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 Steenberge 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

A substantial portion of global biodiversity evolved through adaptive radiation. However, the effects of explosive speciation on species interactions remain poorly understood. Metazoan parasites infecting radiating host lineages could improve our knowledge because of the intimate host-parasite relationships. Yet limited molecular, phenotypic, and ecological data have discouraged multivariate analyses of evolutionary patterns and encouraged the use of discretised characters. Here, we highlight the limitations of these methods and propose new methods optimised for high-dimensional datasets. We apply multivariate phylogenetic comparative methods (PCMs) and statistical classification with support vector machines (SVMs) using morphometric and host range data widely inferred from a species-rich lineage of parasites (*Cichlidogyrus*, Platyhelminthes: Monogenea) infecting cichlid fishes. For PCMs, we modelled the attachment/reproductive organ and host range evolution using the data of 134/137 described species and an updated multi-marker (28S/18S rDNA, ITS1, COI mtDNA) phylogenetic reconstruction of 58 species. Through a cluster analysis, SVM-based classification, and taxonomic literature survey, we inferred the systematic informativeness of discretised and continuous characters. We demonstrate that character treatment for these parasites should be reconsidered as discretisation masks phylogenetic signal while remaining key for characterising species groups. Regarding the morphology, PCMs suggest divergent

evolutionary regimes for at least one lineage. A limited morphological variation indicates stabilising selection pressure through host and environmental parameters. However, high multicollinearity, a moderate 77% accordance rate of literature survey and SVMs, and a low systematic informativeness call for a reconsideration of the morphometric measurements included in species characterisations. We conclude that PCMs and SVM-based approaches are suitable tools to investigate character evolution in systems involving adaptive radiations or host-parasite interactions.

KEY WORDS

Adaptive radiation; *Cichlidogyrus*; Dactylogyridae; Monogenea; 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 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. Katemo Manda, C. Mukweze Mulelenu, M. Kasongo Ilunga Kayaba and C. Kalombo Kabalika (Université de Lubumbashi), M. Collet and P. N'Lemvo (Institut 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 that support the findings of this study are openly available in MorphoBank at www.morphobank.org, at https://dx.doi.org/XXXXXXXX. The DNA sequence data are openly available in the GenBank Nucleotide Database at https://www.ncbi.nlm.nih.gov/genbank, accession numbers XXXXXX–XXXXXX. Phylogenetic trees and data matrices are openly available in TreeBase at https://treebase.org, accession number XXXXXX.

INTRODUCTION

Adaptive radiations and host-parasite interactions

Adaptive radiation is one of the most important processes of species formation. These explosive speciation events might explain a substantial part of the biodiversity on the planet (Glor, 2010). Adaptive radiations are characterised by a rapid diversification resulting from adaptation to newly available ecological niches (Losos, 2010). Famous examples include the Darwin's finches (Grant, 1999), Carribean lizards of the genus *Anolis* Daudin, 1802 (Mahler et al., 2013), cichlid fishes (Salzburger, 2018), and Hawaiian silverworths (Landis et al., 2018). Despite the attention these examples have received, few studies have investigated the evolution of ecological interactions involving these groups and other organisms of other groups beyond their feeding ecology (e.g. Guerrero and Tye, 2009; Takahashi and Koblmüller, 2011; but see Karvonen and Seehausen, 2012; Blažek et al., 2018). How do organisms evolve that accompany the rapid diversification process of an adaptive radiation?

Metazoan parasites could provide answers to this question. Parasites have an intimate relationship with their hosts. As such, evolutionary patterns can be strongly impacted by the host evolutionary history (Huyse et al., 2005). In addition, parasites likely encompass a substantial part of the metazoan biodiversity (Poulin, 2014) and are known to shape species communities (Gómez and Nichols, 2013) from the tropics (Dunne et al., 2013) to the Arctic (Amundsen et al., 2009). Metazoan parasites also impact human health and food security (Thompson, 2013; Webster et al., 2016). Therefore, macroevolutionary processes of these organisms matter to society, especially in the context of human-wildlife spillover events (Thompson, 2013; Webster et al., 2016). Host-parasite interactions have also far-reaching potential as eco-evolutionary model systems in speciation research (see Toju et al., 2017) as

these associations are usually well-preserved in natural history collections (Harmon et al., 2019; Jorissen et al., 2020) because of parasites that infected collected specimens.

Traditionally, metazoan parasite macroevolution has been studied through morphological characters. Many species are anatomically adapted to their lifestyle by presenting prominent attachment and reproductive organs (Littlewood, 2006) to hold onto the host while creating offspring. Hence, these organs are frequently used not only for species descriptions but also for the study of parasite evolution (Kearn, 1968; Benkirane et al., 1999; Šimková et al., 2002; Matsuo et al., 2014). But how do attachment and reproductive organs evolve if their hosts rapidly diverge into a multitude of species?

Multiple scenarios apply to parasites infecting radiating host lineages. First, parasites that are strongly dependent on their hosts, might experience a selection pressure to remain competitive in the arms race with the host's defences (Kaltz and Shykoff, 1998). Structures relevant to this arms race such as attachment organs might experience a strong stabilising selection pressure as they are the parasite's main physical connection to the host. In this scenario, the attachment organ evolution would be heavily impacted by host and environmental parameters (Kaltz and Shykoff, 1998). Second, parasites could follow the example of their hosts in terms of an (adaptive) radiation. Such co-radiations have previously been reported for insects parasitic on plants or other insects (Weiblen and Bush, 2002; Forbes et al., 2009) but quantitative support for these hypotheses in the form of evolutionary models has not been provided to date. Third, parasites could have structures that are not under strong selection pressure and, therefore, their morphology might randomly diverge over time following a pattern associated with a 'random walk', i.e. genetic drift or randomly fluctuating selection (Losos, 2008). Reproductive organ structures, e.g. in flatworms infecting the gills of cyprinid fish (Šimková et al., 2002), have been suggested to follow this pattern causing reproductive isolation between species.

Phylogenetic comparative methods and statistical classification: a new approach to parasite taxonomy

Many methods have been used to study the evolution of morphological characters. The exponential increase of computational power has amplified the use of computationally intensive approaches to quantitatively assess evolutionary patterns. In this study, we focus on two such methods: 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 (e.g. Pennell et al., 2014; Clavel et al., 2015) and even models including multiple regimes (e.g. Beaulieu et al., 2012; Khabbazian et al., 2016). These models reflect evolutionary processes such as random walks, stabilising natural selection, and adaptive radiation (see Pennell et al., 2014; Clavel et al., 2019). To account for potential interactions between characters, new software packages implement multivariate models (Khabbazian et al., 2016; Goolsby et al., 2017; Clavel et al., 2019). Recently, models have been optimised for high-dimensional analyses (Clavel et al., 2019) to address the sensitivity of previous approaches to trait covariation, data orientation, and trait dimensions (Adams and Collyer, 2018). Yet to our knowledge, no study has applied this multivariate approach to parasitic organisms to date. Only univariate PCMs, i.e. for single parameters, have been used for monogenean flatworms infecting mugilid fishes in combination with geomorphometric analyses (Rodríguez-González et al., 2017).

Recent years have also seen an increase in 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). Amongst existing methods, supervised machine learning through support vector machines (SVMs) has been applied in taxonomic studies most frequently (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 input data. In the last step, the validated algorithm can be applied to a test dataset of samples with unknown group affiliation.

Species-rich but missing data: a case study for host-parasite systems

Phylogenetic comparative methods and machine learning both require large datasets that remain scarce for parasitic organisms. Metazoan parasites are often small, which makes identifying these organisms to species level notoriously difficult (de Meeûs et al., 2007). Furthermore, only a fraction of the known species is molecularly characterised (Poulin et al., 2019) and most species are yet to be discovered and described (Poulin et al., 2020). Of the best-known models in adaptive radiation research, African cichlids (Cichliformes, Cichlidae) are possibly the best-studied system (Salzburger, 2018) with approximately 2000 (described and undescribed) species reported from Eastern Africa alone (Turner et al., 2001; Salzburger et al., 2014). This knowledge is rooted in a long and productive tradition of international cichlid research (see Ronco et al., 2020; Van Steenberge et al., 2011). Resulting from this scientific interest, one group of gill parasites (Fig. 1a) infecting these fishes, the monogenean flatworms belonging to *Cichlidogyrus* sensu Paperna, 1960 (incl. the nested genus *Scutogyrus* Pariselle & Euzet, 1995 (Wu et al., 2007), together termed *Cichlidogyrus* in the following) (Fig. 1b, c), has been studied in more depth than other species-rich genus of parasites, in

particular from the African continent (Cruz-Laufer et al., 2021). Species of *Cichlidogyrus* rival (Cruz-Laufer et al., 2021) and possibly exceed (see Poulin, 2014) their hosts in terms of species numbers: 137 parasite species have been described on 126 fish species (Cruz-Laufer et al., 2021). The cichlid-*Cichlidogyrus* species network has even been proposed as model system for macroevolutionary research in host-parasite systems (Pariselle et al., 2003; Vanhove et al., 2016) because of this diversity. 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).

The species-richness of *Cichlidogyrus* has previously been attributed to co-divergence with their hosts (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 other parasites, morphological characters of attachment and reproductive organ structures have formed an essential part of species descriptions of monogenean flatworms. Evolutionary studies have paid considerable attention to the evolutionary patterns of these (mostly) sclerotised structures (e.g. Šimková et al., 2002; Mendlová et al., 2012). For species of *Cichlidogyrus*, the attachment organ includes a haptor with several anchors, hooks, and bars and the reproductive organs encompass the male copulatory organ (MCO) and the vagina (Fig. 1b, c). Based on the similar morphologies observed in related species, some studies have suggested phylogenetic constraints, i.e. ancestry of the parasite species, as 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). These host switches are considered rare as monogenean representatives have been reported for 27 (Carvalho Schaeffner, 2018) of 48 (Lévêque et al., 2008) fish families in Africa and most harbour some other dactylogyridean lineage of gill parasites (see Carvalho Schaeffner, 2018). 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 present study aims to investigate the evolutionary processes that have shaped the evolution of species of *Cichlidogyrus*. However, out of the 137 species that are currently described, DNA sequences of only 18 species were included in the most recent phylogenetic study (Messu Mandeng et al., 2015). Furthermore, morphological and ecological data have not been collected in centralised databases inhibiting the progress of large-scale meta-analytical studies (Cruz-Laufer et al., 2021). Previous studies have evaded this scarcity of available information by discretising the attachment organ measurements or host range data. Vignon et al. (2011) and Mendlová et al. (2012) grouped species by the relative size of the sclerotised structures of the attachment organ (for an overview of the terminology, see Fig. 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). 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.

To address data scarcity, we provide new morphological, molecular, and host range data which are made available in public databases. We apply PCMs and SVMs to an updated

phylogeny and the morphological and host range data to investigate evolutionary mechanisms. With these approaches, we want to answer questions on the systematic informativeness of these characters and on the evolutionary patterns of phenotypic characters. These questions are:

(i) How systematically informative are morphometric measurements that are currently widely used and can we increase phylogenetic signal by reducing the numbers of input variables through a cluster analysis?

(ii) Does discretisation lead to information loss in morphometric and host range data?

(iii) As species of *Cichlidogyrus* infect a host lineage known for its rapid speciation, which evolutionary processes have shaped the attachment and reproductive organ morphology? Are these structures under stabilising selection pressure (H_2), do the parasites mirror the radiations of their hosts either as an early (H_3) or a late burst (H_4) in the character evolution, or do these structures follow a pattern associated with a random walk (H_1)? Furthermore, do parasites infecting the Eastern African lineages of cichlids follow a different evolutionary regime than the species infecting other hosts (H_5)?

(iv) 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 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 instead of 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 at GenBank (see Appendix 2). The alignment includes partial ITS1, 18S rDNA, 28S rDNA, and COI mtDNA sequence data. Partial DNA sequence data (i.e. specimens with less than those four loci sequenced) were included with the lacking fragments coded as missing data. We aligned the sequences of each locus using the algorithm *L-INS-i* 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 root the phylogenetic trees 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 software 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 with 500 random sequence addition (RAS) replicates, tree bisection and reconnection (TBR) branch swapping, and 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 verify 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 CADM 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* v4.1.0 (R Core Team, 2021) to calculate phylogenetic pair-wise distance matrices and to conduct the CADM test.

Morphological and host range data collection

We assembled morphometric (Fig. 1c) and host range data for all species belonging to *Cichlidogyrus* and *Scutogyrus.* We assembled morphometric data for 134 species of *Cichlidogyrus* and *Scutogyrus* including the 58 species represented in the phylogenetic analysis (Fig. 2a). Raw measurements taken through light microscopy were assembled 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; Kmentová et al., 2016a, 2016b, 2016c; Rahmouni et al., 2017, 2018; Igeh et al., 2017; 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., 2018b, 2018a) without standard errors. If no mean was reported, we inferred the measures from drawings provided in previous publications (Dossou, 1982; Dossou and Birgi, 1984; Birgi and Lambert, 1986; Muterezi Bukinga et al., 2012) by calibrating the images size through the scale bars provided in these studies. 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, 1960, 1968; Paperna and Thurston, 1969) 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). Hook configurations were coded according to the secondary growth in the first and third hook pair (‘small-small’, ‘small-large’, ‘large-small’, ‘large-large’) that can make these pairs larger than the second pair, which retains it embryonal size in the adult organism (Llewellyn, 1963). The anchor pairs were categorised as ‘similar’ or ‘dissimilar’ in shape and size based on the original species descriptions. The ventral bar shape was defined according to the presence of membranous extensions (see Mendlová et al., 2012) and size (‘with membranous extensions’,

'without membranous extensions', 'massive with membranous extension', 'bar supports large extended plate (*Scutogyrus*)'). We inferred IS data from the host-parasite interactions listed by Cruz-Laufer et al., (2020) with species exclusively reported from a single species, genus, or tribe classified as species-, genus-, and tribe-specific, or generalist if hosts belonging to two different tribes (or even families) were infected by the same parasite species. For the classification of cichlid species into tribes, we followed Dunz and Schliewen (2013). All discrete and continuous morphometric and host specificity data are made available at TreeBase alongside the DNA sequence alignments and tree topologies (www.treebase.org, accession number: XXXXX). Raw morphometric data are made available at MorphoBank (www.morphobank.org, project XXXXX).

Cluster analysis

To minimise model complexity and avoid overfitting statistical models, we produced a second dataset with a reduced number of continuous variables for downstream analyses, i.e. phylogenetic comparative methods and statistical classification. To achieve this variable reduction, we identified clusters of multicollinear morphometric measurements (Fig. 1b, c) and excluded all but one measurements from each cluster as multicollinearity amongst the variables can lead to unstable and unreliable models (Mundry, 2014). We used a Pearson pairwise correlation matrix (Dormann et al., 2013) and the *ward.d2* clustering algorithm (Murtagh and Legendre, 2014) to detect clustered variables in the *R* package *ComplexHeatmap* v2.8.0 (Gu et al., 2016). If the absolute values of the Pearson correlation coefficients (|r|) exceeded 0.7 (Dormann et al., 2013), we considered measurements multicollinear. Hence, we split the variables into clusters along the dendrograms provided by *ward.d2* until all coefficients within the same cluster exceeded this threshold.

Character evolution of the attachment and reproductive organs

Using the full dataset and the reduced dataset resulting from the cluster analysis, we tested for possible patterns of a random walk (H_1) stabilising selection (H_2) and adaptive radiation with a decelerating (H_3) and accelerating divergence of characters (H_4) on the evolution of the attachment and reproductive organs. We also tested if species from the East African lakes followed a different evolutionary regime (H_5) as a result of the multiple host radiations in this region. To detect these patterns, we fitted a range of multivariate models of continuous character evolution on a single sample of 100 randomly selected tree topologies drawn from the post–burn-in phase of the BI analysis. We excluded all but one specimen per species from the phylogenetic trees to avoid pseudoreplication in the phylogenetic comparative analyses. Specimens included in the subset trees, were chosen at random and are highlighted in Appendix 2. The tested models include:

- The Brownian motion model (BM) (Felsenstein, 1973) simulates a ‘random walk’, i.e. genetic drift or randomly fluctuating selection (see Hansen and Martins, 1996) (H_1). In this case, the parasite morphology would not be affected by strong selection pressures and would randomly diverge over time.
- The Ornstein-Uhlenbeck model (OU) (Butler and King, 2004) approximates a character evolution under stabilising selection (H_2). Here, the parasite morphology experiences a selection pressure towards a selective optimum due to a high host-dependence.
- The Early-Burst (EB) (Harmon et al., 2010) or accelerate-decelerate model (ACDC) (Blomberg et al. 2003) models an early rapid evolution of characters followed by a slow-down such as might occur during adaptive radiation events (H_3). Here, the parasite morphology has rapidly diverged possibly mirroring the explosive speciation events reported for some of the host lineages.
- The Late-Burst model (LB), a variation of the ACDC model with an accelerating divergence of characters (Blomberg et al., 2003) (H_4), simulates evolution as expected

under a recent, ongoing radiation event. The speciation of the parasites would mirror the host radiations but would still be ongoing.

- The multi-rate Brownian motion model (BMM) represents two different multi-selective and multi-rate Brownian motion regimes for East African parasite lineages and other lineages (O'Meara et al., 2006). Here, the parasite morphology would have undergone an evolutionary rate shift because species infecting the host radiations in Eastern African have developed differently than their congeners in the rest of Africa. The regimes were defined using the function *paintSubTree* in the *R* package *phytools* v0.7-80 (Revell, 2012).

We fitted all models using the *mvgls* function (Clavel et al., 2019) in the *R* package *mvmorph* v1.1.4 (Clavel et al., 2015) as recommended for independently evolving measurements with a natural orientation. In addition, we expected the cluster analysis to minimise interdependence of the input 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 *mvgls*, we set the upper bound of the EB search algorithm to 10. For the BMM model, we defined a separate regime for the *EAR* clade (Fig. 2a), which encompasses species infecting cichlids of the East African radiation (Fig. 2b). As traditional information criteria such as the Akaike information criterion (AIC) are not applicable to penalised likelihoods (Clavel et al., 2019), we assessed the model performance with the extended information criterion (EIC) (Ishiguro et al., 1997; Kitagawa and Konishi, 2010) using the *EIC* function in the *R* package *mvmorph* with 100 bootstrap replicates. Smaller values for EIC were considered indicative of a better model performance. To test for phylogenetic signal in the

datasets, we also inferred Pagel's λ (Pagel, 1999) through the *mvgls* function using the *lambda* option.

Finally, we modelled the discrete character evolution of the categorical data (see *Morphological and Ecological Data Collection*) under a univariate Pagel's λ model (Pagel, 1999) using the *fitDiscrete* function in the *R* package *geiger* v2.0 (Pennell et al., 2014) to assess the effects of data discretisation. We applied this function to the same set of 100 randomly selected BI subset trees selected for the morphometric data to account for phylogenetic uncertainty. Model performance was assessed through the sample size–corrected Akaike information criterion in relation to a white noise model (ΔAICc) that assumes absolute phylogenetic independence (Pennell et al., 2014).

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 77/79 species not included in the phylogenetic analysis in the species groups inferred from the tree topology in Fig. 2a. 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/family of non-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. Host ranges (natural and invasive) and geographical ranges by country were taken into account as we expected the host environment to be a relevant factor in speciation processes of these parasites (see McCoy, 2003; Huyse et al., 2005). We used the host-parasite list and literature databases provided by Cruz-Laufer et al. (2021) 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). Missing data in the measurements were imputed 100-fold through an expectation-maximisation (EM) algorithm as implemented in the *R* package *amelia* (Honaker et al., 2011). 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 function *tune* of the *R* package *e1071* (Meyer et al., 2020). This optimisation was conducted on a subset of 10 randomly sampled imputed datasets. Furthermore, overfitting has been suggested to be a major hurdle for machine learning approaches (Jabbar and Khan, 2015). Therefore, we applied SVMs to both full and reduced datasets (see *Cluster analysis*) of the combined attachment and reproductive organ measurements to test the effect of overfitting. To assess the performance, we computed the area under the receiver operating curve (AUROC) as widely applied for machine learning algorithms (Huang and Ling, 2005). Classification errors were inferred from the results suggested by the taxonomic classification approach. Finally, we permuted each variable 100-fold for every imputed dataset to infer variables importance for SVM predictions of the best dataset measured as the z-scores of AUROC values in relation to the non-permuted data.

Graphing

We plotted the graphs and phylogenetic trees using the *R* packages *ggplot2* v3.3.5 (Wickham, 2016) and *ggtree* v3.13 (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. 2a) and duplicate sequences of the

same species dropped from the topology similar to the phylogenetic comparative analysis (highlighted in Appendix 2). Node states were inferred from the output of the best-fitting model. To map 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* v0.7-80 (Revell, 2012). Heatmaps were plotted using the *R* package *ComplexHeatmap* v2.8.0 (Gu et al., 2016).

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 mtDNA. The combined data set includes sequences of 103 specimens belonging to 58 species (see Appendix 2), and has a length of 2173 base pairs following the removal of poorly aligned positions and divergent regions (missing sequence data were treated as gaps). 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 codon positions 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, χ^2 = 2506, p < 0.01), ML (Kendall's W = 0.75, χ^2 = 2523, p < 0.01), and MP (Kendall's W =0.65, χ^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, χ^2 = 11044, p < 0.01), we only displayed the BI phylogram but report support values of BI, ML, and MP analyses in the phylogram (Fig. 2a).

We found 10 well-supported clades within *Cichlidogyrus* 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) (see *Characterisation of Species Groups* for details of morphology and host range). 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., 2018b) 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 the sister group to other species of *Scutogyrus* with moderate support (0.99/*/90/*).

Cluster Analysis

Based on a Pearson pairwise correlation matrix (Fig. 3), we detected 15 clusters of multicollinear measurements. Based on the minimum amount of missing data, we selected one proxy variables for each cluster of attachment and reproductive organ measurements (Fig. 3) omitting 14 variables and resulting in a reduced dataset with 15 variables.

Character evolution of the attachment and reproductive organs

For the attachment organ measurements, the multivariate PCM performed significantly better with the reduced dataset as EIC estimates were consistently lower than for the full dataset (Fig. 4a). We also detected a strong phylogenetic signal in the reduced (λ = 0.94; CI: 0.81–1.00) and full datasets (λ = 0.91; CI: 0.77–1.00). However, Brownian Motion (BM), Ornstein-Uhlenbeck (OU), and Early-Burst (EB) models all performed equally in the reduced dataset (Fig. 4b). The Late Burst (LB) model performed worse. Only the multi-rate BM model (BMM) performed better albeit with a slight overlap in EIC estimates. For the reproductive organ measurements, EIC estimates increased for the reduced dataset (Fig. 4a) and the phylogenetic signal was considerably weaker for the reduced (λ = 0.74; CI: 0.39–1.00) and full datasets (λ = 0.73; CI: 0.43–1.00). Model fits of the full dataset showed a similar pattern to the reduced dataset of the attachment organs (Fig. 4a). Out of the discrete characters, only the hook configuration (HC) and ventral bar shape (VBS) produced strong phylogenetic signals (Fig. 4c). 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. 2a), we found shared morphological characteristics in eight. We were able to affiliate all but 12 species without available DNA sequences to these clades (Appendix 3). Some morphological structures were particularly relevant for characterising species groups of *Cichlidogyrus*. Character states of these structures were frequently shared and features that were key for the characterisation of the proposed species groups (Appendix 3; also Fig. 6, Fig. 7, and Fig. 8). Based on the number of species groups defined by a structure, the male copulatory organ (MCO) was the most distinctive structure (eight species groups) as phenotypes were often specific to a species group or subgroup (Fig. 6). The size and shape of the first hook pair (U1) were 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 the detailed morphological characterisations of all mentioned groups and a discussion on the host and geographical ranges, see below and the overview of the morphological features highlighted in Fig. 6 (full species names are cited in Appendix 3 and reported alongside key features and host names; for abbreviations of morphometric measures, see Fig. 1c).

Accordance of the proposed group affiliation of the species with the SVM-based statistical classification approach (Fig. 9a) was overall acceptable (mean AUROC: 77%, CI: 72%– 80%) when using the reduced dataset of the attachment organ with 1014 observations in the training data and 638 observations in the test data (not including observations of species without proposed species group). Model performance did not improve considerably when using the full dataset (AUROC: 78%, CI: 75%– 80%). Model tuning mostly provided

the best result for $C = 8$ and $\gamma = 0.125$ for both datasets. Furthermore, the error rate of the SVM-based approach varied substantially between the different species groups (Fig. 9a) ranging from 89.3% correctly classified specimens on average for species in the group of *Scutogyrus* to 12.4% and 1.0% respectively for species classified in the *Tylo* and *Bulb* species groups. Species of the *Oreo2+* and the *EAR* groups shared no apparent features in their attachment and reproductive organs with species of the same group apart from the characteristics mentioned in the genus diagnosis (Pariselle and Euzet, 2009). Yet the accordance between taxonomic and statistical classification approaches was above average for the *EAR* group with 82.7%. For the reduced dataset, the most important character was the length of dorsal bar auricles, followed by the length of the first hook pair (U1) (Fig. 9b). Permutations of all other variables and in particular measurements of the anchors in the attachment organ (VA) did not significantly lower model performance (Fig. 9b).

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. 6) 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*) (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 (Dossou and Birgi, 1984; Pariselle and Euzet, 2004; Vanhove et al., 2018) 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) and a ventral bar that supports a large plate (Fig. 6).

3. *Cop+: Species infecting coptodonine cichlids*. Anchor pairs dissimilar in shape. Dorsal anchor with V-shaped indentation between the roots and an elongated inner root at least four times longer (Fig. 6) 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 penis. 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. Species of the *Cop+* clade have previously been grouped together (Pariselle and Euzet, 2003; Jorissen et al., 2018b) but without *C. berminensis*.

4. *Oreo1+: Species infecting oreochromine and coptodonine cichlids*. Anchor pairs similar in shape and size but dorsal anchors 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 of accessory piece attached to penis bulb and distal end holds the distal end or a mid-portion of 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 from 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 Fig. 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 for the whole group. However, some 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. (2018b).
- *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. 2a).

- *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, both variably shaped; the heel is finger-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. Accessory piece simple, elongated with few remarkable structures, sometimes with thickened (*C. gistelincki*), 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. 2a) but also displays a bifurcate distal end. According to the phylogenetic analyses (Fig. 2a), *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 accessory piece 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; Vanhove et al., 2018). *Cichlidogyrus amphoratus, C. bulbophallus, C. flagellum, C. giostrai, C. karibae, C. lagoonaris, C. njinei, C. ornatus*, and *C. sanjeani* 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 dorsal 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, i.e. a 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), and *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 this group (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. (2018b).

Other species. For 12 species, we found no 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 4). On average a majority of observations for *C. fontanai* (9.5/10), *C. rognoni* (14.3/20), *C. sanjeani* (6.9/10) were classified in *Oreo1+* group. All 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 9.3/10 observations of *C. albareti* in the *EAR* group. For *Cichlidogyrus arfii* classifications were more ambiguous with 4.8/11 added to 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 most observations of *C. bulbophallus* (55.7/66) and *C. lobus* (4.7/5) unlike *C. flagellum* (3.9/8) and *C. ranula* (1.4/3) which were classified in the *CPO+3* group. Only a single observation could be provided for *C. karibae* (*EAR* group, Appendix 5). Classifications for *C. tilapiae* through SVMs revealed a morphological similarity with species of the *EAR* group (13/20) in contrast with the results of the phylogenetic analysis (Fig. 2a). 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

We investigated the evolution of the attachment and reproductive organs of parasites infecting a host lineage that has undergone multiple adaptive radiation events (Seehausen, 2006). Species of *Cichlidogyrus* (incl. the nested *Scutogyrus*) form a morphologically diverse lineage of parasitic flatworms infecting the gills of African cichlid fishes. As such, these monogeneans are a proposed model system for parasite macroevolution (Pariselle et al., 2003; Vanhove et al., 2016). Here, we employ 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 African metazoan parasites to date. We were able to identify and characterise 10 different species groups (Fig. 2a; Fig. 6) amongst the 137 species that are currently known (Cruz-Laufer et al., 2021). For some of these groups some member species have 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 58 species and morphometric

data of 134 species. We here 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 yet undiscovered (Cruz-Laufer et al., 2021), the relationship between the proposed species groups remains unresolved with only moderate support for a few macroclades (Fig. 2a).

Systematic informativeness of morphometric measurements and host range data

To identify and classify species of *Cichlidogyrus*, previous publications have applied a specific set of measurements of the sclerotised attachment and reproductive organs (Cruz-Laufer et al., 2021). We investigated the systematic informativeness of the 29 measurements most frequently provided in the literatures (Fig. 1c). A cluster analysis revealed that many of the measurements (14) might be redundant due to an abundance of multicollinear relationships (Fig. 3). Therefore, we used a fraction of these measurements (15/29) as an alternative reduced dataset for downstream analyses. This approach resulted in improved performance for some models in the PCM approach (Fig. 4b) and no decreased performance in all other cases (Fig. 4a) and the machine learning approach. Furthermore, the phylogenetic signal of the data increased as a result of variable reduction but the signal was weak for reproductive organ measurements.

The low systematic informativeness of the measurements is also reflected by the performance of a classification analysis using support vector machines (SVMs). The trained algorithm was not able to predict species group affiliation accurately based on the most informative (reduced) or full dataset. Consequently, accordance of the taxonomic literature survey and SVM-based approaches in this study never exceeded 89.3% for any of the proposed species groups. One cause for this moderate performance might be the low number

of systematically informative measurements in the dataset as only nine of them caused a lower performance if permuted (Fig. 9b). Notably, the phylogenetic signal was also low for any of the evolutionary models applied to reproductive organ measurements despite their importance to the machine learning algorithm. This observation contradicts a taxonomic literature survey, which suggested that the phenotype of the male copulatory organ was key for characterising groups of related species. However, the taxonomic assessment of the morphology allowed us to consider shape information of the reproductive organs available in the literature in contrast to a SVM algorithms that is based purely on numeric values. Therefore, we could affiliate all but 12 species with the proposed species groups. This relative success highlights the value of phylogenetic information conveyed by reproductive organ structures; e.g. Rahmouni et al. (2017) noted that the heel of the male copulatory organ was shown to be similarly shaped in possibly related Eastern African species. Similar conclusion have been drawn for other dactylogyrid monogeneans such as species of *Characidotrema* Paperna & Thurston, 1968 (Řehulková et al., 2019). Regarding the lack of a phylogenetic signal of reproductive organ measurements, Pouyaud et al. (2006) proposed that this issue might be explained by a fast evolutionary rate but we suggest that this observation might be rather linked to a limited taxon coverage as more species have been described ever since. Instead, the low systematic informativeness indicated here more likely reflects a lack of systematically informative metrics as, for example, the number of reproductive organ measurements is low compared to the attachment organ (5 vs. 24). In conclusion, the measurements that are currently employed, fail to incorporate the systematic information offered by attachment and reproductive organs.

The low predictive power of the morphometrics has encouraged the use of discretisation in the past. Several studies have categorised species according to attachment

organ measurements (Vignon et al., 2011; Mendlová et al., 2012) and host range data (Mendlová and Šimková, 2014). Discretisation 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 but the phylogenetic informativeness of the ‘shape of anchors’ character was never investigated further. In fact, the cluster analysis of the present study shows that dorsal anchor measurements and the respective ventral anchor measurements are highly multicollinear (Fig. 3) with the exception of the inner roots (DAd and VAd). The ventral bar shape was also introduced as variable for character mapping but no ancestral state could be inferred (Mendlová et al., 2012). Yet species of *Scutogyrus* present a unique character state with an extended plate associated to the bar structure (Pariselle and Euzet, 1995a). 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 founded on 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, hook measurements were informative in a machine learning context (Fig. 9b) 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*, 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 studies.

A discrete variable proposed to represent 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., 2021), 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 again reflect a study bias in cichlid parasitology, in this case towards large economically relevant host species (Cruz-Laufer et al., 2021). Because of their importance to fisheries, these fish species unlike others have been introduced to various new habitats. 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) as well as temporal fluctuations (Brooks et al., 2019). We suggest that these parameters could be included, e.g., in a genus-wide comparative study in the future.

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

Attachment organ evolution: Rate heterogeneity 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. 4b). However, we found no significantly increased performance of the Ornstein-Uhlenbeck (OU), Early Burst (EB), and Late Burst (LB) models compared to the Brownian Motion (BM) model. Only the multi-rate BM (BMM) model performed slightly but significantly better than the other models (Fig. 4b). The OU, EB, and LB models present extensions of the Brownian Motion (BM) model (Butler and King 2004; Harmon et al. 2010). Therefore, the character evolution of the attachment organs measurements might simply follow a pattern associated with genetic drift or randomly fluctuating natural selection (Losos, 2008). This pattern might also be reflected in morphological features that are frequently shared by related species (Fig. 6). The relatedness of the parasite species could determine the evolution of attachment organ as suggested by a previous study (Vignon et al. 2011). Studies on a related lineage of monogeneans (*Ligophorus* Euzet & Suriano, 1977), also indicate a correlation of morphometrics and phylogenetic relationships (Rodríguez-González et al. 2017). However, the increased support for the BMM model suggests that the evolution of these parasites might have been shaped by different evolutionary regimes. This rate heterogeneity appears logical as ancestral host lineages have also developed differently compared to the East African radiations of cichlids (see Brawand et al., 2015). In fact, purely single-rate models appear increasingly unlikely to capture the real evolutionary history of *Cichlidogyrus* because of its pan-African distribution (Cruz-Laufer et al., 2021) and ecological diversity of hosts (Burress, 2014). In the present study, we only modelled a multi-rate BM model because a lack of software packages suited for high-dimensional dataset with pre-set hypothesis. Yet other multi-rate models including

OU, EB, and LB processes could better fits. In particular, species of the *EAR* group (Fig. 2a) that infect cichlid radiations in East Africa (Fig. 2b) might mirror the explosive speciation of their hosts (see Vanhove et al., 2015). Recent years, have produced multiple tools to investigate these more complex models but often these methods represent maximum likelihood (ML) approaches (e.g. Beaulieu et al., 2012; Ingram and Mahler, 2013; Puttick, 2018) that are sensitive to trait covariation, data orientation, and trait dimensions, particularly for high-dimensional datasets (Adams and Collyer, 2018). Other methods were developed to detect evolutionary rate shifts (e.g. Ingram and Mahler, 2013; Uyeda and Harmon, 2014; Khabbazian et al., 2016) rather than to test hypotheses based on previous knowledge (such as for the *EAR* clade in the present case). Therefore, penalised log-likelihood approaches that address these issues remain currently limited to multi-rate BM models but future updates will most likely close this technological gap.

For the time being, the taxonomic literature survey paints a more nuanced picture of the attachment organ morphology. For instance, we detected patterns that would be expected under stabilising selection as the number of different phenotypes appears relatively restricted. Highly similar 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 occurred within species groups appear to be adaptations to a diverging host range or environment (see Messu Mandeng et al., 2015). Host- and environmentally induced shifts in the attachment organ morphology have also been observed amongst other dactylogyrids, e.g. species of *Kapentagyrus* (Kmentová et al., 2018) and *Thaparocleidus* (Šimková et al., 2013), and other 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. 2b; 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 possibly 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. 2b), 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 ecological and morphological diversity of the hosts, e.g. related to feeding behaviour (Burress, 2014). The parasites, which we grouped into subgroups of the *EAR* clade, might have adapted to these ecologically diverse hosts resulting in divergent attachment and reproductive organ morphologies.
- Co-infections of different species of *Cichlidogyrus* can result in niche segregation on the gills of a host such as in species infecting Lake Victoria cichlids (Gobbin et al., 2020). At least some co-infecting species can be told apart based on their attachment organ morphology (see Gobbin et al., 2021). Thus, phenotypic differences might be linked to specific infection sites. However, morphological adaptations to gill microhabitats have been reported for other monogeneans (Rohde, 1976; Ramasamy et al., 1985) but no study has so far investigated these mechanisms for species of *Cichlidogyrus*.

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). Furthermore, 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*) (Paperna, 1960; Jiménez-Garcia et al., 2001; Šimková et al., 2019). While the literature suggests that most African fish families have their own monogenean lineages (Carvalho Schaeffner, 2018) and the number of similar host switches might, hence, be limited, 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 sampling, which has affected similar studies in the past. For instance, previous studies were limited to single structures, i.e. the anchors (Rodríguez-González et al., 2017), or to the much lower number of species described at the time (77 vs. 130 today), which comprised less phenotypical diversity than is currently known to exist (Vignon et al., 2011). In fact, 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. Similarities of the attachment organ morphology in related species reflected the specificity of the parasites to the respective host lineage rather than only the relatedness of the monogeneans. While the present analysis includes more species, 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., 2021). This data deficiency likely affects the performance of the models implemented through PCMs as, e.g., the lack of lineages might lower the resolution of phylogenetic inference and increase phylogenetic uncertainty. A more complete taxon coverage of morphological, molecular, and ecological data will 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 induced by phenotypic adaptation to the rapidly speciating host lineages. A slightly better performance of the multi-rate model, provides reason to further investigate the evolutionary history as the evolution of this group might fall under a divergent evolutionary regime. Lastly, convergent evolution of morphological characters might also have affected PCM model performance. Convergence can mask evolutionary processes that have occurred in the past as phenotypes become more similar (Losos, 2008). Monogenean parasites infecting the haplochromine radiations in the Eastern African lakes might be particularly impacted by this effect because of the convergent evolution of their hosts. These fishes have evolved via replicated adaptive radiation events, independently resulting in similar yet unrelated species occupying the same niches in different lakes (McGee et al., 2016). This setting makes their parasites potential targets to study convergent evolution in parasites. Conversely, the Eastern African diversity of *Cichlidogyrus* has mostly been investigated in Lake Tanganyika, the oldest lake, limiting the

current possibilities to study parasite evolution in the context of replicate host radiations in this region (but see Gobbin et al. (2021) for species from Lake Victoria).

Reproductive organ evolution

Unlike the attachment organ, the reproductive organs show no character changes correlated to 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 determined 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. We detected a higher support for the BMM model (Fig. 3b) similar to the attachment organ suggesting that the *EAR* group, with its host lineages that have resulted from adaptive radiation events, has likely experienced evolutionary regime divergent from its congeners. However, the low phylogenetic signal in these data provides far less confidence in this result. Previous studies on the reproductive organs of monogenean flatworms have suggested that the complex surface of the reproductive organ might correlate with environmental factors, e.g. 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. 6, *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 higher number of observations are needed. The results of the taxonomic literature survey and the 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., 2021) 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 as this study highlights the limitations of using morphological characters to infer the evolution of monogenean parasites. Nonetheless, 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., 2021).

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. Notably, these species belong to one of the more extensively investigated lineages in close interaction with an adaptive radiation of other species, namely cichlid fishes. As adaptive radiations

might explain a substantial part of the biodiversity on the planet, many species potentially interact with radiating lineages. Therefore, understanding evolutionary history in this context could be key for understanding the diversity of species interactions in general.

Similar to many parasite systems, the limited availability of phenotypic and genotypic data for the 137 species of *Cichlidogyrus* 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. 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 in this taxon, 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 and reproductive organs suggested by previous studies. No support for alternative models suggests that the variation of the morphology of these organs across the genus has been shaped by different evolutionary regimes. Morphological patterns in different species groups indicate a more nuanced picture with the attachment organ morphology being constrained by host parameters. These constraints do not apply to the reproductive organs that appear to follow a more random evolutionary regime. 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 picture 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

Adams, D.C., Collyer, M.L., 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. . Syst. Biol. 67, 14–31.

Altman, D.G., Royston, P., 2006. The cost of dichotomising continuous variables. . Br. Med. J. 332, 1080.

Amundsen, P.-A., Lafferty, K.D., Knudsen, R., Primicerio, R., Klemetsen, A., Kuris, A.M., 2009. Food web topology and parasites in the pelagic zone of a subarctic lake. . J. Anim. Ecol. 78, 563–572.

Baillie, C., Welicky, R.L., Hadfield, K.A., Smit, N.J., Mariani, S., Beck, R.M.D., 2019. Hooked on you: shape of attachment structures in cymothoid isopods reflects parasitic strategy. . BMC Evol. Biol. 19, 207.

Beaulieu, J.M., Jhwueng, D.C., Boettiger, C., O'Meara, B.C., 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. . Evolution 66, 2369–2383.

Benkirane, O., Coste, F., Raibaut, A., 1999. On the morphological variability of the attachment organ of Lernaeopodidae (Copepoda: Siphonostomatoida). . Folia Parasitol. 46, 67–75.

Birgi, E., Euzet, L., 1983. Monogènes parasites des poissons des eaux douces du Cameroun. Présence des genres *Cichlidogyrus* et *Dactylogyrus* chez *Aphyosemion* (Cyprinodontidae). . Bull. la 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.

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

Bitja-Nyom, A.R., Agnèse, J.F., Pariselle, A., Bilong-Bilong, C.F., 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 5, 3779–3803.

Blažek, R., Polačik, M., Smith, C., Honza, M., Meyer, A., Reichard, M., 2018. Success of cuckoo catfish brood parasitism reflects coevolutionary history and individual experience of their cichlid hosts. . Sci. Adv. 4, eaar4380.

Blomberg, S.P., Garland, T., Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. . Evolution 57, 717–745.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O., Di Palma, F., 2015. The genomic substrate for adaptive radiation in African cichlid fish. . Nature 513, 375–381.

Brooks, D.R., Hoberg, E.P., Boeger, W.A., 2019. The Stockholm Paradigm: climate change

and emerging disease. . University of Chicago Press, Chicago, USA.

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

Butler, M.A., King, A.A., 2004. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. . Am. Nat. 164, 683–695.

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.

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

Carvalho Schaeffner, V.C., 2018. Part 5: Host-parasite list. . In: Scholz, T., Vanhove, M.P.M., Smit, N., Jayasundera, Z., Gelnar, M. (Eds.), A Guide to the parasites of African freshwater fishes. CEBioS, Royal Belgian Institute of Natural Sciences, Brussels, Belgium, pp. 361–402.

Chernomor, O., von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. . Syst. Biol. 65, 997–1008.

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.

Clavel, J., Escarguel, G., Merceron, G., 2015. mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. . Methods Ecol. Evol. 6, 1311–1319.

Cruz-Laufer, A.J., Artois, T., Smeets, K., Pariselle, A., Vanhove, M.P.M., 2021. The cichlid–

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 *Cichlidogyrus* network: a blueprint for a model system of parasite evolution. . Hydrobiologia 848, 3847–3863.

de Meeûs, T., McCoy, K.D., Prugnolle, F., Chevillon, C., Durand, P., Hurtrez-Boussès, S., Renaud, F., 2007. Population genetics and molecular epidemiology or how to *débusquer la bête*. . Infect. Genet. Evol. 7, 308–332.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., Mcclean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. . Ecography 36, 27–46.

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

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

Douëllou, L., 1993. Monogenean of the genus *Cichlidogyrus* Paperna, 1960 (Dactylogyridae: Ancyrocephalinae) from cichlid fishes of Lake Kariba (Zimbabwe) with description of five new species. . Syst. Parasitol. 25, 159–186.

Dryden, I.L., Mardia, K. V, 2016. Statistical shape analysis: with applications in R, 2nd ed. , Wiley series in probability and statistics. John Wiley & Sons, Chichester, UK.

Dunne, J.A., Lafferty, K.D., Dobson, A.P., Hechinger, R.F., Kuris, A.M., Martinez, N.D., McLaughlin, J.P., Mouritsen, K.N., Poulin, R., Reise, K., Stouffer, D.B., Thieltges,

D.W., Williams, R.J., Zander, C.D., 2013. Parasites affect food web structure primarily through increased diversity and complexity. . PLoS Biol. 11, e1001579.

Dunz, A.R., Schliewen, U.K., 2013. Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as *"Tilapia."* . Mol. Phylogenet. Evol. 68, 64–80.

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.

Fannes, W., Vanhove, M.P.M., Huyse, T., Paladini, G., 2015. A scanning electron microscope technique for studying the sclerites of *Cichlidogyrus*. . Parasitol. Res. 114, 2031–2034.

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.

Forbes, A.A., Powell, T.H.Q., Stelinski, L.L., Smith, J.J., Feder, J.L., 2009. Sequential sympatric speciation across trophic levels. . Science 323, 776–779.

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

García-Vásquez, A., Shinn, A.P., Bron, J.E., 2012. Development of a light microscopy stain

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 for the sclerites of *Gyrodactylus* von Nordmann, 1832 (Monogenea) and related genera. . Parasitol. Res. 110, 1639–1648.

Geraerts, M., Muterezi Bukinga, F., Vanhove, M.P.M., Pariselle, A., Chocha Manda, A., Vreven, E.J., 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, 187.

Gillardin, C., Vanhove, M.P.M., Pariselle, A., Huyse, T., Volckaert, F.A.M., 2012. Ancyrocephalidae (Monogenea) of Lake Tanganyika: II: description of the first *Cichlidogyrus* spp. parasites from Tropheini fish hosts (Teleostei, Cichlidae). . Parasitol. Res. 110, 305–313.

Glor, R.E., 2010. Phylogenetic insights on adaptive radiation. . Annu. Rev. Ecol. Evol. Syst. 41, 251–270.

Gobbin, T.P., Vanhove, M.P.M., Seehausen, O., Maan, M.E., 2020. Microhabitat distributions and species interactions of ectoparasites on the gills of cichlid fish in lake Victoria, Tanzania. . Int. J. Parasitol.

Gobbin, T.P., Vanhove, M.P.M., Seehausen, O., Maan, M.E., Pariselle, A., 2021. Four new species of *Cichlidogyrus* (Platyhelminthes, Monogenea, Dactylogyridae) from Lake Victoria haplochromine cichlid fishes, with the redescription of *C. bifurcatus* and *C. longipenis* [preprint]. . bioRxiv 2021.01.29.428376.

Goloboff, P.A., Mattoni, C.I., Quinteros, A.S., 2006. Continuous characters analyzed as such. . Cladistics 22, 589–601.

Gómez, A., Nichols, E., 2013. Neglected wild life: parasitic biodiversity as a conservation

target. . Int. J. Parasitol. Parasites Wildl. 2, 222–227.

Goolsby, E.W., Bruggeman, J., Ané, C., 2017. Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within--species variation. . Methods Ecol. Evol. 8, 22–27.

Goren, M., 2014. *Tristramella sacra*. . IUCN Red List Threat. Species 2014 e.T61372A19010617.

Grant, P.R., 1999. Ecology and evolution of Darwin's finches, 2nd ed. . Princeton Science Library, Princeton University Press, Princeton, USA.

Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. . Bioinformatics 32, 2847–2849.

Guerrero, A.M., Tye, A., 2009. Darwin's finches as seed predators and dispersers. . Wilson J. Ornithol. 121, 752–764.

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.

Hansen, T.F., Martins, E.P., 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. . Evolution 50, 1404–1417.

Harmon, A., Littlewood, D.T.J., Wood, C.L., 2019. Parasites lost: using natural history collections to track disease change across deep time. . Front. Ecol. Environ. 17, 157–166.

Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., Purvis, A.,

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

Ricklefs, R.E., Schluter, D., Schulte Ii, J.A., Seehausen, O., Sidlauskas, B.L., Torres-Carvajal, O., Weir, J.T., Mooers, A.Ø., 2010. Early bursts of body size and shape evolution are rare in comparative data. . Evolution 64, 2385–2396.

Hassouna, N., Michot, B., Bachellerie, J.-P., 1984. The compelte nucleotide sequences 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.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Le Vinh, S., 2018a. UFBoot2: Improving the Ultrafast Bootstrap Approximation. . Mol. Biol. Evol. 35, 518–522.

Hoang, D.T., Le Vinh, S., Flouri, T., Stamatakis, A., von Haeseler, A., Minh, B.Q., 2018b. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. . BMC Evol. Biol. 18, 11.

Honaker, J., King, G., Blackwell, M., 2011. Amelia II: a program for missing data. . J. Stat. Softw. 45, 1–45.

Hsu, C., Chang, C., Lin, C., 2003. A practical guide to support vector machines. Technical report. Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan. Available at: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (accessed Mar 6, 2021).

Huang, J., Ling, C.X., 2005. Using AUC and accuracy in evaluating learning algorithms. . IEEE Trans. Knowl. Data Eng. 17, 299–310.

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

Igeh, P.C., 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.

Igeh, P.C., Dos Santos, Q.M., Avenant-Oldewage, A., 2017. Redescription of *Cichlidogyrus philander* (Monogenea, Ancyrocephalidae) using scanning electron microscopy (SEM) and molecular analysis. . Parasite 24, 49.

Ingram, T., Mahler, D.L., 2013. SURFACE: Detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. . Methods Ecol. Evol. 4, 416–425.

Ishiguro, M., Sakamoto, Y., Kitagawa, G., 1997. Bootstrapping log likelihood and EIC, an extension of AIC. . Ann. Inst. Stat. Math. 49, 411–434.

Jabbar, H.K., Khan, R.Z., 2015. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). 163–172.

Jiménez-Garcia, M.I., Vidal-Martínez, V.M., Lopez-Jiménez, S., Jiménez-García, M.I., Vidal-Martínez, V.M., López-Jiménez, S., 2001. Monogeneans in introduced and native cichlids in México: Evidence for transfer. . J. Parasitol. 87, 907.

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

Jorissen, M.W.P., Pariselle, A., Huyse, T., Vreven, E.J., Snoeks, J., Decru, E., Kusters, T., Wamuini Lunkayilakio, S., Muterezi Bukinga, F., Artois, T., Vanhove, M.P.M., 2018a. Six new dactylogyrid species (Platyhelminthes, Monogenea) from the gills of cichlids (Teleostei, Cichliformes) from the Lower Congo Basin. . Parasite 25, 64.

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

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

Jorissen, M.W.P., Vanhove, M.P.M., Snoeks, J., Vreven, E.J., Šimková, A., Volckaert, F.A.M., Wamuini Lunkayilakio, S., Chocha Manda, A., Kapepula Kasembele, G., Muterezi Bukinga, F., Artois, T., Huyse, T., (in prep.). Parasite gene flow along the Congo Basin: a phylogenetic approach. .

Kalafi, E.Y., Tan, W.B., Town, C., Dhillon, S.K., 2016. Automated identification of monogeneans using digital image processing and K-nearest neighbour approaches. . BMC Bioinform. 17, 511.

Kaltz, O., Shykoff, J.A., 1998. Local adaptation in host–parasite systems. . Heredity 81, 361–370.

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

Karvonen, A., Seehausen, O., 2012. The role of parasitism in adaptive radiations-when might parasites promote and when might they constrain ecological speciation? . Int. J. Ecol. 2012.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. . Mol. Biol. Evol. 30, 772–780.

Kearn, G., Whittington, I., 2015. Sperm transfer in monogenean (platyhelminth) parasites. .

Acta Parasitol. 60, 567–600.

Kearn, G.C., 1968. The development of the adhesive organs of some diplectanid, tetraonchid and dactylogyrid gill parasites (Monogenea). . Parasitology 58, 149–163.

Khabbazian, M., Kriebel, R., Rohe, K., Ané, C., 2016. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. . Methods Ecol. Evol. 7, 811–824.

Kitagawa, G., Konishi, S., 2010. Bias and variance reduction techniques for bootstrap information criteria. . Ann. Inst. Stat. Math. 62, 209–234.

Klingenberg, C.P., 2010. Evolution and development of shape: integrating quantitative approaches. . Nat. Rev. Genet. 11, 623–635.

Kmentová, N., Gelnar, M., Koblmüller, S., Vanhove, M.P.M., 2016a. Deep-water parasite diversity in Lake Tanganyika: Description of two new monogenean species from benthopelagic cichlid fishes. . Parasit. Vectors 9, 1–9.

Kmentová, N., Gelnar, M., Koblmüller, S., Vanhove, M.P.M., 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 2016, e1629.

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

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, M.P.M., 2020. Failure to diverge in African Great Lakes: the case of

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 *Dolicirroplectanum lacustre* gen. nov. comb. nov. (Monogenea, Diplectanidae) infecting latid hosts. . J. Great Lakes Res. 46, 1113–1130.

Kmentová, N., Van Steenberge, M., Raeymaekers, J.A.M., Koblmüller, S., Hablützel, P.I., Bukinga, F.M., N'sibula, T.M., Mulungula, P.M., Nzigidahera, B., Ntakimazi, G., Gelnar, M., Vanhove, M.P.M., 2018. Monogenean parasites of sardines in Lake Tanganyika: diversity, origin and intraspecific variability. . Contrib. Zool. 87, 105–132.

Landis, M.J., Freyman, W.A., Baldwin, B.G., 2018. Retracing the Hawaiian silversword radiation despite phylogenetic, biogeographic, and paleogeographic uncertainty. . Evolution 72, 2343–2359.

Legendre, P., Lapointe, F.-J., 2004. Assessing congruence among distance matrices: single-malt scotch whiskies revisited. . Aust. N. Z. J. Stat. 46, 615–629.

Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M.L.J., Tedesco, P.A., 2008. Global diversity of fish (Pisces) in freshwater. . Hydrobiologia 595, 545–567.

Littlewood, D.T.J., 2006. The evolution of parasitism in flatworms. . In: Maule, A.G., Marks, N.J. (Eds.), Parasitic flatworms: molecular biology, biochemistry, immunology, and physiology. CAB International, Wallingford, UK, pp. 1–31.

Littlewood, D.T.J., Rohde, K., Clough, K.A., 1997. Parasite speciation within or between host species? - Phylogenetic evidence from site-specific polystome monogeneans. . Int. J. Parasitol. 27, 1289–1297.

Llewellyn, J., 1963. Larvae and larval development of monogeneans. . In: Dawes, B. (Ed.), Advances in parasitology. Academic Press, London and New York, pp. 287–326.

Lockyer, A.E., Olson, P.D., Østergaard, P., Rollinson, D., Johnston, D.A., Attwood, S.W.,

Southgate, V.R., Horak, P., Snyder, S.D., Le, T.H., Agatsuma, T., McManus, D.P., Carmichael, A.C., Naem, S., Littlewood, D.T.J., 2003. The phylogeny of the Schistosomatidae based on three genes with emphasis on the interrelationships of *Schistosoma* Weinland, 1858. . Parasitology 126, 203–224.

Losos, J.B., 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. . Ecol. Lett. 11, 995–1003.

Losos, J.B., 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism: American society of naturalists E. O. Wilson award address. . Am. Nat. 175, 623–639.

Mahler, D.L., Ingram, T., Revell, L.J., Losos, J.B., 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. . Science 341, 292–295.

Marchiori, N.C., Pariselle, A., Pereira, J., Agnèse, J.F., Durand, J.D., Vanhove, M.P.M., 2015. A comparative study of *Ligophorus uruguayense* and *L. saladensis* (Monogenea: Ancyrocephalidae) from *Mugil liza* (Teleostei: Mugilidae) in southern Brazil. . Folia Parasitol. 62, 024.

Matejusová, I., Gelnar, M., McBeath, A.J.A., Collins, C.M., Cunningham, C.O., 2001. Molecular markers for gyrodactylids (Gyrodactylidae: Monogenea) from five fish families (Teleostei). . Int. J. Parasitol. 31, 738–745.

Matsuo, K., Hirose, Y., Johnson, N.F., 2014. A taxonomic issue of two species of *Trissolcus* (Hymenoptera: Platygastridae) parasitic on eggs of the brown-winged green bug, *Plautia stali* (Hemiptera: Pentatomidae): Resurrection of *T. plautiae*, a cryptic species of *T. japonicus</. . Appl. Entomol. Zool. 49, 385–394.*

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

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 *Parasitol. 19, 400–404.*

McGee, M.D., Neches, R.Y., Seehausen, O., 2016. Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. . Mol. Ecol. 25, 260–268.

Mendlová, M., Desdevises, Y., Civáňová, K., Pariselle, A., Šimková, A., 2012. Monogeneans of West African cichlid fish: evolution and cophylogenetic interactions. . PLoS One 7.

Mendlová, M., Šimková, A., 2014. Evolution of host specificity in monogeneans parasitizing African cichlid fish. . Parasit. Vectors 7, 69.

Messu Mandeng, F.D., Bilong Bilong, C.F., Pariselle, A., Vanhove, M.P.M., Bitja Nyom, A.R., 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, 1–12.

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. Vienna, Austria. Available at: https://cran.r-project.org/web/packages/e1071/e1071.pdf (accessed Oct 27, 2020).

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. . 2010 Gatew. Comput. Environ. Work. GCE 2010.

Mundry, R., 2014. Statistical issues and assumptions of phylogenetic generalized least squares. . In: Garamszegi, L.Z. (Ed.), Modern phylogenetic comparative methods and their application in evolutionary biology. Springer, Berlin/Heidelberg, Germany, pp.

131–153.

Murtagh, F., Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? . J. Classif. 31, 274–295.

Muterezi Bukinga, F., Vanhove, M.P.M., 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.

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

O'Meara, B.C., Ané, C., Sanderson, M.J., Wainwright, P.C., 2006. Testing for different rates of continuous trait evolution using likelihood. . Evolution 60, 922.

Pagel, M., 1999. Inferring the historical patterns of biological evolution. . Nature 401, 877–884.

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 17, 107–111.

Paperna, I., 1979. Monogenea of inland water fish in Africa. . Ann. Mus. R. Afr. Cent. Tervuren Belg. Ser. IN-8 Sci. Zool. 226, 1–131.

Paperna, I., Thurston, J.P., 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.

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

Parins-Fukuchi, C., 2018. Use of continuous traits can improve morphological phylogenetics. . Syst. Biol. 67, 328–339.

Pariselle, A., Bitja Nyom, A.R., Bilong Bilong, C.F., 2014. Four new species of Cichlidogyrus (Monogenea, Ancyrocephalidae) from Sarotherodon mvogoi and Tylochromis sudanensis (Teleostei, Cichlidae) in Cameroon. . Zootaxa 3881, 258–266.

Pariselle, A., Bitja Nyom, A.R., Bilong Bilong, C.F., Bilong, C.F., 2013. Checklist of the ancyrocephalids (Monogenea) parasitizing Tilapia species in Cameroon, with the description of three new species. . Zootaxa 3599, 78–86.

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.

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 . . J. Helminthol. Soc. Washingt. 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.

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.

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.

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.

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

Pariselle, A., Morand, S., Deveney, M.R., 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. (Eds.), Taxonomie, écologie et évolution des métazoaires parasites: Livre-hommage à Louis Euzet. Presses Universitaires de Perpignan, Perpignan, France, pp. 147–166.

Pariselle, A., Muterezi Bukinga, F., Van Steenberge, M., Vanhove, M.P.M., 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.

Pariselle, A., Van Steenberge, M., Snoeks, J., Volckaert, F.A.M., Huyse, T., Vanhove, M.P.M., 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.

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

Pennell, M.W., Eastman, J.M., Slater, G.J., Brown, J.W., Uyeda, J.C., Fitzjohn, R.G., Alfaro, M.E., Harmon, L.J., 2014. Geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. . Bioinformatics 30, 2216–2218.

Poulin, R., 2014. Parasite biodiversity revisited: frontiers and constraints. . Int. J. Parasitol. 44, 581–589.

Poulin, R., Hay, E., Jorge, F., 2019. Taxonomic and geographic bias in the genetic study of helminth parasites. . Int. J. Parasitol. 49, 429–435.

Poulin, R., Krasnov, B.R., Mouillot, D., 2011. Host specificity in phylogenetic and geographic space. . Trends Parasitol. 27, 355–361.

Poulin, R., Presswell, B., Jorge, F., 2020. The state of fish parasite discovery and taxonomy: a critical assessment and a look forward. . Int. J. Parasitol. 50, 733–742.

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.

Price, C.E., Kirk, R.G., 1967. First description of a monogenetic trematod from Malawi. . Rev. Zool. Bot. Afr. 76, 137–144.

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

Puttick, M.N., 2018. Mixed evidence for early bursts of morphological evolution in extant clades. . J. Evol. Biol. 31, 502–515.

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

for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/ (accessed Jul 5, 2021).

Rahmouni, C., Van Steenberge, M., Vanhove, M.P.M., Šimková, A., 2020. Intraspecific morphological variation in Cichlidogyrus (Monogenea) parasitizing two cichlid hosts from Lake Tanganyika exhibiting different dispersal capacities. . Hydrobiologia in press.

Rahmouni, C., Vanhove, M.P.M., 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 . . Parasit. Vectors 10, 591.

Rahmouni, C., Vanhove, M.P.M., 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.

Ramasamy, P., Ramalingam, K., Hanna, R.E.B., Halton, D.W., 1985. Microhabitats of gill parasites (Monogenea and copepoda) of teleosts (Scomberoides spp.). . Int. J. Parasitol. 15, 385–397.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. . Syst. Biol. 67, 901–904.

Řehulková, E., Kičinjaová, M.L., Mahmoud, Z.N., 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.

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., 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.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). . Methods Ecol. Evol. 3, 217–223.

Rodríguez-González, A., Sarabeev, V., Balbuena, J.A., 2017. Evolutionary morphology in shape and size of haptoral anchors in 14 Ligophorus spp. (Monogenea: Dactylogyridae). . PLoS One 12, e0178367.

Rohde, K., 1976. Monogenean gill parasites of Scomberomorus commersoni Lacépède and other mackerel on the Australian East Coast. . Zeitschrift für Parasitenkd. 51, 49–69.

Ronco, F., Büscher, H.H., 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.

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

Salzburger, W., 2018. Understanding explosive diversification through cichlid fish genomics. . Nat. Rev. Genet. 19, 705–717.

Salzburger, W., van Bocxlaer, B., Cohen, A.S., 2014. Ecology and evolution of the African Great Lakes and their faunas. . Annu. Rev. Ecol. Evol. Syst. 45, 519–545.

Schedel, F.D.B., Musilova, Z., Schliewen, U.K., 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, 1–25.

Seehausen, O., 2006. African cichlid fish: a model system in adaptive radiation research. . Proc. R. Soc. B Biol. Sci. 273, 1987–1998.

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.

Šimková, A., Plaisance, L., Matějusová, I., Morand, S., Verneau, O., 2003. Phylogenetic relationships of the Dactylogyridae Bychowsky, 1933 (Monogenea: Dactylogyridea): The need for the systematic revision of the Ancyrocephalinae Bychowsky, 1937. . Syst. Parasitol. 54, 1–11.

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

Šimková, A., Serbielle, C., Pariselle, A., Vanhove, M.P.M., Morand, S., 2013. Speciation in Thaparocleidus (Monogenea: Dactylogyridae) parasitizing Asian pangasiid catfishes. . Biomed Res. Int. 2013, 353956.

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

Takahashi, T., Koblmüller, S., 2011. The adaptive radiation of cichlid fish in Lake Tanganyika: a morphological perspective. . Int. J. Evol. Biol. 2011, 1–14.

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

Thompson, R.C.A., 2013. Parasite zoonoses and wildlife: One Health, spillover and human

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 *activity. . Int. J. Parasitol. 43, 1079–1088.*

Thurston, J.P., 1970. The incidence of Monogenea and parasitic Crustacea on the gills of fish in Uganda. . Rev. Zool. Bot. Afr. 82, 111–130.

Toju, H., Yamamichi, M., Guimarães, P.R., Olesen, J.M., Mougi, A., Yoshida, T., Thompson, J.N., 2017. Species-rich networks and eco-evolutionary synthesis at the metacommunity level. . Nat. Ecol. Evol. 1, 24.

Turner, G.F., Seehausen, O., Knight, M.E., Allender, C.J., Robinson, R.L., 2001. How many species of cichlid fishes are there in African lakes? . Mol. Ecol. 10, 793–806.

Uyeda, J.C., Harmon, L.J., 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. . Syst. Biol. 63, 902–918.

Van Steenberge, M., Pariselle, A., Huyse, T., Volckaert, F.A.M., Snoeks, J., Vanhove, M.P.M., 2015. Morphology, molecules, and monogenean parasites: an example of an integrative approach to cichlid biodiversity. . PLoS One 10, e0124474.

Van Steenberge, M., Vanhove, M.P.M., Muzumani Risasi, D., Mulimbwa N'Sibula, T., Muterezi Bukinga, F., Pariselle, A., Gillardin, C., Vreven, E.J., Raeymaekers, J.A.M., Huyse, T., Volckaert, F.A.M., Nshombo Muderhwa, V., Snoeks, J., 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.

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

1–16.

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

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

Vanhove, M.P.M., Van Steenberge, M., Dessein, S., Volckaert, F.A.M., Snoeks, J., Huyse, T., Pariselle, A., 2013. Biogeographical implications of Zambezian Cichlidogyrus species (Platyhelminthes: Monogenea: Ancyrocephalidae) parasitizing Congolian cichlids. . Zootaxa 3608, 398–400.

Vanhove, M.P.M., Volckaert, F.A.M., 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.

Vignon, M., Pariselle, A., Vanhove, M.P.M., 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.

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.

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

Wäldchen, J., Mäder, P., 2018. Machine learning for image based species identification. . Methods Ecol. Evol. 9, 2216–2225.

Webster, J.P., Gower, C.M., Knowles, S.C.L., Molyneux, D.H., Fenton, A., 2016. One health - an ecological and evolutionary framework for tackling Neglected Zoonotic Diseases. . Evol. Appl. 9, 313–333.

Weiblen, G.D., Bush, G.L., 2002. Speciation in fig pollinators and parasites. . Mol. Ecol. 11, 1573–1578.

Wickham, H., 2016. ggplot2: Elegant graphics for data analysis, 2nd ed. , Use R! Springer, New York, USA.

Wu, X.Y., Zhu, X.Q., Xie, M.Q., Li, A.X., 2006. The radiation of Haliotrema (Monogenea: Dactylogyridae: Ancyrocephalinae): molecular evidence and explanation inferred from LSU rDNA sequences. . Parasitology 132, 659–668.

Wu, X.Y., Zhu, X.Q., Xie, M.Q., Li, A.X., 2007. The evaluation for generic-level monophyly of Ancyrocephalinae (Monogenea, Dactylogyridae) using ribosomal DNA sequence data. . Mol. Phylogenet. Evol. 44, 530–544.

Yu, G., Lam, T.T.-Y., Zhu, H., Guan, Y., 2018. Two methods for mapping and visualizing associated data on phylogeny using ggtree. . Mol. Biol. Evol. 35, 3041–3043.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-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.

Zhi, T., Xu, X., Chen, J., Zheng, Y., Zhang, S., Peng, J., Brown, C.L., 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.

Zischke, M.T., Litherland, L., Tilyard, B.R., Stratford, N.J., Jones, E.L., 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.

LIST OF 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 the IQ-TREE ModelFinder manual (Kalyaanamoorthy et al., 2017).*

LIST OF 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. *Bayesian inference phylogram of monogenean flatworms belonging to Cichlidogyrus and Scutogyrus (Platyhelminthes, Monogenea, Dactylogyridae) (a) and host range matrix of the species by tribe or subfamily of cichlid hosts and family of non-cichlid hosts (b) (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 have been generated for this study. Node support values: Bayesian posterior probabilities (PP) followed by ultrafast bootstrap values (UFBoot) 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, UFBoot < 95, SH-aLRT < 80, MPBoot < 95); black dots, internal nodes with strong support across all support values; highlighted clades, monophyletic 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 cichlids 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, pelmatolapiine 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. *Cluster analysis of morphometric measurements of attachment and reproductive organs of species of Cichlidogyrus. Clusters were detected using the ward.d2 clustering algorithm. Proxy variables (in bold) included in the reduced dataset used for in downstream analyses to minimise model overfitting were selected on the amount of missing data to maximise data input. For abbreviation of measurements, see Fig. 1.*

Figure 4. *Results of phylogenetic comparative analysis applied to one set of 100 randomly sampled tree topologies from the Bayesian post–burn in phase. Fitted continuous models include Brownian Motion (BM), Ornstein-Uhlenbeck (OU), Early-Burst (EB), Late-Burst (LB), and multi-rate Brownian Motion (BMM) models simulating a random walk (H_1), stabilising selection (H_2), adaptive radiation with a decelerating (H_3) and accelerating (H_4) divergence of characters, and divergent BM regimes in parts of the phylogeny (H_5). (a) Multivariate analysis of continuous characters before and after variable reduction resulting from cluster analysis (Fig. 3). Variable reduction through cluster analysis (Fig. 3) resulted in improved multivariate models fits for attachment organ but reduced model fits for the reproductive organ measurements. Model fits were assessed through the Extended Information Criterion (EIC) for multivariate analysis. (b) All models assuming a single evolutionary regime across the phylogeny performed similarly (LB performed worse). However, the BMM model suggesting a different regime for the EAR clade outperformed the latter. (c) Univariate PCMs for discrete characters including 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). HC and VBS show a strong phylogenetic signal, AS and IS shows no detectable phylogenetic signal. Model performance*

is assessed through difference in the sample size–corrected Akaike Information Criterion (ΔAICc) compared to a white noise model under the assumption of a star-like phylogeny (absolute phylogenetic independence).

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 Fig. 2a with only one specimen per species. Numbers on nodes refer to clades in Fig. 2a. Hook 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-, and tribe-specific, or 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 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 little unifying or contrasting morphological patterns were detected between groups, see "Characterisation of species groups"). Arrows indicate key features of each species group (blue) and subgroup of the EAR group (yellow) (see Appendix 3), 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) illustrating 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 bar (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 trees in Fig. 2a with only one specimen per species. Numbers on nodes refer to clades in Fig. 2a. The sizes of the outer root of the dorsal anchor (DAc), the first (U1) and third hook pair (U3), and the auricles of the dorsal bar (DBh) display a limited phylogenetic information on their own as extreme values (dark blue or dark red) are spread out across the phylogenetic tree. Node states inferred here under the assumption of neutral evolution do likely not reflect real evolutionary history as a multi-rate model showed more support than single-rate models for PCM (Fig. 4).*

Figure 9. *Results of statistical classification of specimens belonging to Cichlidogyrus into proposed species groups (taxonomic classification) using support vector machine (SVM) algorithm and proxy variables (reduced dataset) of attachment and reproductive organ metrics (Fig. 3): (a) Confusion matrix including percentage of measured specimens classified in each group (missing data imputed 100 fold). (b) Variable importance assessed through z-scores of area under the operator receiver curve (AUROC) after permuting each variable 100-fold. The lower z-score the more important the variable. For abbreviation of measurements, see Fig. 1. Species groups were proposed based on phylogenetic analysis (Fig. 2) and taxonomic literature survey.*

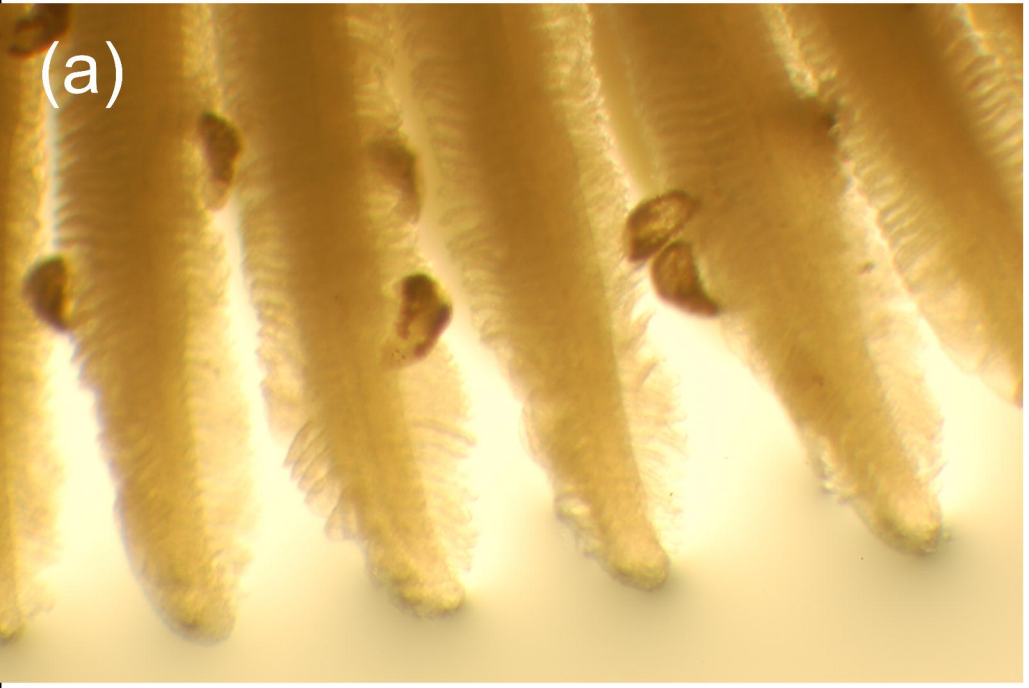
LIST OF APPENDICES (IN PRINT)

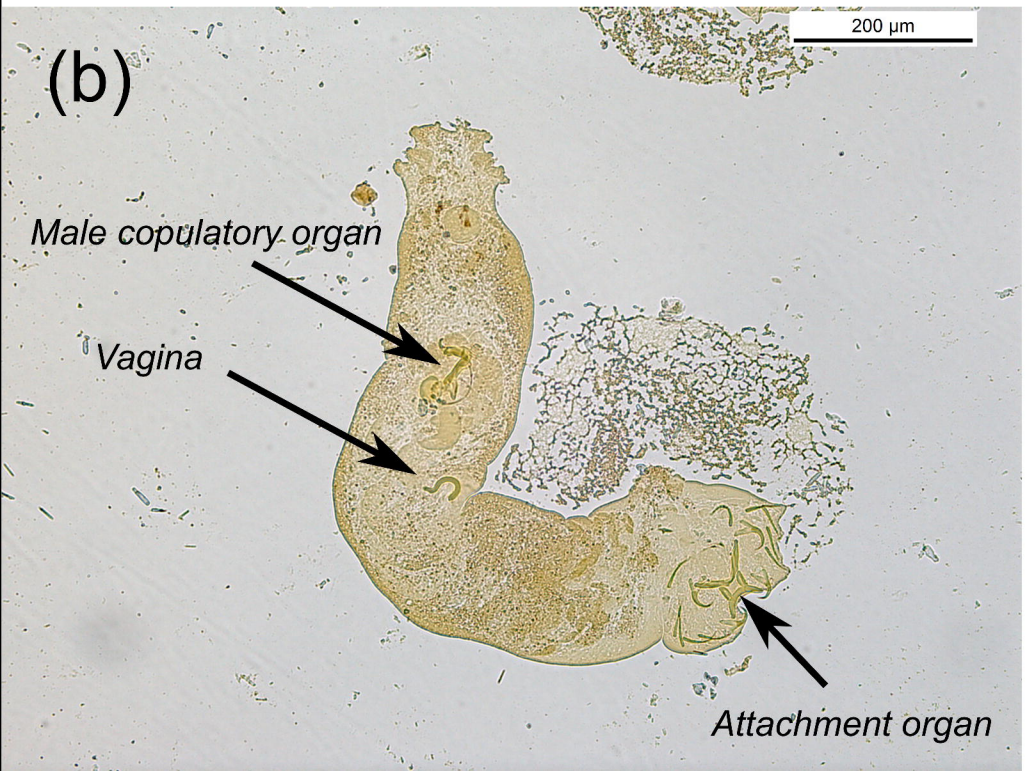
Appendix 1. *Sampling localities and dates of specimens belonging to Cichlidogyrus and Scutogyrus collected for this study including a reference for the samples previously included in taxonomic studies. Freshwater ecoregions are assigned according to Thieme et al. (2005).*

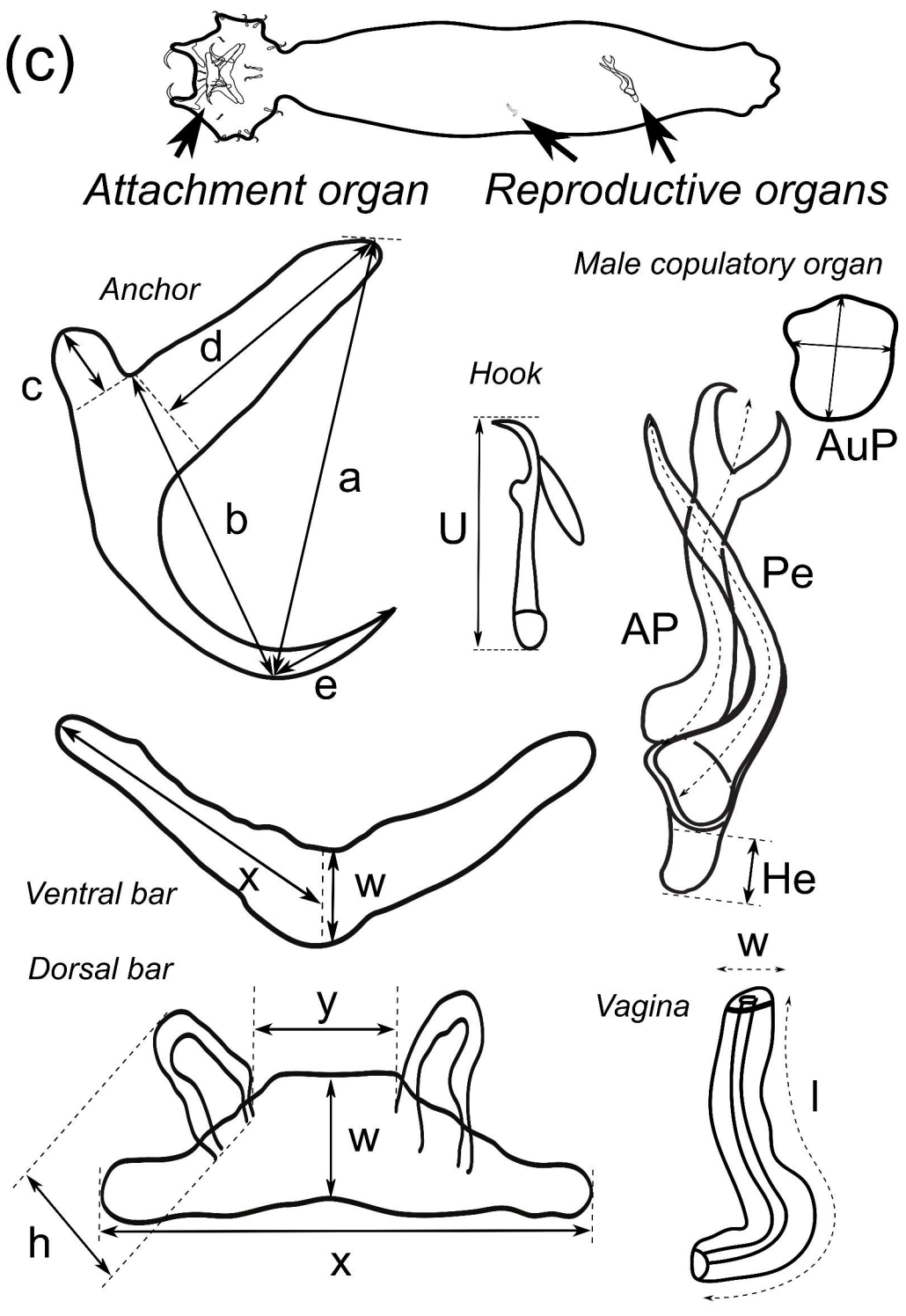
Appendix 2. *Specimen data of cichlid parasites of the genera Cichlidogyrus and Scutogyrus used for phylogenetic analyses including host species, GenBank accession numbers, locality by country, and reference. Voucher/isolate ID and accession numbers in italics indicate specimens not included in subset trees used for phylogenetic comparative methods.*

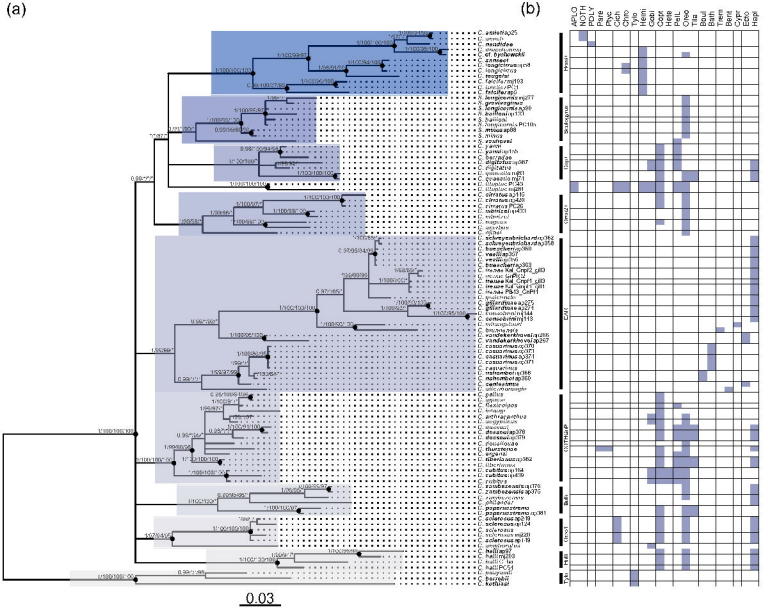
Appendix 3. *Species groups of Cichlidogyrus with key features, species included, species potentially included, and the respective host ranges reported in the taxonomic literature including host species of candidate species.*

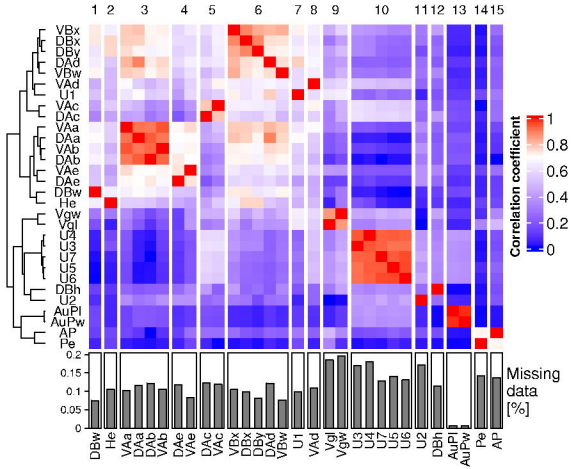
Appendix 4. *Statistical classification of species of Cichlidogyrus (incl. Scutogyrus) in species groups resulting from phylogenetic analyses (Fig. 2a) using support vector machines and morphometric data. Values refer to number of specimens classified in each category on average for 100 imputed datasets. These datasets were computed to address missing data for some specimens. For proposed classification of species according to taxonomic literature review, see Appendix 3. For species in bold, we proposed no affiliation with any of the proposed species groups.*

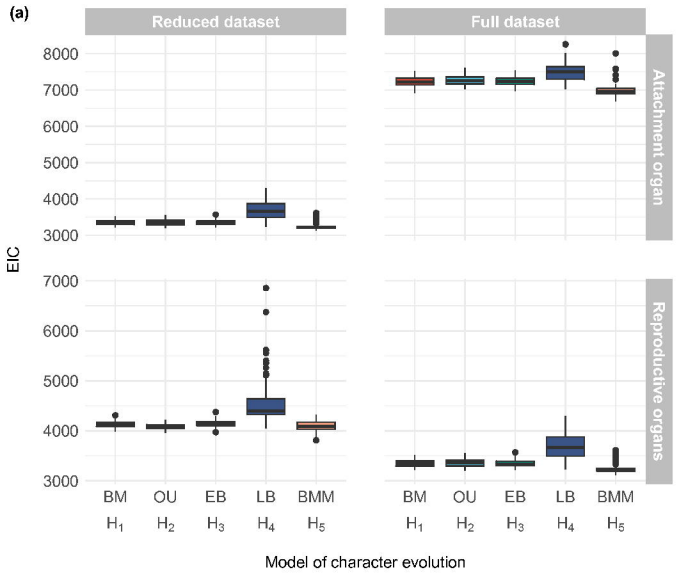

(a)

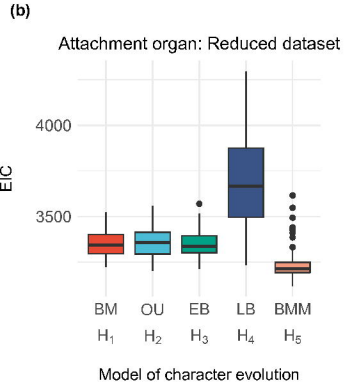

(b)
200 µm
Male copulatory organ
Vagina
Attachment organ

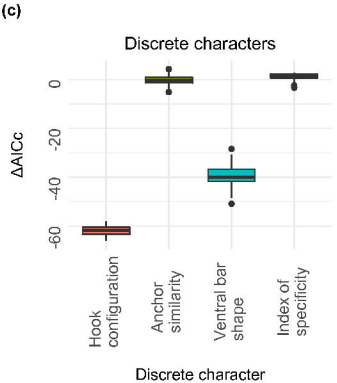

(c)
Attachment organ
Reproductive organs
Anchor
c
d
b
a
e
Hook
U
Male copulatory organ
AuP
Pe
AP
He
Ventral bar
x
w
Dorsal bar
y
w
h
x
Vagina
w
l

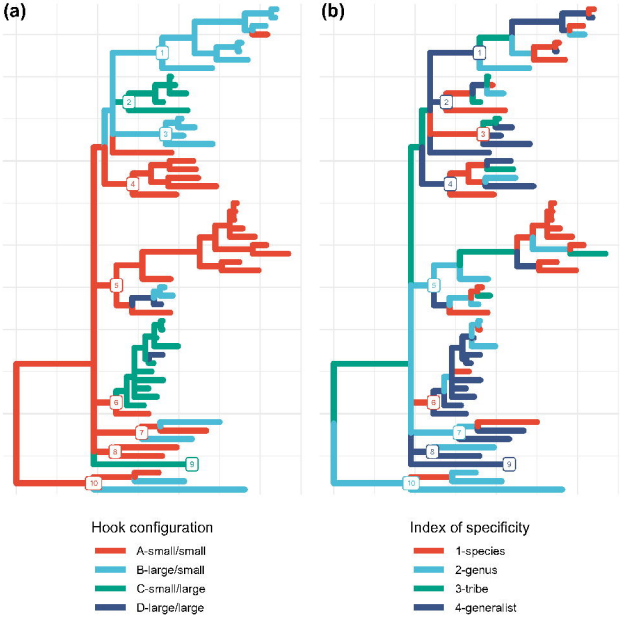

(a)
(b)
APLO
NOTH
POLY
Pare
Pyx
Rich
Chlo
Tylo
Hemi
Gcal
Ccal
Hete
Pelt
Ttyp
Tiza
Bual
Bath
Tiem
Bent
Cyr
Ebo
Hagi
0.03

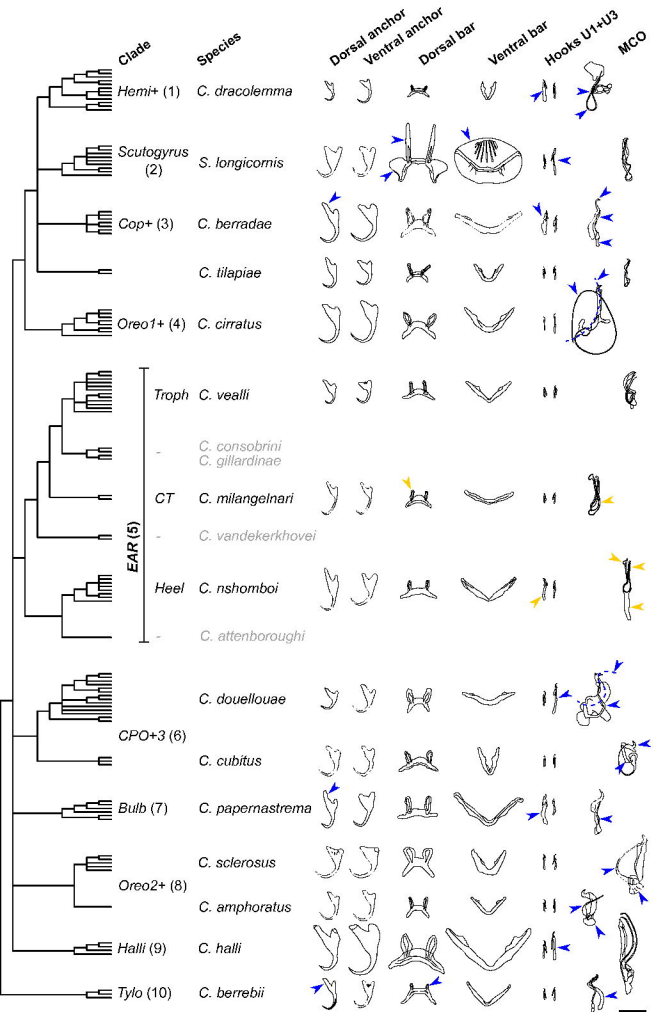

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
VBx
DBx
DBy
DAd
VBw
VAd
U1
VAc
DAc
VAa
DAa
VAb
DAb
VAe
DAe
DBw
He
Vgw
Vgl
U4
U3
U7
U5
U6
DBh
U2
AuPl
AuPw
AP
Pe
Correlation coefficient
1
0.8
0.6
0.4
0.2
0
Missing data [%]
0.2
0.15
0.1
0.05
0
DBw He VAa DAa DAb VAb DAe VAe DAc VAc VBx DBx DBy DAd VBw U1 VAd Vgl Vgw U3 U4 U7 U5 U6 U2 DBh AuPl AuPw Pe AP

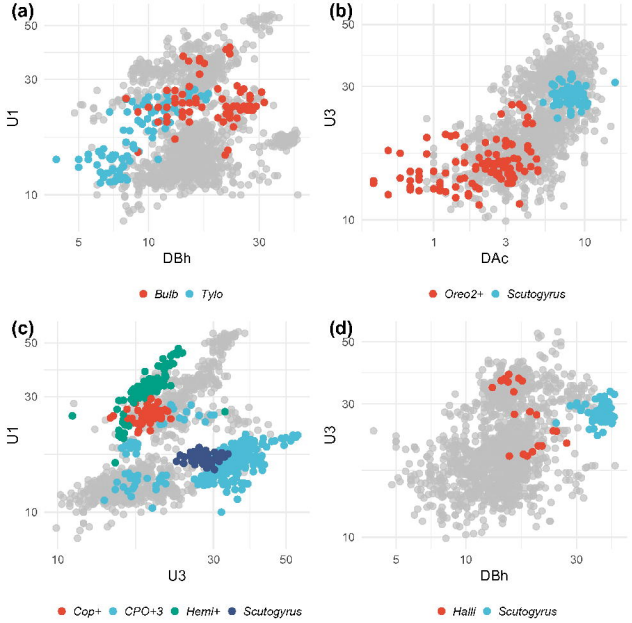

(a)
Reduced dataset
Full dataset
Attachment organ
Reproductive organs
EIC
8000
7000
6000
5000
4000
3000
BM
OU
EB
LB
BMM
H1
H2
H3
H4
H5
Model of character evolution

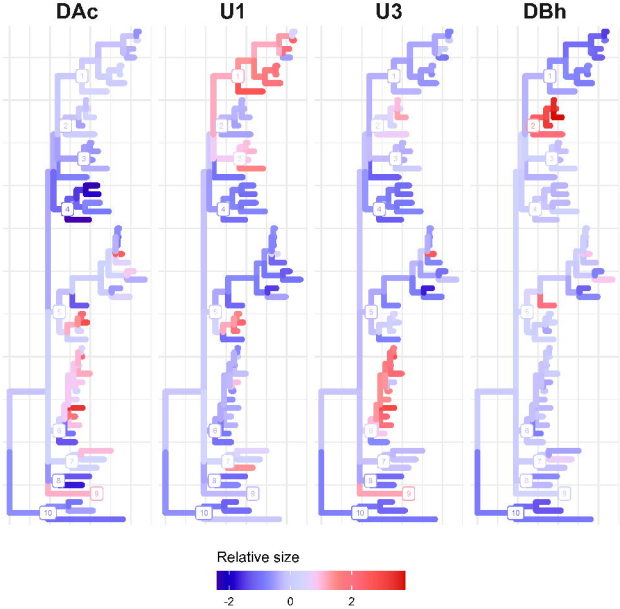

(b)
Attachment organ: Reduced dataset
EIC
4000
3500
BM
OU
EB
LB
BMM
H1
H2
H3
H4
H5
Model of character evolution

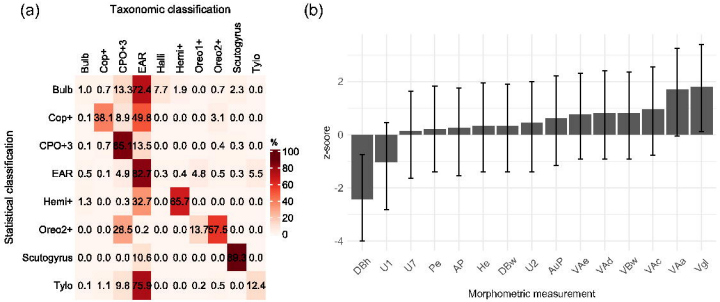

(c)
Discrete characters
ΔAICc
0
-20
-40
-60
Hook configuration
Anchor similarity
Ventral bar shape
Index of specificity
Discrete character


(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


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)
U1
DBh
Bulb
Tylo
(b)
U3
DAc
Oreo2+
Scutogyrus
(c)
U1
U3
Cop+
CPO+3
Hemi+
Scutogyrus
(d)
U3
DBh
Halli
Scutogyrus


DAc
U1
U3
DBh
Relative size
-2
0
2


(a)
Taxonomic classification
Statistical classification
Bulb
Cop+
CPO+3
EAR
Halli
Hemi+
Oreo1+
Oreo2+
Scutogyrus
Tylo
Bulb 1.0 0.7 13.3 72.4 7.7 1.9 0.0 0.7 2.3 0.0
Cop+ 0.1 38.1 8.9 49.8 0.0 0.0 0.0 3.1 0.0 0.0
CPO+3 0.1 0.7 85.1 13.5 0.0 0.0 0.0 0.4 0.3 0.0
EAR 0.5 0.1 4.9 92.1 0.3 0.4 4.8 0.5 0.3 5.5
Hemi+ 1.3 0.0 0.3 32.7 0.0 65.7 0.0 0.0 0.0 0.0
Oreo2+ 0.0 0.0 28.5 0.2 0.0 0.0 13.7 57.5 0.0 0.0
Scutogyrus 0.0 0.0 10.6 0.0 0.0 0.0 0.0 0.0 89.3 0.0
Tylo 0.1 1.1 9.8 75.9 0.0 0.0 0.2 0.5 0.0 12.4
%
100
80
60
40
20
0
(b)
z-score
2
0
-2
-4
DBh
U1
U7
Pe
AP
Hb
DBw
U2
AuP
VAe
VAd
VBw
VAc
VAa
Vgl
Morphometric measurement