The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

Biological Journal of the Linnean Society, 2015, ••, ••–••. With 5 figures.

High levels of genetic structure and striking phenotypic variability in a sexually dimorphic suckermouth catfish from the African Highveld

JAKE MORRIS[1†], ANTONIA G. P. FORD[1†], JAROME R. ALI[1], CLAIRE R. PEART[1,2], ROGER BILLS[3] and JULIA J. DAY[1*]

[1]*Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, UK*
[2]*Department of Life Sciences, Natural History Museum, Cromwell Road, London, SW7 5BD, UK*
[3]*South African Institute for Aquatic Biodiversity, Private Bag 1015, Grahamstown, 6140, South Africa*

Received 20 June 2015; revised 29 July 2015; accepted for publication 29 July 2015

Uncovering biological diversity to more accurately understand diversity patterns, and ultimately the processes driving diversification, is important not only from an evolutionary perspective but also a conservation perspective. This is particularly pertinent in Africa's rivers in which overall diversity, as well as how it arose, is poorly understood in comparison with lacustrine environments. Here we investigate population divergence in the sexually dimorphic suckermouth catfish species *Chiloglanis anoterus* (Crass, 1960) from the African Highveld, in which we observe striking variability in exaggerated male caudal fins across its range. As this trait is likely to be indirect evidence for sexual selection by female choice, a mechanism that has been shown to increase species diversity in different taxa, we used an integrated approach to test if current diversity in this species is underestimated. Results based on phylogenetic inference, population genetics and geometric morphometrics indicate that the recognized species *C. anoterus* represents five distinct lineages that may be considered confirmed candidate species. We suggest that diversification in these highland catfish has been facilitated through geographical isolation in upper river catchments, and that sexual selection through female choice has probably driven variation in male caudal fin morphology. In contrast to the relatively large range size of the currently recognized species (*C. anoterus*), our findings highlight highly restricted ranges of the lineages identified here, indicating that these highland habitats may harbour higher levels of endemic diversity than previously thought. © 2015 The Authors. Biological Journal of the Linnean Society published by John Wiley & Sons Ltd on behalf of Linnean Society of London, *Biological Journal of the Linnean Society*, 2015, **00**, 000–000.

ADDITIONAL KEYWORDS: Africa – amplified fragment length polymorphism – biodiversity – geometric morphometrics – mitochondrial DNA – rivers – sexual selection.

INTRODUCTION

Uncovering biological diversity to more accurately understand diversity patterns, and ultimately the processes driving diversification, is important not only from an evolutionary perspective but also a conservation perspective. This is particularly pertinent where a single species, with a broad geographical range and considered to be a species of least concern, may actually represent multiple endemic species with restricted ranges and consequently could be of high concern (Bickford *et al.*, 2007; Funk, Caminer & Ron, 2012). This problem is exacerbated for environments and regions in which diversity is poorly known, such as freshwater habitats (Dudgeon *et al.*, 2006) and nowhere is this more apparent than in Africa (Darwall *et al.*, 2011). Although recent DNA studies focusing on African riverine fish radiations have started to uncover potential cryptic diversity (e.g. Feulner *et al.*, 2007; Goodier *et al.*, 2011; Schwarzer *et al.*, 2011; Day *et al.*, 2013; Schmidt *et al.*, 2014), these environments remain substantially understudied compared with lacustrine systems.

*Corresponding author. E-mail: j.day@ucl.ac.uk
†These authors contributed equally to this work.
The copyright line for this article was changed on 29 September 2015 after original online publication.

Unlike the East African Great Lakes where a complex interplay of speciation mechanisms has been put forward to explain the elevated diversity of the cichlid fish adaptive radiations, including ecological and sexual selection (Wagner, Harmon & Seehausen, 2012), the diversity of African riverine fish is generally attributed to vicariance and geographical isolation at continental (Day *et al.*, 2013), regional (e.g. Collier, Murphy & Espinoza, 2009; Schmidt *et al.*, 2014) and through to local scales (Markert, Schelly & Stiassny, 2010). However, this mechanism alone does not explain why some clades are more species-rich than others, and further investigation of riverine clades is required to better elucidate the generation of ichthyological diversity in these systems.

Evidence of speciation via sexual selection in African rivers is significantly less well documented than in lakes, but has been demonstrated in a sympatric radiation of mormyrids from the Ogooué River system in Gabon where sexual selection on electric signal pulses is thought to be an important driver of species radiation in these fishes (Arnegard *et al.*, 2010). In African catfish sexual dimorphism is considered most striking and pervasive in the species-rich family Mochokidae (Friel & Vigliotta, 2006). Adult males may display elongate anal and/or caudal fins (Roberts, 1989; Friel & Vigliotta, 2006, 2011), and hypertrophied humeral processes (Roberts, 1989), while denser aggregations of head tubercles are reported in *Microsynodontis* (Ng, 2004). As such this is a potentially useful group for investigating the extent to which sexual selection may have promoted diversity. In particular, sexual dimorphism occurs within the genus *Chiloglanis* (suckermouth catfish), in which male caudal fin shape is highly divergent in some species, and differs from the typical homocercal condition displayed by females. Depending on species, exaggerated male caudal fins may include elongation of one or other caudal lobes, whereas others display a trilobate caudal fin in which the middle rays are extended (Friel & Vigliotta, 2011).

Within southern Africa the pennant-tailed suckermouth catfish (*Chiloglanis anoterus*) is the only sexually dimorphic species of the six described from this region (Skelton, 2001). This species was originally described from the Phongolo (Pongola) river (Crass, 1960). However, it occupies a much larger range across the highlands of Swaziland and adjoining regions of South Africa as shown from a recent biodiversity survey by one of us (R. Bills, Supporting Information Fig. S1). Notably this survey also revealed considerable variation of male caudal fin shape across its range. The elongation of the male caudal fin is probably highly disadvantageous considering that these small catfish (< 100 mm) occur principally in the upper catchments of the Highveld (1000–1860 m, Fig. S1), and are restricted to rapids and complex rocky habitats with fast-flowing water (Bills *et al.*, 2004). To our knowledge there are no studies investigating form and function of fin extensions in catfishes, although the presence of an elongate fin extension in male swordtails (*Xiphophorus*), albeit a different shape to the focal group, is known to incur serious fitness costs (Rosenthal *et al.*, 2001) as well as energetic costs as these fins are less hydrodynamically efficient (Basolo & Alcaraz, 2003).

The exaggeration of male fish fins in terms of both size and/or colour has been posited as being driven by sexual as opposed to natural selection. Certainly, the caudal fin extensions ('sword') of *Xiphophorus* species are inferred as indirect evidence for sexual selection, as such exaggerated ornaments are considered detrimental to their survival (Basolo & Alcaraz, 2003). Although several studies demonstrated differences in morphological traits, including caudal fin size, in stream fish populations from fast versus slow water flow (e.g. Pakkasmaa & Piironen, 2001; Cureton & Broughton, 2014), differences are not reported between males or females within the same habitat. Given that *C. anoterus* occur as mixed sexed populations in fast-flowing upland stream habitats, it is highly plausible that the exaggeration of this trait is likely to be under female choice, as opposed to natural selection. We therefore predict that diversity within this lineage may be underestimated, as sexual selection has been shown to increase diversity across different taxonomic groups (e.g. Masta & Maddison, 2002; Ritchie *et al.*, 2005, 2007; Boul *et al.*, 2007; Seldon, Merrill & Tobias, 2008).

The present study is the first to investigate genetic and phenotypic divergence in *C. anoterus*, in which we include sampling from eight rivers across its range that comprise the following three river basins: (1) Incomati, (2) Mbuluzi and (3) Maputo (Fig. S1). As these systems output into the Maputo region of Mozambique (in which regular floods interconnect all these rivers) we test the null hypothesis that there is minimal structure between populations. To investigate the genetic structure of this species, we generated independent genetic markers including mitochondrial cytochrome c oxidase 1 (CO1) and the control region (CR), and nuclear data from amplified fragment length polymorphisms (AFLPs) analysed using phylogenetic inference and population genetic methods. In addition, we quantified and visualized shape differences between sexes and between populations using a geometric morphometric approach.

MATERIAL AND METHODS

Taxon sampling and DNA extraction

A total of 134 genetic samples and 176 voucher specimens of *C. anoterus* were collected from 54

sites across eight rivers in Swaziland and South Africa (Fig. 1) representing three field expeditions (2002, 2003, 2012; Table S1). One sample from the closely related species *C. pretoriae* was selected as an outgroup based on a phylogeny of southern Africa *Chiloglanis* (J. J. Day, unpubl. data.). Fish were sampled using electrofishing and euthanized following capture with clove oil. Voucher specimens were stored in industrial methylated spirits (IMS) and sub-sampled tissues were stored in 96–100% ethanol. Total DNA was extracted from white muscle tissue using a DNeasy Blood and tissue kit

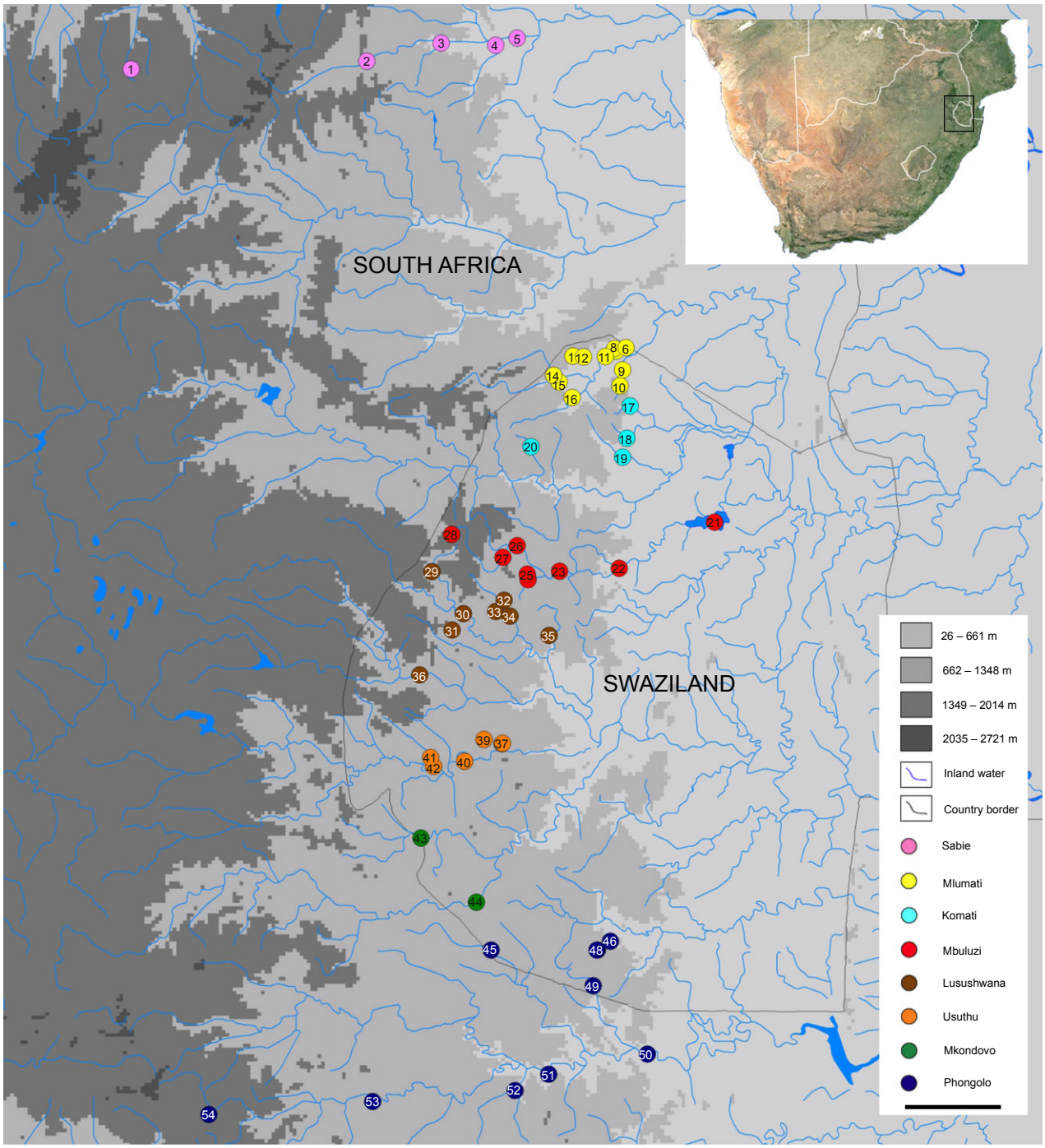


Figure 1. Map of Swaziland and surrounding South Africa detailing collection sites. Site numbers correspond to tissue samples in Table S1. Scale bar = 25 km. Inset: map of southern Africa highlighting the focal region.

(Qiagen). DNA was successfully extracted from most samples. However, extractions were highly variable for some sites (Mkondvo, Komati and Usuthu) leading to fewer individuals being sequenced, while samples collected from the Sabie river in 2002–2003 failed. The poor quality of DNA from these samples led to their exclusion from AFLP analyses. Additional samples from this river were collected during 2012 and were included in subsequent mtDNA analyses, but 2012 samples are not represented in AFLP analyses.

Mitochondrial DNA amplification and sequencing

Two mitochondrial markers, CR (690 bp) and the barcoding gene CO1 (679 bp), were selected for this study based on their performance in resolving both population- and species-level relationships in catfishes (e.g. Montoya-Burgos, 2003; Day *et al.*, 2013; Peart *et al.*, 2014). These markers were sequenced for 80 taxa, although some amplifications/sequences were not successful, leading to 68 CR and 71 CO1 sequences (80 individuals have one or both markers sequenced). Published primers were used for the amplification of CO1 (Ward *et al.*, 2005), while novel primers ChiloDLR (5′-CTTGCCTGGTTTAGGGGT TT-3′) and ChiloDLF (5′-CTCACCCCTAGCTCCCAA AG-3′) were used to amplify the CR, designed using the program primer3 (Steve & Skaletsky, 2000). PCR amplifications were performed in 25-μL volumes with 2 μL of genomic DNA, 18.0 μL double distilled H_2O, 2.5 μL 10× buffer, 1.0 μL $MgCl_2$, 0.25 μL dNTP, 1.0 μL of each primer (10 mM) and 0.25 μL Taq DNA polymerase (Bioline reagents). PCR conditions for CO1 were: denature at 94 °C for 1 min, annealing at 52 °C for 1 min and extension at 72 °C for 90 s for 35 cycles; CR: denature at 94 °C for 1 min, annealing at 66 °C for 30 s and extension at 72 °C for 2 min for 35 cycles. PCR products were cleaned using microclean, and sequenced on an ABI 3730 sequencer (Applied Biosystems).

Genotyping with AFLP markers

AFLP profiles were generated for 120 samples following Vos *et al.* (1995). Seven selective amplifications using restriction enzymes *Eco*RI and *Mse*I (see Table S2) were selected giving clearly defined (98.4%) reproducible markers. Fragments were denatured and separated on an AB3730 automated sequencer. Preliminary presence and absence of fragments was resolved using GeneMapper 3.7 (Applied Biosystems) within a range of 50–500 bp. AFLPscore (Whitlock *et al.*, 2008) was then used to generate AFLP marker profiles for all samples using the Bayesian method implemented by Hadfield, Richardson & Burke (2006). In this way we were able to maximize the number of loci included in the analysis (307 variable markers) while keeping the global Bayesian error rate (5%) below an acceptable threshold. This 5% global Bayesian error rate threshold equates to a mismatch error rate of ≈1%. Three samples (Phongolo 70788-030, Mkondvo 67210-309, Mkondvo 67210-55) were removed from subsequent analyses due to high levels of missing data, resulting in a final dataset of 117 individuals (Dryad DOI: 10.5061/dryad.661jb; Table S1).

Sequence alignment and phylogenetic analysis

Chromatograms of sequence data were checked and edited using ChromasPro v1.33 (Technelysium Pty Ltd) and aligned in the program BioEdit (Hall, 1999) using ClustalW (Thompson, Higgins & Gibson, 1994). All alignments were secondarily checked by visual inspection.

Phylogenetic analysis was performed on the mtDNA dataset (Dryad DOI: 10.5061/dryad.661jb), consisting of concatenated sequence data from the two mtDNA markers CO1 + CR (1369 bp) from 79 *C. anoterus* samples with *C. pretoriae* selected as the outgroup. Partitioning schemes and appropriate models of nucleotide substitution for the mtDNA data were determined using the program PartitionFinder 1.0 (Lanfear *et al.*, 2012) under the Bayesian Information Criterion (BIC) metric for model selection, using a greedy algorithm with four possible subsets defined (each codon position for the protein coding gene, plus the control CR). Phylogenetic inference using Bayesian inference was implemented in the program MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001). Two separate analyses were run for 5 000 000 generations, sampling every 100 generations, with an initial burn-in set to 5000 (chain temperature 0.2, four chains). Support was provided with Bayesian Posterior Probabilities (BPPs) and convergence of Monte Carlo Markov chains was checked using Tracer 1.5 (Rambaut & Drummond, 2009) and ensuring that all effective sample size values were > 200. Sequence divergence for the focal group was generated using Kimura-2-parameter (K2P) distances for the CO1 data set using PAUP* (Swofford, 2003).

Haplotype analysis

Haplotype networks based on concatenated, CR and CO1 sequences were constructed for 53, 59 and 70 *C. anoterus* samples respectively, from all sampled rivers to identify consistently supported haplogroups. Smaller datasets were used here (compared with

phylogenetic analyses) due to some taxa not having sequence data for both markers and/or missing data, which has the potential to reduce resolution in these analyses (Joly, Stevens & van Vuuren, 2007). The sequence data were initially collapsed into haplotypes using the program DnaSP 5.0 (Librado & Rozas, 2009) and sites with missing data were ignored. Network 4.6.1.2 (Fluxus Technology Ltd) was used to build a network of these resultant haplotypes using a median joining algorithm (Bandelt *et al.*, 1995; Bandelt, Forster & Röhl, 1999) with epsilon set to 0. The full median network was then simplified using maximum parsimony to produce the final haplotype network.

Genetic Clustering Analyses

Population admixture and cluster assignment to differentiated genetic clusters (K) was assessed using the Bayesian program Structure 2.3.4 (Pritchard, Stephens & Donnelly, 2000; Falush, Stephens & Pritchard, 2007) for our AFLP data. All four model variations (admixture/correlated allele frequencies; no admixture/uncorrelated allele frequencies; no admixture/correlated allele frequencies; no admixture/uncorrelated allele frequencies) were tested in preliminary runs without a priori population designation. Other parameters used default settings, with the allele frequency distribution parameter λ set to one and the admixture parameter α set to be inferred from the data. Preliminary runs were repeated four times for each value $K = 1$–8, with a burn-in of 10 000 generations followed by a further 10 000. From these runs the mean Ln$P(D)$ and ΔK (Evanno, Regnaut & Goudet, 2005) were calculated to determine the value of K with the highest likelihood. From these initial runs, use of the admixture model and correlated allele frequencies were selected as the most suitable models based on mean Ln$P(D)$. These models were then used to run for the range of K values with the highest likelihood based on mean Ln$P(D)$ and ΔK (Evanno *et al.*, 2005), and subsequent runs were for $K = 2$–7 with burn-in set to 100 000 generations, with a further 1 000 000 afterwards, for five iterations of each run. Output was compiled and averaged using Structure Harvester (Earl & vonHoldt, 2011) and run permutations were clustered using Clumpp 1.1.2 (Jakobsson & Rosenberg, 2007). The clustered data were visualized using the program Distruct 1.1 (Rosenberg, 2003). As an alternative to a model-based approach (i.e. Structure), variance within all datasets (i.e. CO1, CR and AFLPs) was further investigated using principal component analysis (PCA) in the R package adegenet (Jombart, 2008) using the same matrices used for phylogenetic inference.

Analysis of genetic variation

Hierarchical analysis of molecular variance (AMOVA) of AFLP marker profiles was calculated in the EXCEL add-in GenAlEx6.501 (Peakall & Smouse, 2006, 2012) with the number of permutations set to 9999. The AMOVAs were used to measure level of molecular variation within and between samples based on two separate grouping criteria: (1) rivers (collection locality), and (2) the genetic population clusters identified by the Structure analysis in which four separate clusters were analysed: Mlumati + Komati, Phongolo, Mbuluzi (plus an additional six Lusushwana samples that showed majority assignment to this cluster), and a combined cluster of the remaining Lusushwana samples together with Usuthu and Mkondvo samples (see Results). The AMOVA conducted in GenAlEx quantifies genetic differentiation between populations using the statistic Φ_{PT}, which we report here, and is analogous to F_{ST}.

Simple and partial Mantel tests were conducted to test for correlations of geographical and genetic distance (isolation by distance) as well as correlation of morphological and genetic distance whilst controlling for geographical distance (isolation by adaptation). For completeness, all possible comparisons (geographical, morphometric, genetic) were tested for both mtDNA (CR only) and AFLP data. Tests were conducted using the vegan package (Oksanen *et al.*, 2015) in R 3.0.1 (R Core Team, 2013) using the Pearson method, with 5000 permutations to test for significance. Tests used Φ_{PT} (AFLP) and Φ_{ST} (mtDNA) for genetic distances, Procrustes distances for the morphometric data, and straight-line geographical distances. Although only a crude measure of geographical separation, straight-line distances between sampling sites were used here given the lack of detail available for potential river connections in the Highveld. Any attempt to trace distances in the Highveld between rivers (without a more detailed hydrological and geological survey) would result in arbitrary distances being applied between sampling sites. Therefore, while we acknowledge the imperfect nature of using a straight-line approach, this measure can at least provide an indication of the comparative scale of distance between the different rivers in the analysis. A single central collection site from each river system was used to calculate geographical distances, using the following sites: Phongolo (51), Mkondvo (44), Usuthu (40), Lusushwana (30), Mbuluzi (23), Komati (18), Mlumati (11) and Sabie (2) (Fig. 1).

We also assessed the geographical component of genetic variation using Tess 2.3.1 (Chen *et al.*, 2007), which implements a Bayesian clustering algorithm

for spatial population analysis. Sampling site coordinates for each individual were included as prior information, and run with the admixture model for 50 000 sweeps (10 000 burn-in) for 50 iterations at each value of K_{max} = 2–7. We ran the analyses with the BYM admixture model and did not specify the recessive allele (O. Francois, pers. comm.). The most likely number of clusters was assessed using the deviance information criterion (DIC), and the runs with 20% lowest DIC values were averaged and visualized using Clumpp and Distruct, respectively, as for the Structure analysis. Tess output was further visualized in a geographical perspective using Tess Ad-Mixer 1.0 (Mitchell *et al.*, 2013).

Population differentiation and molecular diversity were assessed for the mtDNA dataset using Arlequin 3.5 (Excoffier, Smouse & Quattro, 1992; Excoffier & Lischer, 2010) for the same sequence data as used in the phylogenetic analysis. As certain samples had data available for only one mtDNA marker, we analysed CO1 and CR datasets separately to maximize sample numbers available for inclusion. The most appropriate substitution model for each locus, available in Arlequin, was selected using jModelTest2 (Guindon & Gascuel, 2003; Darriba *et al.*, 2012) using Akaike's Information Criteria (AIC). While molecular diversity was calculated across all samples, to calculate population differentiation samples were grouped by river. Both pairwise F_{ST} (variation in haplotype frequencies) and Φ_{ST} (sequence divergence between samples) comparisons were calculated. All analyses were run allowing a maximum threshold of missing data at the default level of 0.05, and with significance tested using 10 000 permutations. AMOVA was conducted to partition variance within and among populations for both Φ_{ST} and F_{ST}, with Φ_{ST} distances calculated using pairwise differences. Furthermore, we also analysed the mtDNA dataset with and without the Sabie population, to provide analogous results to the AFLP analyses as this river was not included in the AFLP dataset. Where appropriate, Bonferroni correction was applied when testing multiple pairwise comparisons.

Geometric morphometrics

Body and caudal fin shape were analysed using two-dimensional (2D) geometric morphometric analysis of landmark and semi-landmark data, respectively. These datasets were analysed separately, as the more readily observable differences in caudal fin shape could mask the signal of relatively cryptic body shape in a combined dataset. Specimens for morphometric analysis were designated as either male or female based on the presence or absence, respectively, of pronounced genital papillae (Crass, 1960). Our main morphometric analyses principally used the following four clades: Sabie, Mlumati + Komati, Phongolo, and 'mixed' (Lusushwana, Mbuluzi, Usuthu and Mkondvo), with clades further sub-divided by sex to give eight separate groups for analysis. Although we also analyse these data by river, lower sample numbers and lack of genetic data for all specimens included in the morphological analyses means that results should be viewed more cautiously.

Morphometric data collection

All 176 voucher specimens (Table S1) for morphometric analysis had been formalin fixed and later transferred to IMS, so were expected to show minimal shrinkage effects from storage. Specimens were pinned to a polystyrene board to clarify anatomical features and minimize any deformation due to preservation. Digital 2D photographs were taken from a set distance of 0.5 m using a tripod and Canon EOS 2OD DS126061 camera with Macro lens EF 100 m 1:2.8 USM. Photographs were taken of both lateral body views to minimize error in downstream landmark placement from any small difference in specimen placement on the pinning board (the broad head and large lips of *C. anoterus* may prevent perfectly flat placement). Caudal fins were placed between microscope slides to prevent warping (allowing easier placement for 2D imaging), and as the tails are flat structures only the left side was photographed.

Morphometric analysis

Body landmarks were selected based on previous catfish studies (Reis, Trajano & Hingst-Zaher, 2006; Schmidt & Pezold, 2011) totalling 15 landmarks (Fig. S2). A total of 70 semi-landmarks (Bookstein, 1996; Viscosi & Cardini, 2011) were resampled from the outline curve of the caudal fin (reduced from a higher number of arbitrarily placed points describing the outline). We used a comparatively high number of semi-landmarks relative to sample number, as this was the smallest number of points needed to define the outline of the caudal fin. Previous studies have shown that measurement error increases with reduction in number of points used to describe a curve (Sheets *et al.*, 2006) and that a larger number of semi-landmarks reduces differences in the fitting of points along the curve (Perez, Bernal & Gonzalez, 2006). All landmarks and semi-landmarks were digitized using the program tpsDig (Rohlf, 2013). Specimens for which landmarks were not discernible due to damage or warping were

removed from the dataset giving 171 body specimens (both sides) and caudal fin images for 163 specimens (left-side only).

Morphometric analyses were conducted using MorphoJ (Klingenberg, 2011) and tpsRegr (Rohlf, 2003) software. The body dataset was averaged so that the left and right side data were combined (after reflection) to form a single configuration for each individual, thus reducing potential measurement error in the placement of landmarks and reducing the effects of any fluctuating asymmetry within individuals (e.g. Klingenberg & McIntyre, 1998; Leamy & Klingenberg, 2005). The coordinates of landmarks for each individual were subjected to a Procrustes fit, which removes all features of size, location and superimposes coordinates on the consensus configuration using least squares methods (Rohlf & Slice, 1990).

The data were explored for allometric variation by using a regression of Procrustes coordinates against centroid size in MorphoJ. Only five of 16 subgroups (body and caudal fin data for each group of clade and sex) showed any significant correlation between shape and size. Additionally, canonical variate analysis (CVA) of size-corrected data after multivariate regression pooled within subgroups (e.g. Drake & Klingenberg, 2008) showed no appreciable change in ordination of groups. Given the inconsistent signal of allometry and the lack of improvement in discriminatory power after size correction, subsequent analyses were performed without correction.

PCA was used to identify the principal axes of variation in shape, producing principal component (PC) scores for the full dataset and for subsets of data (for specific comparisons). As PCA does not necessarily recover differences in the data (being a single-group procedure; Strauss, 2010), Procrustes coordinates were used to describe differences between groups defined a priori: (1) CVA, which is an ordination method that maximizes between-group differences relative to within-group variation (Zelditch *et al.*, 2004), and (2) discriminant function analysis (DFA) with cross-validation, which indicates if two groups can be reliably distinguished from each other. We tested the significance of these shape differences using permutation tests performed in MorphoJ.

RESULTS

Phylogeny, uncorrected pairwise distances and haplotype network

Bayesian inference of mtDNA data identifies six well-supported clades (98–100 BPP) that comprise populations from single or combined rivers that are in close proximity (Fig. 2). The mtDNA phylogeny identifies a clade composed of the adjacent Mlumati and Komati river populations (northern Swaziland border) as sister to all other populations (93 BPP). This clade is sister to a Sabie river clade (situated ~100 km north in South Africa), and a larger clade composed of the Phongolo clade and a more complex geographical clade ('mixed' clade) that contains samples from multiple rivers in close proximity from the Highveld of western Swaziland, although the latter 'mixed' clade is poorly supported. Within this geographically diverse grouping, a clade composed largely of Mbuluzi samples, but also containing samples from the Lusushwana (sites 29, 33, 34), is sister to a clade (UMkLu) comprising two subclades. One subclade contains samples predominately from the Usuthu river and a single sample from the Mkondvo river (43), while the other comprises the remaining Lusushwana samples (sites 30, 31, 32, 34, 36) and several samples from the Mkondvo (44). A separate analysis of the CO1 data (data not shown) resolves a similar topology to the concatenated tree, albeit with the Phongolo clade and the three subclades within the 'mixed' clade forming a polytomy.

Uncorrected pairwise distances (K2P) of the barcoding gene (CO1) generated from our data are low, with an average of 1.05% across the clades highlighted in Figure 2. Between-lineage sequence divergence is 0.01–0.02, but 0.005 between the clades Mbuluzi + Lusushwana and UMkLu. Aside from comparison of the latter clades, between-lineage distances are higher than those estimated for within-lineage distances (0.001–0.0038).

The haplotype network based on the concatenated mtDNA sequence data from CO1 and the CR identifies 44 haplotypes from across the eight rivers (Fig. 3). When these genes were analysed separately, 29 CO1 and 34 CR haplotypes were identified (Fig. S3A, B). This is reflected in the finding that molecular diversity was greater for the CR (Theta-pi = 24.300) than for CO1 (Theta-pi = 13.218). All network analyses identified four distinct strongly differentiated haplogroups, which correspond to the main clades identified from phylogenetic inference. Haplogroups consist of samples from the Mlumati + Komati rivers, the Phongolo river, the Sabie river, and the mixed geographical group composed of the multiple rivers (Mbuluzi, Lusushwana, Usuthu and Mkondvo). The concatenated mtDNA analysis shows clear clustering within the mixed group, as identified by phylogenetic inference, so that the groupings Mbuluzi + Lusushwana and UMkLu are distinct, whereas analyses of separate markers indicate haplotype sharing within these clusters composed from different rivers.

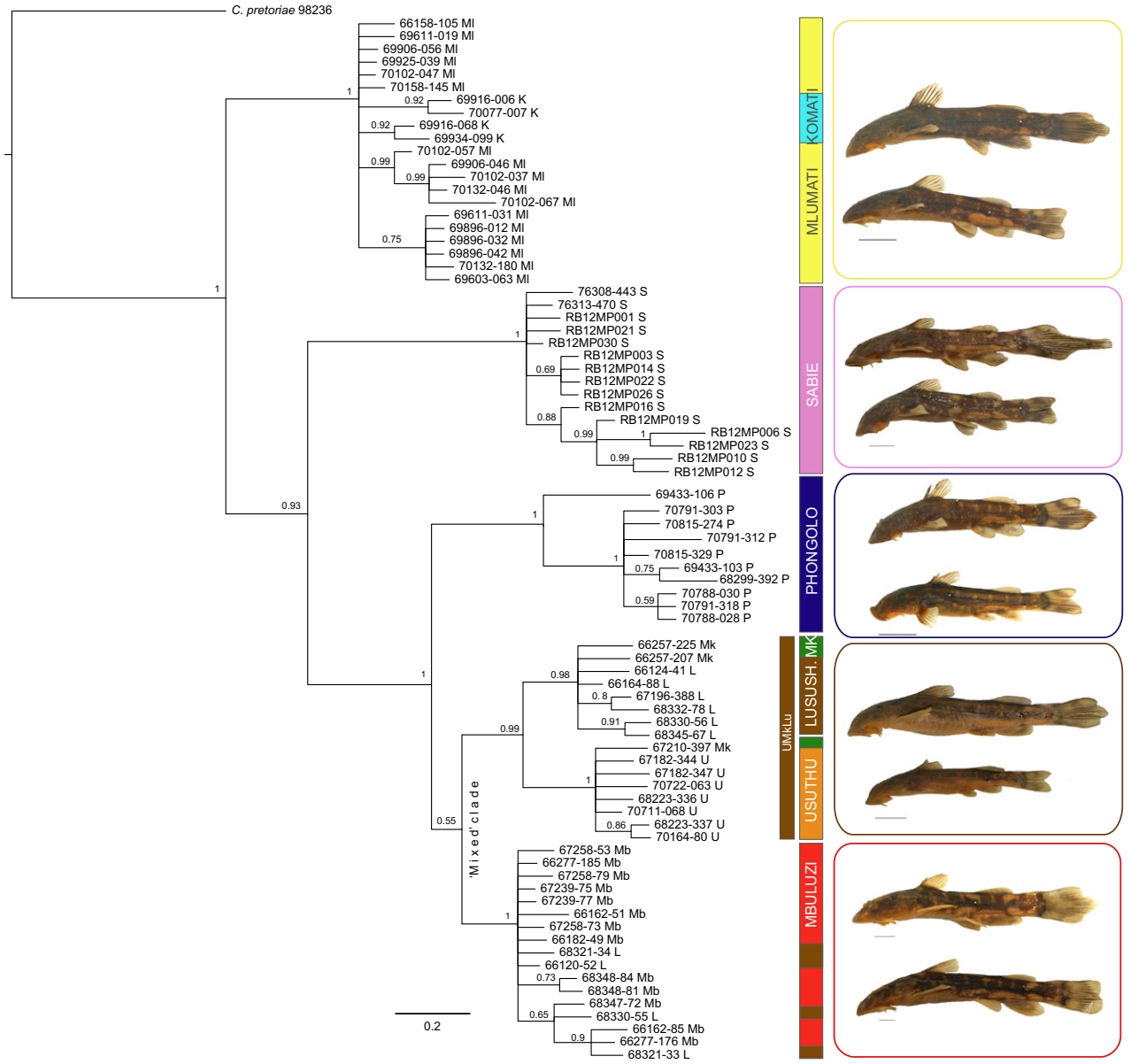


Figure 2. Mitochondrial DNA phylogeny (CO1 and the CR) inferred from Bayesian analysis. Numbers above the branches are Bayesian Posterior Probabilities (BPPs). Photos show males (top) and females (bottom) for each clade (top to bottom): Mlumati + Komati (yellow); Sabie (pink); Phongolo (blue, type locality); Usuthu + Mkondvo + Lusushwana (brown); Mbuluzi + Lusushwana (red). Taxa are coloured by river collection locality as shown in Fig. 1: Mlumati (yellow); Komati (aqua); Sabie (pink); Phongolo (blue); Usuthu (orange); Mkondvo (green); Lusushwana (brown); Mbuluzi (red). See Table S1 for photographed specimen numbers (scale bar = 1 cm).

Low levels of gene flow between phylogenetic lineages

The extended Structure runs, of 1 000 000 generations, gave a likelihood optimum at $K = 5$ (Fig. 4), with the ΔK method giving a likelihood optimum at $K = 3$, with a second ΔK peak at $K = 5$ (see Fig. S4 for probability scores). At $K = 5$ (Fig. 4) discrete clusters are identified for the Phongolo and Mlumati + Komati clades as observed in the mtDNA phylogeny (Fig. 2). Additionally, Mbuluzi samples appear genetically differentiated from the other populations in the 'mixed clade', while the remaining three populations, namely Lusushwana, Usuthu and Mkondvo, show very similar cluster membership. However, these three rivers had the fewest samples included, so it is possible that there is undescribed

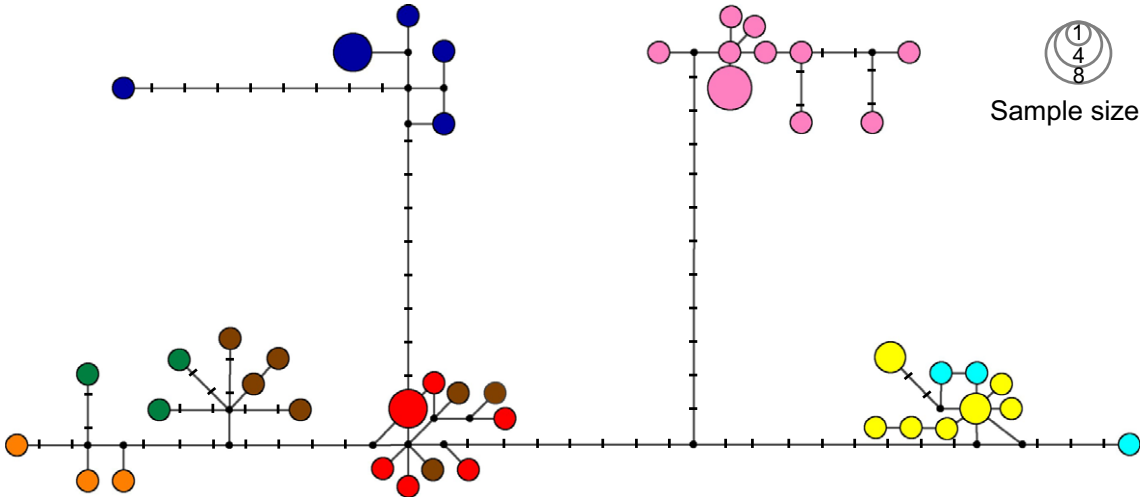


Figure 3. Median-joining haplotype networks based on the concatenated mtDNA sequence data. Haplotypes are coloured by river collection locality as shown in Figure 1: Mlumati (yellow), Komati (aqua), Sabie (pink), Phongolo (blue), Lushushwana (brown), Mkondvo (green), Usuthu (orange) and Mbuluzi (red).

differentiation between these systems that we did not detect in our analysis due to low sample number. Variable probability of assignment to a fifth clade (orange in Fig. 4) occurs at low frequencies across all clades, suggesting shared ancestral variation or historical admixture. Six Lusushwana individuals showed predominant membership assignment to the Mbuluzi clade. The three sampling sites from which these samples were collected from the Lusushwana river (32–34) are in close geographical proximity to each other and distant from the other sites in this river (Fig. 1). The results at $K = 3$ (the most likely K_{max} based on the Evanno method) follow a similar pattern, but differentiation between the Lusushwana, Usuthu and Mkondvo rivers is lacking with samples demonstrating complete shared cluster membership with Phongolo samples (Fig. S4). As for $K = 5$, the Mlumati + Komati and Mbuluzi (with six Lusushwana samples) form well-differentiated clades, and there is less evidence of shared cluster membership across clades. The results at $K = 5$ correlate most closely with the mtDNA results, with the clusters identified in the Structure analysis corresponding to the groups seen in the haplotype network of concatenated mtDNA: Phongolo, Mbuluzi (+ some Lusushwana samples), Mlumati + Komati, mixed clade.

The TESS Ad-Mixer results had lowest DIC values (highest probability) at $K_{max} = 5$–7, and the lowest 20% of DIC values were all within these K-values (Fig. S5). Cluster membership was almost exactly equivalent to the results from Structure runs. However, when mapped in geographical space these results showed slightly different cluster membership probabilities (although this is modelled over the whole map area rather than just considered at actual sampling sites), but exhibited a general trend to higher levels of admixture at higher values of K_{max}.

Analysis of hierarchical molecular variance (AMOVA) of AFLP data demonstrated that a significant proportion of molecular variation was due to differentiation between populations, when grouped by either river (23% of total variation; $P < 0.005$) or by

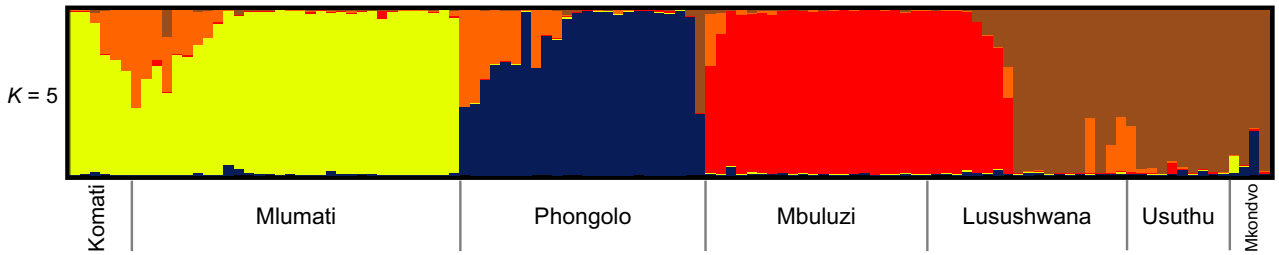


Figure 4. Structure analysis of AFLP data (307 loci) with cluster assignment probabilities for $K = 5$. Individuals (117) are arranged by river system and cluster membership proportion. Likelihood values and additional cluster membership results for other K-values are given in Fig. S4.

Table 1. Population pairwise comparisons among samples grouped by river system for mtDNA CR Φ_{ST} (top right) and AFLP data Φ_{PT} (bottom left)

River system	Komati	Mlumati	Mbuluzi	Lusushwana	Usuthu	Mkondvo	Phongolo	Sabie
Komati	–	0.475	0.894*	0.878*	0.888	0.859	0.834*	0.912*
Mlumati	0.046	–	0.862*	0.840*	0.851	0.824	0.841*	0.893*
Mbuluzi	0.283*	0.276*	–	0.141	0.341*	0.303	0.622*	0.872*
Lusushwana	0.235*	0.231*	0.169*	–	0.119	−0.050	0.577*	0.857*
Usuthu	0.193*	0.240*	0.258*	0.041	–	0.018	0.615*	0.859*
Mkondvo	0.229*	0.266*	0.293*	0.092	0.026	–	0.547	0.862*
Phongolo	0.264*	0.272*	0.288*	0.178*	0.160*	0.155*	–	0.859*

*Significant following Bonferroni correction ($\alpha = 0.002$).

genetic clade identified in Structure analysis (25%; $P < 0.005$). The AMOVA of mtDNA CR and CO1 data based on sequence divergence between samples, revealed an even greater proportion of variation (88%, $P < 0.001$ for CO1; and 83%, $P < 0.001$ for the CR) was explained by differences among samples from different rivers.

Similar patterns are seen in the pairwise comparisons of Φ_{PT} and Φ_{ST} (Table 1), where genetic differentiation was higher between rivers based on the CR (including Sabie samples: mean $\Phi_{ST} = 0.657$; SD = 0.303; excluding Sabie samples: mean $\Phi_{ST} = 0.689$; SD = 0.244) than on AFLP data (mean $\Phi_{PT} = 0.200$; SD = 0.086). As might be expected, when testing only the genetic clades identified by Structure analysis for the AFLP data, group differentiation was slightly higher than for individual rivers (mean $\Phi_{PT} = 0.241$; SD = 0.038; Table S3). Pairwise population comparisons based on mtDNA CO1 showed Φ_{ST} values similar to those from the CR dataset with patterns of significance also largely congruent (Table S4). Pairwise comparisons of F_{ST} (rather than Φ_{ST}) for the mtDNA datasets exhibited lower levels of pairwise differentiation although still demonstrated high levels of differentiation between the Sabie river and other populations (Table S5), and a lower proportion of variance was explained by differences among populations (CR = 8%; CO1 = 36%; both $P < 0.001$).

In PCA of the CR and CO1 datasets, the first three PCA axes represent 18.52, 15.98 and 9.45%, and 16.43, 10.17 and 7.23% of the variation, respectively. The PCA plots support phylogenetic and haplotype analysis as they show clustering of Mlumati + Komati, Sabie, Phongolo and the 'mixed clade' populations (Fig. S6A–D), although within the PC axes displayed (i.e. PC1–3) there is no separation of populations within this last grouping. For the AFLP dataset, which does not include Sabie samples, the first three PCA axes represent 22.47, 19.12 and 16.04% of the variation, respectively. The PCA plots, based on the first three axes, support the separation of the Mlumati + Komati and Mbuluzi + Lusushwana populations, but a third cluster includes both Phongolo and UMkLu populations (Fig. S6A–F).

Table 2. Simple and partial Mantel test results

	mtDNA		AFLP	
Comparison	*R*	*P*-value	*R*	*P*-value
Simple Mantel tests				
Genetic × Geographical	0.528	< 0.01	0.535	0.01
Genetic × Morphometric	0.553	< 0.01	0.134	n.s.
Morphometric × Geographical	0.600	0.02	0.321	n.s.
Partial Mantel tests				
Genetic × Geographical | Morphometric	0.294	n.s.	0.524	0.01
Genetic × Morphometric | Geographical	0.348	n.s.	−0.047	n.s.
Geographical × Morphometric | Genetic	0.436	n.s.	0.297	n.s.

Correlation between genetic (mtDNA and AFLP), morphometric (Procrustes distance of male caudal fin shape) and geographical (km) distances between *C. anoterus* populations. Significance tested using 5000 permutations. Note AFLP data do not include the Sabie site. n.s., Not significant.

Distance correlation

Simple Mantel tests demonstrated a significant relationship between geographical distance and genetic distance between rivers for both AFLP and mtDNA (Table 2). These results suggest an effect of isolation by distance, despite the AFLP dataset not including the most geographically isolated population (Sabie). Tests also exhibited significant correlations between morphometric and both geographical and genetic distance for the mtDNA data but not the AFLP data. However, partial Mantel tests revealed only one significant correlation while controlling for other variables, that of genetic and geographical distances while controlling for morphometric distances (AFLP data only; Table 2).

Morphological variation

PCA of the entire body dataset showed that the first three PCs explained over 56% of the variation, where: PC1 described a ventral shape change in body depth and a dorsal change of head orientation; PC2 described a change in body depth; and PC3 showed variation in body length (Fig. 5A). Plotting PC1 against PC2 for the body data did not effectively separate clade or sex groups (Fig. 5B), with most data points overlapping around the origin. A DFA separated the sexes, but with considerable overlap [permutation test (10 000 runs) $P < 0.0001$; Fig. 5C]. Analysis of body shape differentiation yielded significant differences in several male comparisons but no female–female comparisons (Procrustes distances, 10 000 permutations; Table 3). Furthermore, pairwise plots of canonical variates 1 and 2 showed no clear distinction of body shape between sexes, with differentiation between clades but not sexes within clades (Fig. 5D).

In analysis of caudal fin shape variation, the first PC explained 70.9% of the variation and the first seven PCs accounted for 96.1% of the variation. The first two PCs described most of the variation, where PC1 described extension of central rays of the caudal fin (for both male and female shapes, but not necessarily resembling either shape), and PC2 described extension of the central caudal fin rays and shape change to a clear trilobate morphology (Fig. 5E). Plotting PC1 against PC2 clearly separated sexes with no overlap between male and female specimens (Fig. 5F). Furthermore, there was little variation in female caudal fin morphology with all clades overlapping in morphospace, but some clear differences in male caudal fin morphology. Male specimens from Sabie formed a distinct well-separated cluster, but some overlap in male caudal fin shape is noted, as although specimens from the Mlumati + Komati and the Phongolo clades formed tight clusters distinct from each other, they both overlapped with part of the mixed clade (Fig. 5F). Specimens from the mixed clade showed a wide variation in PC1 (Fig. 5F), indicating either higher levels of natural variation or suggesting that these populations are further subdifferentiated in caudal morphology.

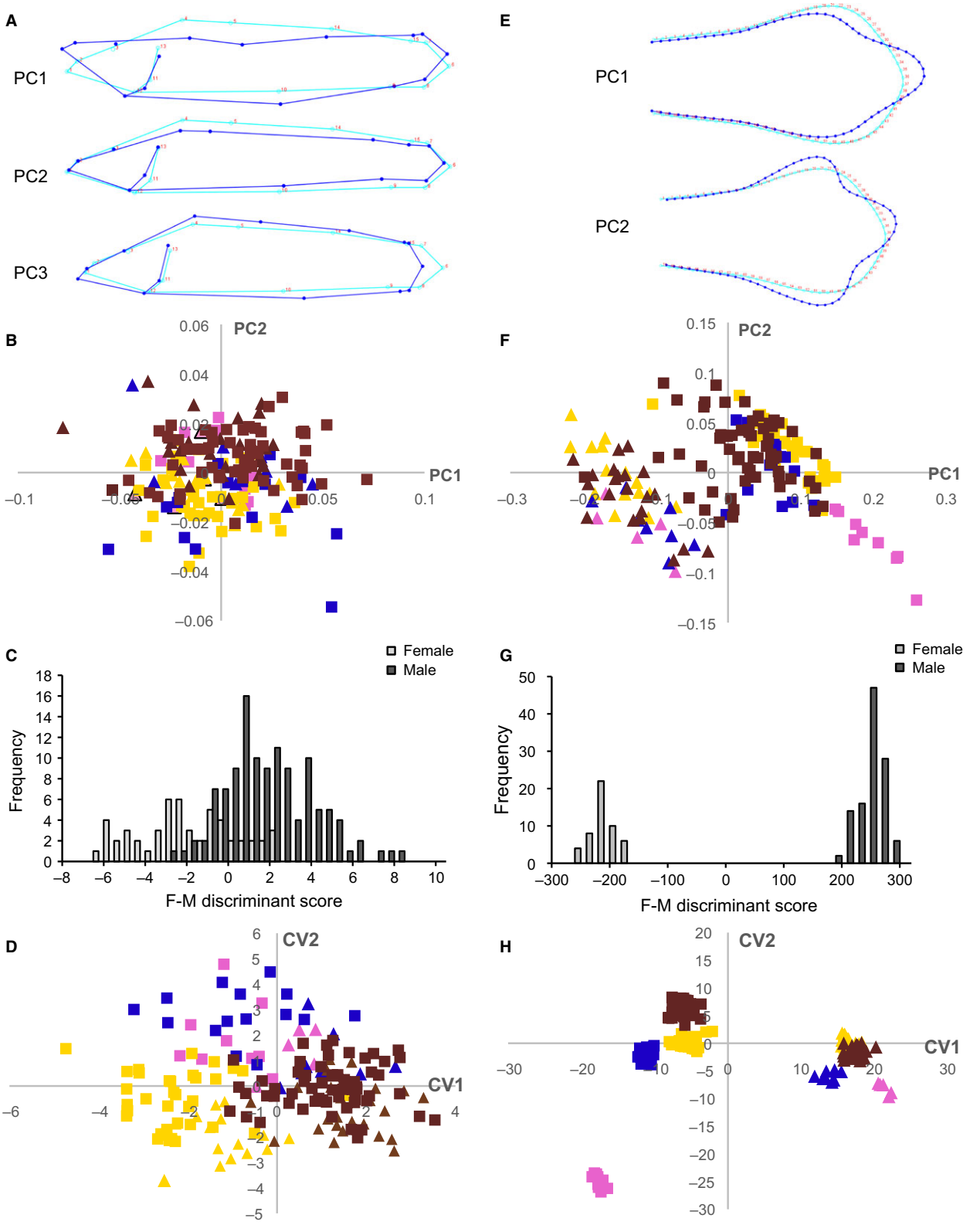
In strong contrast to the body shape data, DFA completely separated the male and female caudal fins (Fig. 5G) and permutation tests (10 000 runs) found their separation to be highly significant ($P < 0.0001$). Analysis of caudal fin shape demonstrated significant differentiation between all male–male and male–female comparisons, but considerably lower differentiation between clades of female caudal fin shape with no female–female comparisons being significant following Bonferroni correction (Procrustes distances, 10 000 permutations; Table 3, Fig. S7). A plot of the two canonical variates accounting for most variation within the dataset (CV1 and CV2; 80.8%; Fig. 5H) shows that specimens grouped by clade and sex are generally well separated. The CVA of caudal fin shape separated all male clades (the mixed clade no longer overlapped with other populations), and two of the female populations (Phongolo and Sabie). The large variation of within-clade caudal fin differentiation between sexes is shown in Fig. S8.

Given the large variation in PC1 of caudal shape across specimens from the mixed clade, we also conducted an analysis by river to see if populations within the mixed clade were differentiated by caudal fin shape. As for the main analysis, for the body data, in CVA males and females clustered by river rather than by sex (Fig. S9A). By contrast, for the caudal fin data, CVA clearly separated the male specimens of Mbuluzi from the other mixed clade rivers, but with some overlap between the remaining populations of Usuthu, Lusushwana and Mkondvo (Fig. S9B). There was also clear separation of female specimens from all rivers, but analysing by river reduced sample number group considerably, so these results should be interpreted cautiously.

DISCUSSION

Unexpected diversity identified from the African Highveld

We tested for genetic and morphological differences based on reasonably dense sampling of the pennant-tailed suckermouth catfish, *Chiloglanis anoterus*, from across its range in the African Highveld. Our integrative analyses uncovered high levels of unexpected genetic diversity, as well as morphological



disparity between males, with females remaining relatively conserved. Using a delimitation criterion based on multiple data sources, the lineages identified here are all reciprocally monophyletic regarding mtDNA, genetically divergent units and morphologically highly differentiated. The majority of genetic analyses

Figure 5. Geometric morphometric analysis of body shape data (plots A–D) and caudal fin shape (plots E–H). For plots B, D, F and H, symbols and colours represent the following clades: squares, male; triangles, female; pink, Sabie; yellow, Mlumati + Komati; blue, Phongolo; brown, 'mixed clade'. A, wireframe shape change graphs for the first three PCs of PCA of body shape data accounting for >56% of the variation. B, PC1 plotted against PC2 showing little differentiation between groups for body shape. C, DFA of male vs. female specimens showing distinct overlap between sexes in body shape. D, CVA: CV1 plotted against CV2 showing specimens clustering by clade rather than by sex. E, wireframe shape change graphs for the first two PCs of PCA of caudal fin data accounting for 81% of the variation. F, PC1 plotted against PC2 showing no overlap between male and female caudal fin shape. G, DFA of male vs. female specimens showing complete separation of sexes in caudal fin shape. H, CVA: CV1 plotted against CV2 showing significant differentiation of all male clades and three of the female clades.

Table 3. Procrustes distances for group pairwise comparisons of body shape (bottom left) and caudal fin shape (top right)

		Sabie		Mlumati + Komati		Phongolo		Mixed	
mtDNA clade	Sex	Female	Male	Female	Male	Female	Male	Female	Male
Sabie	F	–	0.327*	0.077	0.226*	0.049	0.205*	0.059	0.184*
	M	0.031	–	0.356*	0.148*	0.305*	0.160*	0.339*	0.196*
Mlumati + Komati	F	0.023	0.028	–	0.237*	0.071	0.221*	0.034	0.184*
	M	0.029	0.028*	0.016	–	0.203*	0.058*	0.225*	0.067*
Phongolo	F	0.031	0.030	0.026	0.029*	–	0.182*	0.055	0.155*
	M	0.033	0.028	0.030	0.023	0.025	–	0.212*	0.081*
Mixed	F	0.030	0.025	0.018	0.025*	0.020	0.031*	–	0.172*
	M	0.037	0.026	0.027*	0.026*	0.018	0.026*	0.016	–

Significance tested using permutation tests (10 000 rounds). *Significant following Bonferroni correction ($\alpha = 0.002$). The mixed clade is composed of populations from the Mbuluzi, Lusushwana, Usuthu and Mkondvo rivers.

identified five strongly delimited groups composed of populations from the Sabie, Phongolo, Mlumati + Komati, Mbuluzi + Lusushwana, and Usuthu + Mkondvo + Lusushwana (UMkLu) rivers. Although phylogenetic inference based on mtDNA data identifies additional diversity in this last clade (UMkLu; Fig. 2), and Usuthu and Lusushwana populations are significantly differentiated by AMOVA (Table 1), this differentiation is not observed in the population assignments from the AFLP Structure analysis or PCA analysis. It is therefore possible that the Lusushwana, Usuthu and Mkondvo clade is currently undergoing population divergence, although higher gene flow, the possible impact of higher population sizes or even less selection pressure could alternatively be inferred. Additional sampling, particularly from the Mkondvo, would help to clarify group assignment of these rivers that are not well represented in this study. Morphological data based on geometric morphometrics supports our genetic findings, and although there is a much clearer signal regarding differentiation of caudal fin morphotypes when denoting four clades (assuming a single mixed geographical clade), we do find a significant difference between the two subclades Mbuluzi + Lusushwana and UMkLu within this larger grouping when rivers are analysed separately (Fig. S9).

Despite these findings, uncorrected pairwise distances (K2P) of the barcoding gene (CO1) generated from our data are low, suggesting recent origins of these lineages. Although the minimum sequence divergence necessary to consider a clade a distinct species is a controversial issue in species delimitation (e.g. Funk *et al.*, 2012, and references therein), there are many examples of distinct vertebrate species with low mtDNA sequence divergence including some Malagasy (Vieites *et al.*, 2009) and neotropical frogs (Funk *et al.*, 2012), *Thunnus* (Ward *et al.*, 2005) and most notably cichlid fishes (e.g. Sturmbauer *et al.*, 2001). The cited frog studies classify lineages that display > 3% (although sometimes only 1–2%) uncorrected pairwise genetic divergence in mtDNA, as well as morphological and bioacoustic data differences as confirmed candidate species (CCS). Certainly, sexually selected characters, due to their contribution to the reproductive isolation of species, are suggested to more likely represent species-specific differences than naturally selected characters (see Padial *et al.*, 2010). However, despite clear morphological differences between clades we do

not report deep genealogical lineages (> 3% sequence divergence). As such while our data indicate that the identified lineages do fit the criteria to be CCS, further information on female preference tests would provide necessary evidence as to whether the identified clades are young species, or intraspecific genealogical lineages. A conservative approach is to refer to these lineages as evolutionary significant units (ESUs), as they are reciprocally monophyletic based on mtDNA, and show significant divergence of allele frequencies at nuclear loci (Mortiz, 1994).

Geographical isolation and endemism within highland rivers

While a correlation between geographical distance and genetic isolation among populations is generally expected, we find significant genetic isolation by distance for both AFLP and mtDNA data (Table 2), highlighting the genetic isolation of individual rivers, and refuting a hypothesis of panmixia throughout the existing species range. Morphometric distances were also significantly correlated with geographical distance and with genetic distance for mtDNA, but not for AFLP data. As these pairwise correlations could reflect indirect correlation with a third factor, we also used partial Mantel tests to conduct each pairwise comparison while controlling for the third factor. Although partial Mantel tests have been condemned for high Type I error rates, with some authors advocating a more stringent ($P = 0.001$) threshold for determining significance (discussed by Diniz-Filho *et al.*, 2013), they remain a common and useful test for exploring multivariate relationships in environmental and genetic datasets (e.g. Funk, Egan & Nosil, 2011; DeWoody, Trewin & Taylor, 2015). Testing the correlation of phenotype with genotype while controlling for geographical distance between populations has been used to describe isolation by adaptation (Nosil, Egan & Funk, 2008), and to infer the process of adaptive divergence where morphology is correlated to genetic distance beyond any population structure related merely to geography. Here, the partial Mantel tests revealed no significant correlation of genetic and morphometric distances when controlling for geography (Table 2), suggesting the differences in male caudal fin shape are not driven by an adaptive process independent of geography. Thus, the results of simple and partial Mantel tests suggest that geographical isolation, rather than ecological adaptation, is driving morphological divergence in this species, supporting the hypothesis that the male tail shape is a product of sexual rather than natural selection. Furthermore, the caudal fin elongation would not seem to be a product of geographical isolation and subsequent genetic drift or local environmental adaptation given that the trait is only seen in males and thus does not appear to be an ecological trait.

Our results therefore suggest that recent diversification in these highland catfish is likely to have been facilitated through geographical isolation of populations in the upper catchments, as the distributions of the lineages are largely allopatric. The three major river systems of this study output in the Maputo region (Mozambique) in the Lowveld, and interconnect during flooding events that occur regularly across this coastal region. However, the focal taxa are stenotopic, found only in shallow riffle and waterfall habitats (Bills *et al.*, 2004), and are absent even a few metres downstream of this habitat in deeper water, where they are replaced by other *Chiloglanis* species (R. Bills, pers. observ.). As such, these taxa are restricted to high altitude (Bills *et al.*, 2004). This is supported by the extensive biodiversity survey that showed *C. anoterus* to be absent from the Lowveld and only occasionally found in the Midveld (Fig. S1). Thus, dispersal in a downstream direction is unlikely, and instead dispersal via the upper reaches is hypothesized. This may be via river capture, which is a common geomorphological phenomenon, particularly in geologically active regions, and therefore may contribute to the lineage sharing. In addition, as waterfalls generally do not present barriers to these fish (R. Bills, pers. observ.), they therefore potentially have the ability to disperse in the upper reaches.

When considering the three river basins, it would appear that some connectivity in the upper reaches probably occurred between two of these systems as the majority of analyses identified populations from Usuthu + Mkondvo + Lusushwana (which form part of the Maputo river basin) as more closely related to those of the Mbuluzi river basin, than they are to those of the Phongolo river (Maputo river basin). However, two analyses based on AFLPs (PCA, Fig. S6F; Evanno Structure ΔK, Fig. S4) indicated a close assignment between all sampled rivers from the Maputo river basin (i.e. Usuthu, Mkondvo, Lusushwana and Phongolo).

Furthermore, our findings indicate that in contrast to the previously large range size of *Chiloglanis anoterus* (~43 875 km^2), the ESUs have relatively small geographical distributions (~3750–4162 km^2) with several restricted to a single river, although their ranges do not reach the levels of microendemism documented in the Caney Fork barcheek darters that speciated allopatrically within a single highland river system (Hollingsworth & Near, 2009). Nevertheless, our findings have the potential to lead to an increase in diversity of *Chiloglanis* endemics being recognized within this region. While *C. anoterus* is

currently listed as a species of least concern by the IUCN Red List assessment, its conservation status may need reassessment if ESUs are taken into account. In particular, threats to the upper catchments in which these fish occur include: sedimentation from forestry and agriculture, alien predators such as trout, and modifications to river flow in South Africa that may also impact these fish (Engelbrecht, Bills & Cambray, 2007).

Sexual selection in *Chiloglanis anoterus*?

In common with other highland freshwater fish that have diversified on a local scale, e.g. eastern North American barcheek darters that display an apparent lack of morphological and ecological diversification (Hollingsworth & Near, 2009), the suckermouth catfish ESUs also have conserved morphological body diversity between clades. However, in contrast these catfish are highly differentiated by a putative sexually selected trait, with caudal fin shape separating specimens by genetically identified clades, as well as sex (Fig. 5), although females are morphologically conserved. Our data also indicate that the ESUs are likely to have arisen rapidly, which is in contrast to the highland endemic darters of the Caney Fork river system that show deep divergences despite exhibiting morphological stasis. Sexual selection through female choice has been shown to facilitate diversification in several groups (e.g. Ritchie *et al.*, 2005, 2007; Boul *et al.*, 2007; Seldon *et al.*, 2008), and is particularly synonymous with the high diversification rates reported in some cichlid fish clades (e.g. Day, Cotton & Barraclough, 2008; Wagner *et al.*, 2012). As some other *Chiloglanis* species also display sexual dimorphism of their caudal fins (Friel & Vigliotta, 2011) it would be of interest to investigate if sexual selection has promoted their diversity compared with other Mochokidae genera such as *Synodontis* that have been shown to have a constant rate of diversification (Day *et al.*, 2013).

Female preference for a sexually selected trait is classically found among fishes in the genus *Xiphophorus* (swordtails) in which females of species with 'swords' (an extension of caudal fin rays) display a preference for longer swords. However, within *Xiphophorus* species there is considerable heterogeneity in female preferences (e.g. Wong & Rosenthal, 2006), and while not tested here, our results also suggest that preferences for different tails may have arisen within suckermouth catfishes. Among the swordtails, the presence of fin extensions is known to incur serious fitness (Rosenthal *et al.*, 2001) and energetic costs because the 'sword' is less hydrodynamically efficient. Decreased male swimming performance has also been demonstrated in Trinidadian guppies (*Poecilia reticulata*) that display polymorphism in caudal fin shape (Karino, Orita & Sato, 2006) consistent with the handicap hypothesis (Zahavi, 1975). Although male *C. anoterus* caudal fins are a different shape from swordtails and guppies, they are extended asymmetrically, and as these catfish occur in fast-flowing water, it is likely that they are exposed to similar fitness and energetic costs. Although our results suggest that sexual selection through female choice is likely to have been an important mechanism in the diversification of this group, an important next step in demonstrating this should include explicit testing for female-biased migration (e.g. Ritchie *et al.*, 2007) as well as mate choice experiments.

CONCLUSIONS

This study provides valuable insight into how geographical variation associated with a putatively sexually selected trait relates to strong divergence between *C. anoterus* populations that we suggest represent ESUs, and fit the criteria for confirmed candidate species (Padial *et al.*, 2010). The combined use of genetic and phenotypic data provides evidence that elevated regional diversity from Africa's Highveld has been facilitated by geographical isolation, with sexual selection through female choice probably having driven variation in male caudal fin morphology. We highlight the need for further integrative ichthyological investigation of the African Highveld, not only to better assess the diversity of this understudied region, but also to understand the processes driving diversification in these highland rivers.

ACKNOWLEDGEMENTS

We thank Siobhan Cox for helpful assistance in the generation of AFLP data and Mari-Wyn Burley for sequencing. We are grateful to Olivier Francois for advice in the use of the Tess 2.3 software program, and thank two anonymous reviewers whose comments improved the manuscript. The genetic and morphometric data formed the basis of Jake Morris' (M.Res) and Jarome Ali's (B.Sc.) UCL research projects. NERC grant NE/I528169/1 and a UCL Graduate Scholarship supported CRP and AGPF respectively.

REFERENCES

Arnegard ME, McIntyre PB, Harmon LJ, Zelditch ML, Crampton WGR, Davis JK, Sullivan JP, Lavoué S, Hopkins CD. 2010. Sexual signal evolution outpaces

ecological divergence during electric fish species radiation. *The American Naturalist* **176:** 335–356.

Bandelt H-J, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141:** 743–753.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

Basolo AL, Alcaraz G. 2003. The turn of the sword: length increases male swimming costs in swordtails. *Proceedings of the Royal Society of London Series B* **270:** 1631–1636.

Bickford D, Lohman DJ, Sodhi NS, Ng PK, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. *Trends in Evolution and Ecology* **22:** 148–155.

Bills R, Boycott R, Fakudze M, Khumalo N, Msibi J, Scott L, Terry S, Tweddle D. 2004. Fish and fisheries survey of Swaziland (2002–2003). *SAIAB Investigational Report* **70:** 1–174.

Bookstein F. 1996. Applying landmark methods to biological outline data. In: Mardia KV, Gill CA, Dryden IL, eds. *Image fusion and shape variability techniques*. Leeds: Leeds University Press, 59–70.

Boul KE, Funk WC, Darst CR, Cannatella DC, Ryan MJ. 2007. Sexual selection drives speciation in an Amazonian frog. *Proceedings of the Royal Society of London Series B* **274:** 399–406.

Chen C, Durand E, Forbes F, François O. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* **7:** 747–756.

Collier GE, Murphy WJ, Espinoza M. 2009. Phylogeography of the genus *Epiplatys* (Aplocheiloidea: Cyprinodontiformes). *Molecular Phylogenetics and Evolution* **50:** 190–196.

Crass RS. 1960. Notes on the freshwater fishes of Natal with descriptions of 4 new species. *Annals of the Natal Museum* **14:** 405–458.

Cureton JC II, Broughton RE. 2014. Rapid morphological divergence of a stream fish in response to changes in water flow. *Biology Letters* **10:** 20140352.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

Darwall WRT, Smith KG, Allen DJ, Holland RA, Harrison IJ, Brooks EGE. 2011. *The diversity of life in African freshwaters: under water, under threat. An analysis of the status and distribution of freshwater species throughout mainland africa*. Cambridge, UK and Gland, Switzerland: IUCN.

Day JJ, Cotton JA, Barraclough TG. 2008. Tempo and mode of diversification of Lake Tanganyika cichlid fishes. *PLoS ONE* **3:** e1730.

Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T. 2013. Continental diversification of an African catfish radiation (Mochokidae: *Synodontis*). *Systematic Biology* **62:** 351–365.

DeWoody J, Trewin H, Taylor G. 2015. Genetic and morphological differentiation in *Populus nigra* L.: isolation by colonization or isolation by adaptation? *Molecular Ecology* **24:** 2641–2655.

Diniz-Filho JA, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, de Campos Telles MP, Rangel TF, Bini LM. 2013. Mantel test in population genetics. *Genetics and Molecular Biology* **36:** 475–485.

Drake A, Klingenberg C. 2008. The pace of morphometric change: historical transformation of skull shape in St. Bernard dogs. *Proceedings of the Royal Society of London Series B* **275:** 71–76.

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny ML, Sullivan CA. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Review* **81:** 163–182.

Earl DA, vonHoldt BM. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* **4:** 359–361.

Engelbrecht J, Bills R, Cambray J. 2007. *Chiloglanis anoterus*. The IUCN Red List of Threatened Species. Version 2014.2. Available at: www.iucnredlist.org. Downloaded on 7 August 2014.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611–2620.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecological Resourses* **10:** 564–567.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application. *Genetics* **131:** 479–491.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7:** 574–578.

Feulner PGD, Kirschbaum F, Mamonekene V, Ketmaier V, Tiedemann R. 2007. Adaptive radiation in African weakly electric fish (Teleostei: Mormyridae: *Campylomormyrus*): a combined molecular and morphological approach. *Journal of Evolutionary Biology* **20:** 403–414.

Friel JP, Vigliotta TR. 2006. *Synodontis acanthoperca*, a new species from the Ogôoué River system, Gabon with comments on spiny ornamentation and sexual dimorphism in mochokid catfishes (Siluriformes: Mochokidae). *Zootaxa* **1125:** 45–56.

Friel JP, Vigliotta TR. 2011. Three new species of African suckermouth catfishes, genus *Chiloglanis* (Siluriformes: Mochokidae), from the lower Malagarasi and Luiche rivers of western Tanzania. *Zootaxa* **21:** 1–21.

Funk DJ, Egan SP, Nosil P. 2011. Isolation by adaptation in *Neochlamisus* leaf beetles: host-related selection promotes neutral genomic divergence. *Molecular Ecology* **20:** 4671–4682.

Funk WC, Caminer M, Ron SR. 2012. High levels of cryptic species diversity uncovered in Amazonian frogs.

Proceedings of the Royal Society of London Series B **279:** 1806–1814.

Goodier SAM, Cotterill FPD, O'Ryan C, Skelton PH, de Wit MJ. 2011. Cryptic diversity of African Tigerfish (genus *Hydrocynus*) reveals palaeogeographic signatures of linked Neogene geotectonic events. *PLoS ONE* **6:** e28775.

Guindon S, Gascuel O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52:** 696–704.

Hadfield JD, Richardson DS, Burke T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology* **15:** 3715–3730.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis proram for windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Hollingsworth PR Jr, Near TJ. 2009. Temporal patterns of diversification and microendemism in eastern highland endemic barcheek darters (Percidae: Etheostomatinae). *Evolution* **63:** 228–243.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754–755.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Joly S, Stevens MI, van Vuuren BJ. 2007. Haplotype networks can be misleading in the presence of missing data. *Systematic Biology* **56:** 857–862.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24:** 1403–1405.

Karino K, Orita K, Sato A. 2006. Long tails affect swimming performance and habitat choice in the male guppy. *Zoological Science* **23:** 255–260.

Klingenberg CP. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources* **11:** 353–357.

Klingenberg CP, McIntyre GS. 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution* **52:** 1363–1375.

Lanfear R, Calcott B, Ho SY, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29:** 1695–1701.

Leamy LJ, Klingenberg CP. 2005. The genetics and evolution of fluctuating asymmetry. *Annual Review of Ecology, Evolution, and Systematics* **36:** 1–21.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452.

Markert JA, Schelly RC, Stiassny LJ. 2010. Genetic isolation and morphological divergence mediated by high-energy rapids in two cichlid genera from the lower Congo rapids. *BMC Evolutionary Biology* **10:** 149.

Masta SE, Maddison WP. 2002. Sexual selection driving diversification in jumping spiders. *Proceedings of the Natural Academy of Sciences* **99:** 4442–4447.

Mitchell MW, Rowe B, Sesink Clee PR, Gonder MK. 2013. TESS Ad-Mixer: a novel program for visualizing TESS Q matrices. *Conservation Genetic Resources* **5:** 1075–1078.

Montoya-Burgos JI. 2003. Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. *Molecular Ecology* **12:** 1855–1867.

Morris J, Ford AGP, Ali JR, Peart CR, Bills R, Day JJ. 2015. Data from: High levels of genetic structure and striking phenotypic variability in a sexually dimorphic suckermouth catfish from the African Highveld. *Dryad Digital Repository*. doi:10.5061/dryad.661jb.

Mortiz C. 1994. Defining 'Evolutionary Significant Units' for conservation. *Trends in Evolution and Ecology* **9:** 373–375.

Ng HH. 2004. The *Microsynodontis* (Teleostei: Siluriformes: Mochokidae) of the lower Guinea region, west central Africa, with the description of eight new species. *Zootaxa* **531:** 1–52.

Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62:** 316–336.

Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2015. *vegan: Community Ecology, Package. R Package Version 2.2–1*. Available at: http://CRAN.R-project.org/package=vegan.

Padial JM, Miralles A, De la Riva I, Vences M. 2010. The integrative future of taxonomy. *Frontiers in Zoology* **7:** 16.

Pakkasmaa S, Piironen J. 2001. Water velocity shapes juvenile salmonids. *Evolutionary Ecology* **14:** 721–730.

Peakall R, Smouse PE. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6:** 288–295.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28:** 2537–2539.

Peart CP, Bills R, Wilkinson M, Day JJ. 2014. Nocturnal claroteine catfishes reveal dual colonisation but a single radiation in Lake Tanganyika. *Molecular Phylogenetics and Evolution* **73:** 119–128.

Perez SI, Bernal V, Gonzalez PN. 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. *Journal of Anatomy* **208:** 769–784.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Rambaut A, Drummond AJ. 2009. *Tracer v1.5*. Available at: http://beast.bio.ed.ac.uk/Tracer.

Reis RE, Trajano E, Hingst-Zaher E. 2006. Shape variation in surface and cave populations of the armoured catfishes *Ancistrus* (Siluriformes: Loricariidae) from the Sao Domingos karst area, upper Tocantins River, Brazil. *Journal of Fish Biology* **68:** 414–429.

Ritchie MG, Webb SA, Graves JA, Magurran AE, Garcia CM, Macias Garcia C. 2005. Patterns of speciation in endemic Mexican Goodeid fish: sexual conflict or early radiation? *Journal of Evolutionary Biology* **18:** 922–929.

Ritchie MG, Hamill RM, Graves JA, Magurran AE, Webb SA, Macías Garcia C. 2007. Sex and differentiation: population genetic divergence and sexual dimorphism in Mexican goodeid fish. *Journal of Evolutionary Biology* **20:** 2048–2055.

Roberts TR. 1989. Systematic revision and description of new species of suckermouth catfishes (*Chiloglanis*, Mochokidae) from Cameroun. *Proceedings of the California Academy of Sciences, Series 4* **46:** 151–178.

Rohlf FJ. 2003. *tpsRegr: Multivariate multiple regression of shape onto independent variables, Version 1.34*. Stony Brook, NY: Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf FJ. 2013. *tpsDig, digitize landmarks and outlines, Version 2.17*. Stony Brook, NY: Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf FJ, Slice DE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* **39:** 40–59.

Rosenberg NA. 2003. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4:** 137–138.

Rosenthal GG, Flores Martinez TY, Garcia de Leon FJ, Ryan MJ. 2001. Shared preferences by predators and females for male ornaments in swordtails. *The American Naturalist* **158:** 146–154.

Schmidt RC, Pezold F. 2011. Morphometric and molecular variation in mountain catfishes (Amphiliidae: *Amphilius*) in Guinea, West Africa. *Journal of Natural History* **45:** 521–552.

Schmidt RC, Bart HL Jr, Nyingi DW, Gichuk NN. 2014. Phylogeny of suckermouth catfishes (Mochokidae: *Chiloglanis*) from Kenya: the utility of growth hormone introns in species level phylogenies. *Molecular Phylogenetics and Evolution* **79:** 415–421.

Schwarzer J, Misof B, Ifuta SN, Schliewen UK. 2011. Time and origin of cichlid colonization of the Lower Congo rapids. *PLoS ONE* **6:** e22380.

Seldon N, Merrill RM, Tobias JA. 2008. Sexually selected traits predict patterns of species richness in a diverse clade of suboscine birds. *The American Naturalist* **171:** 620–631.

Sheets HD, Covino KM, Panasiewicz JM, Morris SR. 2006. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. *Frontiers in Zoology* **3:** 15.

Skelton P. 2001. *A complete guide to the freshwater fishes of southern Africa*. South Africa: Struik.

Steve R, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, eds. *Bioinformatics methods and protocols: methods in molecular biology*. Totowa, NJ: Humana Press, 365–386.

Strauss RE. 2010. Discriminating groups of organisms. In Elewa AMT, ed. *Morphometrics for nonmorphometricians, lecture notes in earth sciences*. Berlin: Springer, **124:** 73–91.

Sturmbauer C, Baric S, Salzburger W, Rüber L, Verheyen E. 2001. Lake level fluctuations synchronize genetic divergences of cichlid fishes in African lakes. *Molecular Biology and Evolution* **18:** 144–154.

Swofford DL. 2003. *PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4*. Sunderland, MA: Sinauer Associates.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Resources* **22:** 4673–4680.

Vieites DR, Wollenberg KC, Andreone F, Kohler J, Glaw F, Vences M. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the Natural Academy of Sciences USA* **106:** 8267–8272.

Viscosi V, Cardini A. 2011. Leaf morphology, taxonomy and geometric morphometrics: a simplified protocol for beginners. *PLoS ONE* **6:** e25630.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Resources* **23:** 4407–4414.

Wagner CE, Harmon LJ, Seehausen O. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* **487:** 366–369.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society London B: Biological Sciences* **360:** 1847–1857.

Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T. 2008. An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources* **8:** 725–735.

Wong BBM, Rosenthal GG. 2006. Female disdain for swords in a swordtail fish. *The American Naturalist* **167:** 136–140.

Zahavi A. 1975. Mate selection – a selection for a handicap. *Journal of Theoretical Biology* **53:** 205–214.

Zelditch ML, Swiderski DL, Sheets DH, Fink WL. 2004. *Geometric morphometrics for biologists: a primer*. San Diego: Elsevier Academic Press.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Distribution map of *Chiloglanis anoterus* based on the South African Institute of Aquatic Biodiversity (SAIAB) survey data.
Figure S2. Body landmarks used in morphometric analyses.
Figure S3. Median-joining haplotype networks based on CO1 and CR data.
Figure S4. Structure analysis of AFLP data (307 loci, 117 individuals) with cluster assignment probabilities for $K = 2$–7.
Figure S5. TESS analysis of AFLP data.
Figure S6. PCA plots for *C. anoterus* populations for: CR data PC1 vs. PC2; CR data PC1 vs. PC3; CO1 data PC1 vs. PC2; CO1 data PC1 vs. PC3; AFLP data PC1 vs. PC2; and AFLP data PC2 vs. PC3.
Figure S7. Pairwise comparison of caudal fin differentiation, based on Procrustes distance (significance tested using permutation tests, 10 000 rounds).
Figure S8. Wireframe comparisons of male and female caudal fin consensus shape for four principal clades identified from the concatenated mtDNA tree in Figure 2.
Figure S9. Geometric morphometric canonical variate analysis of (a) body shape data and (b) caudal fin shape by river system.
Table S1. Taxon sampling, localities, voucher and GenBank accession numbers (M, males; F, females) used in this study.
Table S2. Adaptors, and pre-amplification and selective amplification primers used for AFLP genotyping.
Table S3. Pairwise comparisons of Φ_{PT} (population differentiation) based on AFLP data among samples grouped by cluster as inferred by Structure analysis.
Table S4. Pairwise comparisons of population differentiation among samples grouped by river system for mtDNA CO1 Φ_{ST}.
Table S5. Pairwise comparisons of population differentiation (F_{ST}) grouped by river system for mtDNA CO1 (grey) and CR markers (white).

SHARED DATA

Data deposited in the Dryad digital repository (Morris *et al*., 2015).

