	*Hemi+*	*Oreo2+*	*Oreo1+*	*Scutogyrus*	*Tylo*	
Bulb	0	0	0	4	0	0	0	0	0	0	0%
Cop+	0	16	0	13	0	0	0	0	0	0	55%
CPO+3	0	2	106	31	0	0	3	6	0	0	72%
EAR	0	0	15	169	0	0	6	0	0	1	88%
Hemi+	0	0	0	8	0	6	0	0	0	0	43%
Oreo1+	0	0	0	2	0	0	10	28	0	0	70%
Scutogyrus	0	0	0	0	0	0	0	0	20	0	100%
Tylo	0	1	3	44	0	0	0	0	0	16	25%

FIGURES

Figure 1. Overview of morphology of species of *Cichlidogyrus* Paperna, 1960 (Platyhelminthes: Monongenea, Dactylogyridae). A, Multiple specimens attached to the gills of *Sarotherodon melanotheron* Rüppel, 1852 (Cichliformes: Cichlidae). B, Microscopic image of *Cichlidogyrus agnesi* Pariselle & Euzet, 1995 with sclerotised structures of reproductive (male copulatory organ and vagina) and attachment organs indicated by arrows. C, Overview of hard part morphology and most widely applied measurements of these structures in taxonomic literature. Abbreviations: a, anchor total length; b, anchor blade length; c, anchor shaft length; d, anchor guard length; e, anchor point length; 1–7, hook lengths; w, bar width; x, bar length; h, auricle length; y, distance between auricles; AuP, surface of auxiliary plate of male copulatory organ; AP, accessory piece length; Pe, penis length; He, heel length; l, length; w, width. Terminology and methodology of measurements according to Fannes et al. (2017).

Figure 2. Midpoint-rooted Bayesian inference phylogram of monogenean flatworms belonging to *Cichlidogyrus* and *Scutogyrus* (Platyhelminthes, Monogenea, Dactylogyridae) and host range matrix of the species by tribe or subfamily of cichlid hosts and family of non-cichlid hosts (recent anthropogenic host range expansions italicised). The phylogram was inferred from three nuclear (18S, 28S, and ITS rDNA) and one mitochondrial (CO1 mtDNA) DNA sequence markers. Sequences of specimens in bold generated for the present study. Node support values: Bayesian posterior probabilities (PP) followed by ultrafast bootstrap values (UBoot) and Shimodaira-Hasegawa-like approximate likelihood ratios (SH-aLRT) inferred from maximum likelihood estimation and ultrafast bootstrap values inferred from maximum parsimony estimation (MPBoot), asterisk (*) indicates low or moderate support below the threshold (PP < 0.95, UBoot < 95, SH-aLRT < 80, MPBoot < 95); black dots, internal nodes with strong support across all support values; highlighted clades, monophyletic

Cruz-Laufer AJ, Pariselle A, Jorissen MWP, Muterezi Bukinga F, Al Assadi A, Van Steeberge M, Koblmüller S, Sturmbauer C, Huyse T, Smeets K, Artois T, Vanhove MPM clades considered strongly supported by at least three node support values: species infecting (mostly) hemichromine cichlids (*Hemi+*), species belonging to *Scutogyrus* (*Scutogyrus*), species infecting (mostly) coptodonine cichilids among others (*Cop+*), the first species group infecting oreochromine cichlids among others (*Oreo1+*), species infecting cichlids belonging to the East African Radiation (*EAR*), species infecting coptodonine, pelmatolapiinem oreochromine, tilapiine, heterotilapiine, and gobiocichline cichlids (*CPO+3*), species from Southern Africa with a bulbous penis (*Bulb*), the second species group infecting mainly oreochromine cichlids among others (*Oreo2+*), *C. halli*–clade (*Halli*), and species infecting tylochromine cichlids (*Tylo*). Abbreviations: *C. schreyenbrichard.*, *C. schreyenbrichardorum;* APLO, Aplocheilidae; NOTH, Nothobranchiidae; POLY, Polycentridae; Pare, Paretroplinae; Ptyc, Ptychochrominae; Cich, Cichlasomatini; Chro, Chromidotilapiini; Tylo, Tylochromini; Hemi, Hemichromini; Gobi, Gobiocichlini; Copt, Coptodonini; Hete, Heterotilapiini, PelL, Pelmatolapiini; Oreo, Oreochromini; Tila, Tilapiini; Boul, Boulengerochromini; Bath, Bathybatini; Trem, Trematocarini; Bent, Benthochromini; Cypr, Cyprichromini; Ecto, Ectodini; Hapl, Haplochromini.

Figure 3. Results of multivariate phylogenetic comparative analysis of proxy variables of attachment organ (Table 2) applied to 100 random samples of tree topologies from Bayesian inference. Fitted models include Brownian Motion (BM), Ornstein-Uhlenbeck (OU), Early-Burst (EB), and Late-Burst (LB) models simulating genetic drift (H_1), stabilising selection (H_2), and adaptive radiation with a decelerating (H_3) and accelerating (H_4) divergence of characters. Model performance is assessed through difference in the Generalised Information Criterion (ΔGIC) compared to a null model (H_0) assuming a total polytomy (Pagel's $\lambda = 0$). All models performed equally well, which gives preference to BM as OU, EB, and LB are generalisations of BM (Blomberg et al. 2003; Butler and King 2004; Harmon et al. 2010).

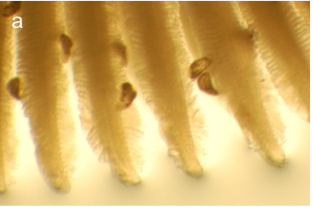
Figure 4. Results of univariate phylogenetic comparative analysis of discrete morphological and environmental characters of *Cichlidogyrus* and *Scutogyrus* under Pagel's λ model using 100 random samples of tree topologies inferred from Bayesian inference. Discrete characters include the hook configuration (HC) (Vignon et al. 2011), anchor similarity (AS), ventral bar shape (VBS) (Mendlová et al. 2012), and index of (host) specificity (IS) (Mendlová and Šimková 2014). Model performance is assessed through difference in the sample size–corrected Akaike Information Criterion (ΔAICc) compared to a null model (H_0) assuming total phylogenetic independence (Pagel's $\lambda = 0$). HC and VBS show a strong phylogenetic signal, AS shows a weak phylogenetic signal, and IS shows no detectable phylogenetic signal.

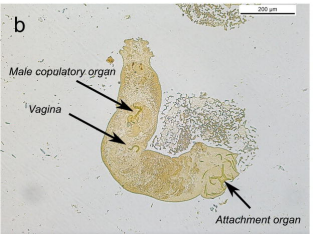
Figure 5. Character maps of discrete characters of haptor configuration (A) and index of specificity (B) of species of *Cichlidogyrus* based on subset of phylogenetic tree in Figure 1 with only one specimen per species. Numbers on nodes refer to clades in Figure 2. Haptor configuration A, B, and C assigned according to criteria by Vignon et al. (2011) with relative size of hook pair 1 vs pairs 3–7; haptor configuration D equates to the configuration of *C. arthracanthus* with large hooks 1 and 3–7. Index of specificity assigned as species, genus, tribe-specific, and generalist according to Mendlová and Šimková (2014).

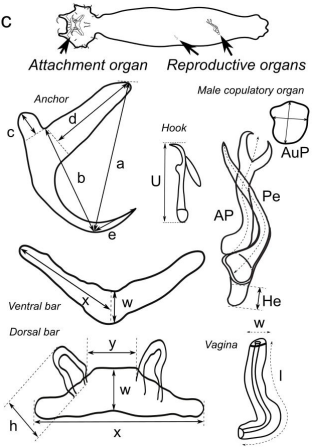
Figure 6. Morphology of attachment and reproductive organs of representative species belonging to proposed species groups and subgroups (Fig. 2) of the genus *Cichlidogyrus* (incl. *Scutogyrus*) including cladogram and illustrations of sclerotised structures of attachment organ (anchors, bars, hooks) and the MCO of selected species (vagina morphology not shown as highly variable between species, see "Characterisation of species groups"). Arrows indicate key features of each species group (blue) and subgroup of the *EAR* group (yellow) (see Appendix 4), dashed lines indicate typical shapes. Species of the *EAR* group that are not displayed are labelled in grey. Scale = 30 µm.

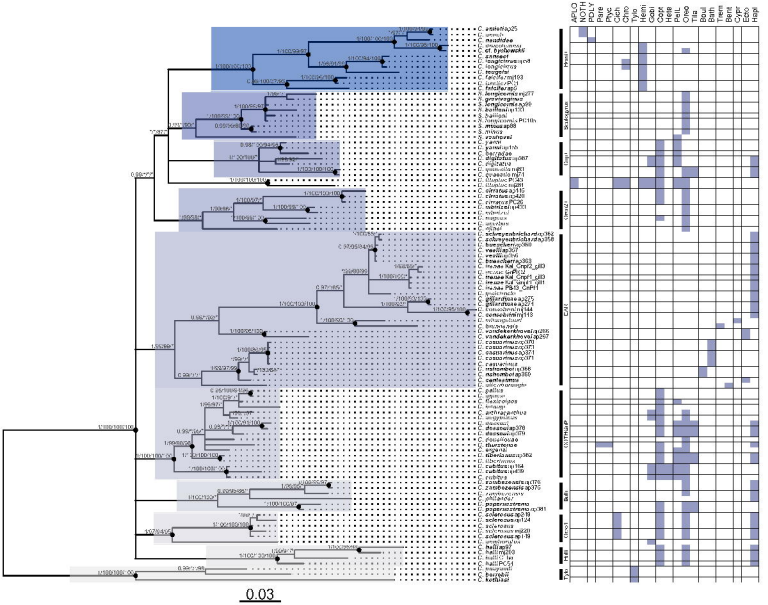
Figure 7. Scatterplots of specimen data (in µm) used to illustrate two-way combinations of morphometric measurements best suited to characterise species groups within *Cichlidogyrus* (incl. *Scutogyrus*), Species groups are highlighted with colours. A. The *Tylo* group is characterised by relatively short auricles fused to the dorsal (DB h), some species display a well-developed first hook pair (U1) similar to species of the *Bulb* group. B. The *Oreo1+* group is characterised by a short inner root of the dorsal bar (DAc) and short hook pairs 3–7 as opposed to the *Scutogyrus* group. C. The *CPO+3* and *Scutogyrus* groups are characterised by well-developed hook pairs 3–7 with the exceptions of some species, the *Cop+* and *Hemi+* groups are characterised by a well-developed first hook pair (U1). D. The *Scutogyrus* and *Halli* groups are characterised by well-developed hook pairs 3–7 and large auricles at the dorsal bar (DB h).

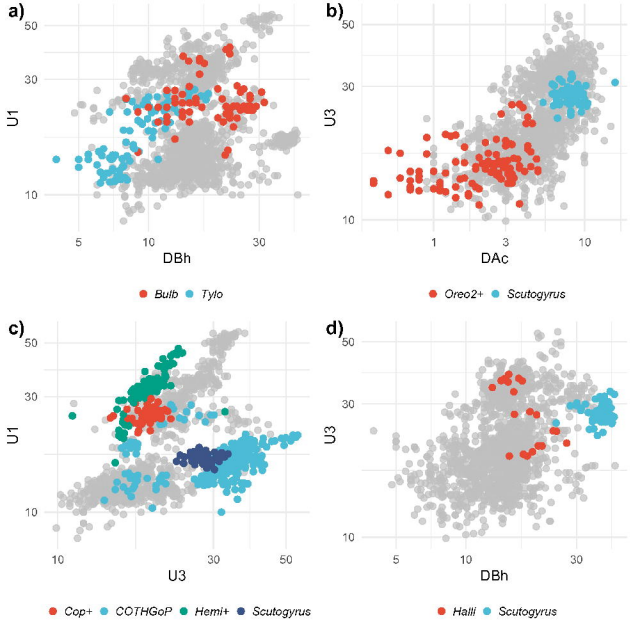
Figure 8. Character maps of scaled means of morphometric measurements of key morphological structures of the attachment organ used to characterise species groups within *Cichlidogyrus* based on subset of phylogenetic tree in Figure 2 with only one specimen per species. Numbers on nodes refer to clades in Figure 2. Ancestral states based on the Brownian Motion model supported by multivariate phylogenetic comparative analysis (Table 3). The sizes of the outer root of the dorsal anchor (DAc), the auricles of the dorsal bar (DBh), and the first (U1) and third hook pair (U3) display a limited phylogenetic information on their own as extreme values (dark blue or dark red) are spread out across the phylogenetic tree.

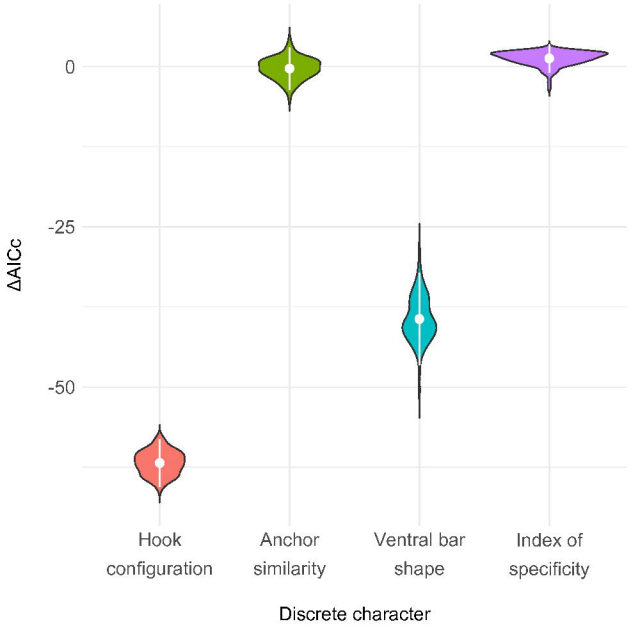

a

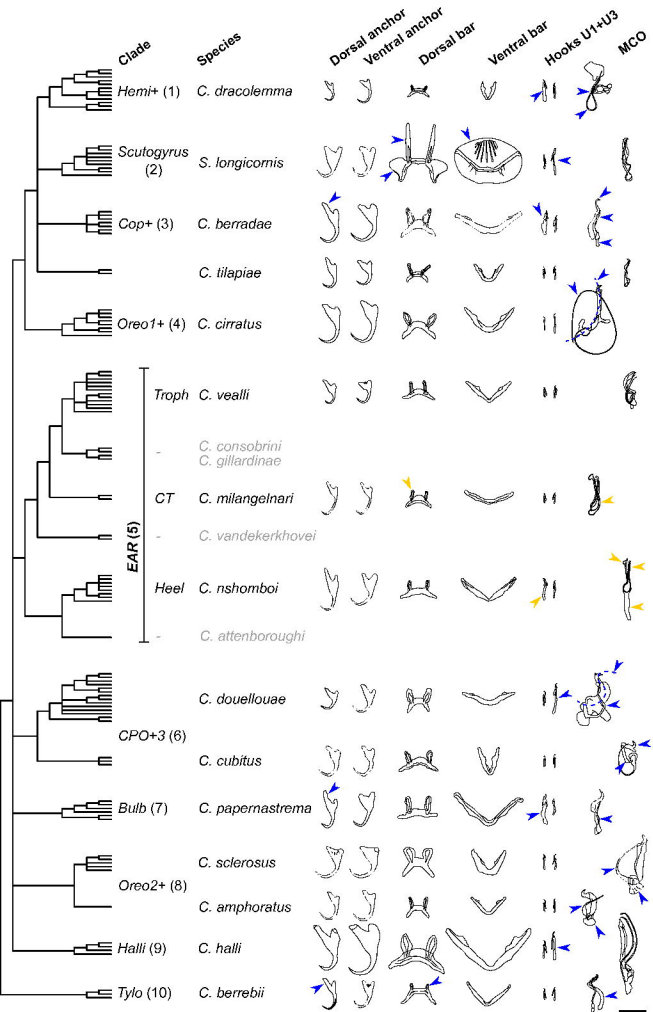

b
200 µm
Male copulatory organ
Vagina
Attachment organ

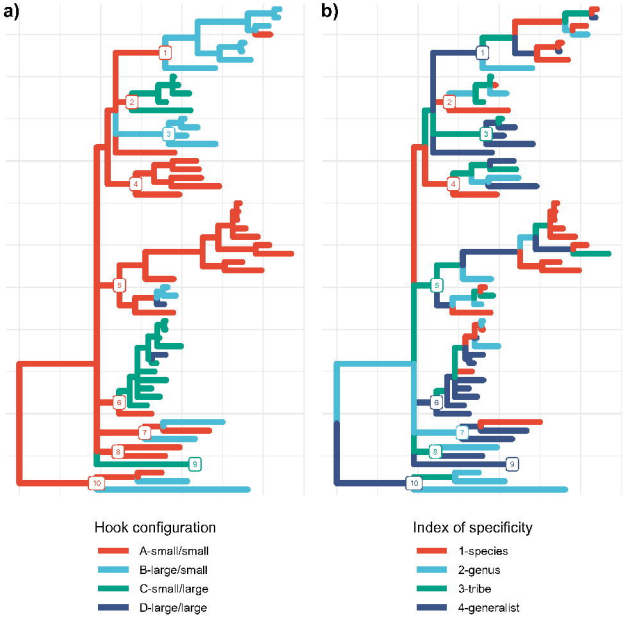

c
Attachment organ
Reproductive organs
Anchor
c
d
b
a
e
Hook
U
Male copulatory organ
AuP
Pe
AP
He
Ventral bar
x
w
Dorsal bar
y
w
h
x
Vagina
w
l


APLO
NOTH
POLY
Pure
Pyle
Cich
Chlo
Tyto
Hemi
Gall
Cept
Hete
Pelt
Tiza
Bual
Bath
Tiem
Bent
Cyst
Ebo
Hagi
Phasi
Sclerogrammi
Bagni
Tiliniformi
Cali
Cistrinaria
Sulfi
Gravi
Pelli
Tyli
0.03


a)
U1
DBh
Bulb
Tylo
b)
U3
DAc
Oreo2+
Scutogyrus
c)
U1
U3
Cop+
COTHGoP
Hemi+
Scutogyrus
d)
U3
DBh
Halli
Scutogyrus


0
-25
-50
ΔAICc
Hook configuration
Anchor similarity
Ventral bar shape
Index of specificity
Discrete character


Clade
Species
Dorsal anchor
Ventral anchor
Dorsal bar
Ventral bar
Hooks U1+U3
MCO
Hemi+ (1)
C. dracolemma
Scutogyrus (2)
S. longicornis
Cop+ (3)
C. berradae
C. tilapiae
Oreo1+ (4)
C. cirratus
EAR (5)
Troph
C. vealli
-
C. consobrini
C. gillardinae
CT
C. milangelnari
-
C. vandekerkhovei
Heel
C. nshomboi
-
C. attenboroughi
C. douellouae
CPO+3 (6)
C. cubitus
Bulb (7)
C. papernastrema
C. sclerosus
Oreo2+ (8)
C. amphoratus
Halli (9)
C. halli
Tylo (10)
C. berrebii


a)
b)
1
2
3
4
5
6
7
8
9
10
Hook configuration
A-small/small
B-large/small
C-small/large
D-large/large
Index of specificity
1-species
2-genus
3-tribe
4-generalist