Genome-wide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River

S. Elizabeth Alter (1,2,3)*, Jason Munshi-South (4), Melanie L. J. Stiassny (3,5)

(1) Department of Biology, York College/The City University of New York, 94-20 Guy R. Brewer Blvd, Jamaica NY 11451

(2) The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016

(3) The Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street and Central Park West, New York, NY 10024

(4) Louis Calder Center—Biological Field Station, Fordham University, Armonk, NY 10504

(5) Department of Ichthyology, American Museum of Natural History, 79th Street and Central Park West, New York, NY 10024

Key words: diversification, phylogeography, fish, RAD-Seq, African biogeography

*Corresponding author:

S. Elizabeth Alter

sealter@gmail.com

Phone: (650) 283-5629

Fax: (718) 262-2700

Running title: Phylogeography in rapids-adapted cichlids

Abstract:

The lower Congo River (LCR) is a freshwater biodiversity hotspot in Africa characterized by some of the world's largest rapids. However, little is known about the evolutionary forces shaping this diversity, which include numerous endemic fishes. We investigated phylogeographic relationships in *Teleogramma*, a small clade of rheophilic cichlids, in the context of regional geography and hydrology. Previous studies have been unable to resolve phylogenetic relationships within *Teleogramma* due to lack of variation in nuclear genes and discrete morphological characters among putative species. To sample more broadly across the genome we analyzed double-digest restriction-associated sequencing (ddRAD) data from 53 individuals across all described species in the genus. We also assessed body shape and mitochondrial variation within and between taxa. Phylogenetic analyses reveal previously unrecognized lineages and instances of microallopatric divergence across as little as ~1.5 km. Species ranges appear to correspond to geographic regions broadly separated by major hydrological and topographic barriers, indicating these features are likely important drivers of diversification. Mitonuclear discordance indicates one or more introgressive hybridization events, but no clear evidence of admixture is present in nuclear genomes, suggesting these events were likely ancient. A survey of female fin patterns hints that previously undetected lineage-specific patterning may be acting to reinforce species cohesion. These analyses highlight the importance of hydrological complexity in generating diversity in certain freshwater systems, as well as the utility of ddRAD-Seq data in understanding diversification processes operating both below and above the species level.

I. Introduction

Sub-Saharan Africa harbors a number of hyperdiverse freshwater systems, many of which represent threatened hotspots of endemism and are subsequently of conservation concern (Darwall *et al.* 2008; Thieme *et al.* 2005). Among these, the hydrologically complex lower Congo River (LCR) is of particular interest because it contains remarkably high levels of fish diversity and endemism for an African fluvial system yet is the proposed site for major dam development (Stiassny *et al.* 2011; Winemiller *et al.* 2016). Despite a very small footprint, representing only about 2% of the total area of the Congo Basin, the LCR contains an estimated 30% of the fish species found in the entire basin (exclusive of the Lake Tanganyikan cichlid radiations) and of these around 25% are endemic to the LCR (Lowenstein *et al.* 2011; Winemiller *et al.* 2016). Because estimates of LCR fish diversity are based almost entirely on morphological assessments, knowledge of the true diversity of this region, as well as the underlying mechanisms shaping biogeography and evolution within the system, remains extremely limited. Applying molecular methods to assess the diversity, evolutionary history and phylogeography of the exceptionally diverse LCR ichthyofauna will aid in focusing regional conservation priorities, and can improve understanding of processes underlying speciation in tropical freshwater systems.

Among fishes, cichlids have emerged as a model system for understanding mechanisms of diversification in these tropical freshwater systems (e.g., Kocher 2004; Santos & Salzburger 2012; Seehausen 2006). However, in contrast to the well-studied lacustrine radiations of the East African Great Lakes (e.g., Genner *et al.* 2015; Keller *et al.* 2013; Kocher 2004; Pauers & McKinnon 2012; Seehausen 2006; Wagner *et al.* 2013), relatively little attention has focused on investigating diversification processes in riverine cichlids. Due to high potential for in-channel connectivity in fluvial environments, fish speciation via allopatry is thought to occur primarily between isolated drainages (e.g., Burridge *et al.* 2008; Griffiths 2015; Tedesco *et al.* 2012), and the relatively few studies that have examined local-scale phylogeography of riverine cichlids have generally found weak structure within drainages (e.g., Ready *et al.* 2006a; Ready *et al.* 2006b; Koblmüller *et al.* 2012; Egger *et al.* 2015). However, rivers characterized by extreme hydrological complexity may represent an exception to this generalization. Recently, several studies have identified the unique bathymetry and hydrology of the LCR, which includes some of the most dynamic and diverse stretches of river on Earth, as a factor potentially driving allopatric diversification of endemic cichlid species (Markert *et al.* 2010; Schwarzer *et al.* 2011; Schwarzer *et al.* 2012). Whether this pattern extends to other LCR cichlids, or whether high-energy rapids impede gene flow even at very small spatial scales (microallopatry) remains unknown.

The genus *Teleogramma* represents a useful case study for investigating patterns of diversification in LCR cichlids. This small clade of rheophilic, rock-dwelling cichlids, recently placed in the chromidotilapiine radiation (Schwarzer *et al.* 2015), is endemic to the LCR region (including one species found just upstream) and another in the Kasai drainage (Fig. 1). Below

Pool Malebo, the river is characterized by a series of extremely high-energy rapids, eddies, and submerged canyons, formed as it drops 280m over 350 km en route to the Atlantic Ocean (Alter *et al.* 2015; Jackson *et al.* 2009; Oberg *et al.* 2009; Robert 1946). In marked contrast to the river upstream of Pool Malebo, the LCR channel is bedrock constrained and flows through a series of intermittently narrow (<0.2 km) and wide (>2 km) gorges in a channel that undergoes numerous changes in direction in response to a highly variable bedrock bathymetry. Littoral habitats are almost entirely rocky and rock strewn, with some intermittent sandy, or occasionally grass fringed or muddy, shorelines. *In situ* measurements have recorded dramatic changes in bathymetry even over very short distances, and this highly irregular bed topography appears to have profound effects on flow dynamics even in the absence of rapids (Jackson *et al.* 2009; Oberg *et al.* 2009). Based on a combination of geomorphology and geology (Robert 1946) divided the “rapids section” of the LCR into three main regions (see also Alter *et al.* 2015; Schwarzer *et al.* 2011). These regions, indicated in Figure 1, broadly correspond to an upper section of about 130 km from just below Pool Malebo to the region of Manyanga, characterized by a relatively small drop in elevation of over 80 m resulting in numerous stretches of surface rapids; an approximately 120 km navigable, middle section from above Manyanga to above Kinganga with few surface rapids, and a lower section of some 95 km ending at Matadi and over which the river drops nearly 170 m in elevation forming a series of massive rapids and cataracts.

Most *Teleogramma* species are notable for putatively rapids-adaptive features, such as extremely dorso-ventrally depressed heads, elongate depressed bodies, hypertrophied pelvic fins, and gas bladder reduction or loss, though the degree of rheophilic adaptation and habitat association varies across the genus (Roberts & Stewart 1976). Notably, the recently described *T. obamaorum* is restricted to deep water, rocky habitats just upstream of Pool Malebo (Stiassny & Alter 2015), and the type species of the genus, *Teleogramma gracile* Boulenger 1899, although exhibiting a number of putatively rapids-adaptive features, is found along the navigable middle section of the LCR, where it is usually associated with rocky outcrops but in the absence of high energy surface rapids.

Teleogramma are pair bonding, highly territorial, cave and crevice spawners and most species are sexually dichromatic (Wickler 1959). Females, rather than males, select spawning sites, initiate courtship and exhibit strikingly differentiated caudal and dorsal fin pigmentation patterning (Stiassny & Alter 2015; Wickler 1959). Although males attain larger adult sizes, other than female fin patterning, and belly/fin coloration during courtship, the sexes are essentially monomorphic. Of the five currently recognized species, three (*T. depressa*, *T. gracile* and *T. brichardi*) appear to occur parapatrically along sections of the LCR, one (*T. obamaorum*) occurs just above Pool Malebo in the lower middle Congo, and one (*T. monogramma*) occurs in a distant tributary (Kasai) (Fig. 1). Most species appear to have extremely narrow geographic distributions and *T. depressa* is the only species thought to occupy a relatively broad distributional range, occurring from just below Pool Malebo to the Inga rapids some 275 river kilometers downstream (Roberts & Stewart 1976). However, preliminary observations suggest extensive morphological and molecular variation across this range, and in the present study we identify two cryptic lineages within *T. depressa* sensu lato, and these are provisionally designated herein as *T.* cf*. depressa* and *T.* cf. *brichardi.*

A thorough understanding of the evolutionary and adaptive history of *Teleogramma* has been hampered by lack of genome-wide data, as traditional nuclear markers have failed to resolve key phylogenetic relationships within the genus. The biogeographic array of species along the middle-lower Congo and the persistent, high in-stream water velocities of the LCR, as it plunges from Pool Malebo towards the Atlantic, suggests a pattern of downstream serial colonization and differentiation. This hypothesis has not been tested for *Teleogramma* with nuclear data (though it has been explored in another genus of LCR cichlids (*Steatocranus*, see Discussion) (Schwarzer *et al.* 2011)). While a previous study found strong support for the monophyly of downstream LCR taxa (*gracile* and *depressa*; (Stiassny & Alter 2015) compared with those found just below and above Pool Malebo (*brichardi*, *obamaorum*, *monogramma*) these results were driven by variation in mitochondrial loci. Virtually no nuclear variation was found across species using several traditionally employed phylogenetic markers (Li *et al.* 2007). Similarly, a comparison of data from three additional nuclear markers from (Schwarzer *et al.* 2015) shows no nuclear variation between the two *Teleogramma* species for which data were available, *T. depressa* and *T. brichardi*.

Reconstructing the history of diversification and identifying lineages can be notoriously challenging for cichlid taxa, many of which have radiated recently and may also continue to share genes across species boundaries (Wagner *et al.* 2013). However, high-throughput sequencing technologies now permit unprecedented resolution in assessing genomic differentiation between taxa along stages of the species continuum, from populations that frequently exchange migrants to completely isolated species (Feder *et al.* 2012). To explore phylogeographic relationships and the evolutionary history of this unusual genus of rheophilic cichlids, we reconstructed population and phylogenetic relationships using genome-wide SNP data via double-digest restriction site associated sequencing (ddRAD-Seq). Our objectives were to elucidate the phylogeographic history of diversification in *Teleogramma* by: (a) testing species boundaries and resolving phylogenetic relationships using genome-wide SNPs; (b) assessing gene flow and connectivity between species and, where sampling allows, populations; (c) quantifying the history of introgression; and (d) measuring morphological variation within and between molecular lineages.

II. Methods

A. Sample collection and library construction

Sampling along numerous stretches of the LCR is logistically challenging and hindered by a combination of geographical isolation, limited river access, and extreme in-stream hydrological conditions. Despite intensive efforts over a ten-year period (2004-2014) there remain reaches along the LCR where no sampling has been possible. These regions are indicated in light grey in the accompanying map and we acknowledge uncertainty regarding the ranges, and extent of overlapping occurrence, of *T.* cf. *brichardi* and *T*. cf. *depressa*, and *T. gracile* and *T. depressa* (Figure 1). However, for most of the LCR, middle Congo, and Kasai Rivers we have been able to sample multiple individuals from across much of the range of all currently recognized *Teleogramma* species (Table S1). In all, we were able to sequence 56 individuals using ddRAD sequencing, including five *T. brichardi*, 15 *T.* cf. *brichardi*, three *T.* cf. *depressa*, four *T. depressa*, 11 *T. gracile*, eight *T. monogramma*, and 10 *T. obamaorum*. Geometric morphometric analysis of 74 adult individuals included 13 *T. brichardi*, 20 *T. cf. brichardi*, 11 *T. cf. depressa*, 10 *T. depressa*, 10 *T. gracile*, five *T. monogramma*, and five *T. obamaorum*.

All specimens were euthanized prior to preservation in accordance with the recommended guidelines for the use of fishes in research (Nickum 2004), and stress was ameliorated by minimizing handling and through the use of anesthetics (MS-222) for euthanasia. All specimens are cataloged and stored in the ichthyology collection of the American Museum of Natural History (AMNH), available online at

http://entheros.amnh,org/db/emuwebamnh/index.php. Tissue samples were taken in the field and immediately preserved in 95% ethanol and the voucher specimens fixed in formalin and later transferred to 70% ethanol for long-term storage. We extracted total genomic DNA from fin clips using the Gentra PureGene tissue extraction protocol (Qiagen). The total amount of genomic DNA was determined using a Qubit dsDNA HS Assay (Life Technologies), and only samples with >500 ng total DNA were used in the rest of the analyses.

We prepared double-digest RAD-Seq libraries for 56 individuals using the protocol outlined in (Peterson *et al.* 2012). In brief, for each sample, 500-1,000 ng of DNA was digested using two restriction enzymes (*MluCI* and *NlaIII*, New England Bioloabs Inc), and adapters with sample-specific barcodes and Illumina multiplexing read indices were ligated to individual samples. Samples with unique adapters were then pooled, and each pool was size-selected for fragments in the range 300-450 bp using a Pippin Prep system (Sage Sciences Inc). Size-selected pooled samples were PCR-amplified using a Phusion polymerase kit and primers that introduce Illumina multiplexing indices (Peterson *et al.* 2012). The final pooled samples were then purified using AMPure XP beads (Beckman Coulter Inc), quantified using qPCR, and sequenced using paired-end sequencing on an Illumina HiSeq 2500 across a single lane at the New York Genome Center (NYGC, New York, NY). A custom script at the NYGC was then used to demultiplex the sequence reads for each sample based on the combination of sample-specific barcode and Illumina index.

B. SNP calling and population genomic statistics

We used Stacks 1.35 to filter and call SNPs from the demultiplexed sequence reads using pipelines for both de novo and reference-aligned data (Catchen et al. 2013a). Stacks calculates the likelihood of the two most common genotypes at each site given a maximum likelihood estimate of the sequencing error rate, and then uses a likelihood ratio test to identify the most likely genotype at that position. Full details of the models underlying Stacks are given in Catchen et al. (2013a), but we describe our general approach below.

First, we used the *process_radtags* script to filter reads with low quality scores or uncalled bases, and truncate all reads to 92 bp. We then used bowtie2 (Langmead & Salzberg 2012) with default settings to align each sample's paired end reads to the Nile tilapia (*Oreochromis niloticus*) reference genome, version OreNil1.1 (NCBI Assembly

GCA_000188235.2). All resulting SAM (Sequence Alignment/Map) files were run through the *ref_map.pl* pipeline in Stacks to identify ddRAD loci and call SNPs. Default settings for the pipeline were used, except for the minimum number of identical reads required to create a 'stack' (m = 10) and the number of mismatches allowed when building the catalog (n = 2). After building the initial catalog of loci, the *populations* script in Stacks was used to filter loci for those that occurred in six out of seven taxa (p = 6), in at least 50% of individuals for each taxon (r = 0.5), and had a minor allele frequency of at least 0.05 (min_maf = 0.05).

Although *Teleogramma* and *Oreochromis* are both members of the same large cichlid subfamily (Pseudocrenilabrinae), they are distantly related and divergence time estimates for these two genera exceed 46 MYA (Schwarzer *et al.* 2011). Thus, using the *O. niloticus* reference genome may have limited our ability to identify SNPs due to a lack of homology between the reference and our study taxa. For comparison with the reference-aligned analysis, we also generated *de novo* SNP calls in Stacks using *denovo_map.pl*. First, we concatenated the forward and reverse reads for each individual into one fastq file because the two paired reads did not overlap and were not aligned. The settings were the same as above except for an additional parameter, i.e. the number of mismatches allowed between loci for a single individual (M = 3).

For both reference-aligned and *de novo* pipelines, the *populations* script in Stacks was used to produce genotype output in multiple formats (i.e. VCF, Genepop, PLINK, Structure, and TreeMix) with one SNP from each locus (--write_single_SNP), and generated summary statistics such as observed heterozygosity, nucleotide diversity, and pairwise F_{ST} between all populations.

We loaded the RAD loci and individual data from Stacks into a MySQL database and visualized the output using the Stacks webserver. Based on the results of the first pipeline runs, we removed three of the 56 individuals because they were genotyped at less than 2,500 SNPs and thus not comparable to the other samples. The mean and range of loci obtained for each sample are given in Table S2. We then reran the pipelines as above with 53 individuals, including two *T. brichardi*, 15 *T.* cf. *brichardi*, three *T.* cf. *depressa*, four *T. depressa*, 11 *T. gracile*, eight *T. monogramma*, and 10 *T. obamaorum*.

C. ddRAD-seq data analyses

Phylogenomic analyses

We estimated phylogenies using two methods based on the SNP-only dataset (invariant sites removed): coalescent-based species tree estimation using the methods of (Bryant *et al.* 2012) and (Chifman & Kubatko 2014). First, we estimated the species tree using SNAPP (Bryant *et al.* 2012) implemented in BEAST v2.2.1 (Bouckaert *et al.* 2014). Because of the computationally intensive nature of species tree estimation, we selected a subset of individuals with the highest numbers of reads and SNP genotypes (n=25; Table S2), with at least two representatives of each species and, based on preliminary results, representatives from across the full geographic range of *T. depressa* (*sensu* Roberts & Stewart, 1976). To assess the effects of variable amounts of missing data on tree topologies, we conducted preliminary runs with 11,459 SNPs (6 or fewer missing genotypes) to 37,826 SNPs (10 or fewer missing genotypes). Species tree estimation was performed on bi-allelic SNP data (37,826 SNPs) in SNAPP v. 1.1.4 (Bryant *et al.* 2012), using gamma priors with a shape parameter of 2, and a scale parameter of 2000, with θ=0.001). SNAPP employs a Yule prior for the species tree and branch lengths (Bouckaert *et al.* 2014) and we used a birth rate (λ) = 0.00765 (Bryant *et al.* 2012). Mutation rates (u and v) were estimated from the data based on equations from (Drummond & Bouckaert 2015) and were fixed at *u*=1.4501 and v=0.8611. Runs were performed with a chain length of 5 million generations, with the first 10% of generations as burn-in. We determined the burn-in, assessed convergence and ensured that effective sample size (ESS)>200 for all model parameters by examining likelihood plots and ESS values using Tracer v1.5 (Rambaut &

Drummond 2007). TreeAnnotator (Rambaut and Drummond 2008) was used to find the Maximum Clade Credibility tree and to estimate posterior probabilities.

In addition to SNAPP, we also estimated a species tree using the program SVDquartets (Chifman & Kubatko 2014), which is based on a coalescent model. This method subsamples four taxa (quartets) at a time from the data matrix, calculates a singular decomposition value for each, and reconstructs the species tree from quartets. Uncertainty is quantified in this method using nonparametric bootstrapping. From our reduced dataset of 25 individuals, we randomly sampled 100,000 quartets using nonparametric bootstrapping with 500 replicates.

Population genomic analyses

To investigate genetic differentiation between *Teleogramma* taxa, we used the model-based evolutionary clustering approach in ADMIXTURE (Alexander et al. 2009) to identify evolutionary clusters and potential introgression. ADMIXTURE computes maximum likelihood estimates of parameters to estimate the most likely number of evolutionary clusters, K. We ran 20 replicates of ADMIXTURE for each value of K from 1 – 15 using SNPs genotyped for the 53 *Teleogramma* individuals described above. The most likely values of K were determined in ADMIXTURE based on minimization of cross-validation error across replicate runs (Alexander and Lange 2011).

For comparison with ADMIXTURE results, we also conducted a principal components analysis (PCA, (Patterson *et al.* 2006)) using the LEA package in R (Frichot and Francois, submitted) on 53 *Teleogramma* individuals. We identified the number of significant principal components using Tracy-Widom tests in LEA, and generated scatterplots using ggplot2 in R.

Both reference-aligned and *de novo* SNP genotypes were analyzed using ADMIXTURE and PCA. We also conducted these analyses on a subset of 17 individuals identified as *T. brichardi* or *T.* cf. *brichardi* to investigate fine-scale structure. These individuals were placed in one of five groups based on sampling locality. We then used the *populations* script from Stacks to filter SNPs for these individuals as above, except SNPs were retained when they occurred in at least three out of five sampling sites, 50% of individuals from each site, and had a minor allele frequency of at least 0.05.

Population introgression

In addition to ADMIXTURE, we assessed gene flow between taxa using two tree-based methods. First, we used Treemix (Pickrell & Pritchard 2012) to assess ancestral admixture between lineages (*N* = 53 individuals). This method first builds a maximum likelihood (ML) phylogeny and subsequently models migration between branches to determine whether migration/admixture events improve the likelihood fit. We computed a basic ML tree using all SNPs independently, and using windows of 500 SNPs grouped together to account for linkage disequilibrium. The two trees were identical so further Treemix analysis used all of the SNPs. We then built ML trees that included from one to four migration events, and used the get_f() script that accompanies Treemix to calculate the percent variance explained by models with different numbers of migration events. To formally test for admixture between *Teleogramma* spp., we used the three-population test (Reich *et al.* 2009) included with Treemix (*threepop* script). In this test, the f3 (X; A,B) statistic is negative when a population X does not form a simple tree with populations A and B, but rather may be a mixture of A and B.

To determine incompatible or ambiguous phylogenetic signal within the dataset, we also used a NeighborNet algorithm (Bryant & Moulton 2004) implemented in Splitstree (4.13.1) (Huson & Bryant 2005). Split network relationships were estimated using uncorrected p-distances with the equal angle algorithm.

Mitochondrial data analyses

We amplified and sequenced two mitochondrial markers: cytochrome oxidase I (COI) and NADH dehydrogenase subunit 2 (ND2) (Kocher *et al.* 1995) across the same representative individuals of all putative *Teleogramma* species, as well as several additional specimens of *T. brichardi*. We used *Chromidotilapia sensu stricto* (*Chromidotilapia kingsleyae*) as an outgroup based on the phylogenetic hypothesis of (Schwarzer *et al.* 2015). The COI marker was amplified using primers VF2_t1, FishF2_t1, FishR2_t1, and FR1d_t1 (Ivanova *et al.* 2007) with the following amplification conditions: 94°C for 2 min, 35 cycles of 94°C for 30 s, 52°C for 40 s, and 72°C for 1 min, with a final extension at 72°C for 10 min. For the ND2 marker, we used primers ND2Trp and ND2Met (Kocher *et al.* 1995) with the following amplification conditions: 94°C for 2 min, 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 10 min. Successful amplifications were sequenced on an ABI 3730 at Genewiz, Inc (South Plainfield, NJ). We aligned and edited forward and reverse chromatograms in Geneious R7.0 (Biomatters Ltd., Aukland, NZ) and MUSCLE v3.5 (Edgar 2004). We determined the best-fit models of nucleotide evolution with the program jModelTest (Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden (Posada 2008)). We performed Bayesian phylogenetic inference on the concatenated loci using Markov Chain Monte Carlo (MCMC) methods in BEAST v2.3 (Bouckaert *et al.* 2014) using 15 million generations with trees sampled every 1000 generations, and a burn-in of 25%, using a lognormal distributed relaxed molecular clock model. As a tree prior we used the calibrated Yule (Heled & Drummond 2011) (xml file available as Supplementary File 1). In order to estimate divergence times between mitochondrial clades, we used the estimated split between *Teleogramma* and *Chromidotilapia* from (Schwarzer *et al.* 2015) of 31.3 (24.3-39.7) Myr with a normal distribution prior, based on the minimum age estimate for *Oreochromis lorenzoi*† from the early Miocene and non-cichlid fossils. While normal distribution priors are not always optimal for use with primary fossil calibrations, they can be appropriate for secondary calibrations as employed here (Ho & Phillips 2009). We checked convergence of parameters using Tracer v.1.5 (Rambaut & Drummond 2007) to assess mixing and effective sample size (ESS) >200 for all model parameters.

E. Morphological analyses

In the absence of evidence of qualitative character differentiation between taxa, geometric mophometrics were used to compare variation in overall head and anterior body shape in individuals of each putative species or population. Specimens were pinned flat, and photographed on the left side with a Nikon Digital SLR camera and 60 mm f/2.8 AF Micro-Nikkor lens. Shape variation was assessed with a set of 8 fixed landmarks and the landmark configuration for each specimen was captured using tpsDIG2 v2.17 (Rohlf 2013) and subsequently translated, scaled to unit centroid size and rotated with Generalized Procrustes Analysis (GPA) in MorphoJ (Klingenberg 2011). After applying GPA, residual disparity among configurations depends entirely on shape variation and consensus configurations or mean shapes are obtained. Aligned shape coordinates were analyzed with Principal Component Analysis (PCA) in MorphoJ in order to summarize major axes of shape variation. For qualitative examination of female fin patterning a minimum of 10 female specimens per taxon were examined, with the exception of *T. cf. depressa* for which only 4 female specimens were available.

III. Results

A. Sequencing and SNP calling

Illumina sequencing of *Teleogramma* ddRAD libraries for 53 individuals resulted in a total of 335.64 million reads after initial filtering for quality. Numbers of reads per individual ranged from 731K – 14.3M, with a mean ± SD = 6.33M ± 2.95M. Mean depth of coverage for each individual ranged from 19X – 74X (mean = 24.3X) for the reference-aligned data. Additional RAD sequencing statistics are provided in Table S2.

The number of nucleotide positions that were polymorphic in at least one taxon ranged from 20.6K – 24.2K for the reference-aligned STACKS pipeline (Table 1), and from 16.7K – 20.2K for the *de novo* pipeline (Table S3). All taxa exhibited similar levels of genetic variation despite variation in sample size, with the exception of consistently lower heterozygosity and nucleotide diversity in *T. monogramma* (Table 1). Nucleotide diversity (0.067-0.131, Table 1) is comparable to results from other isolated fish populations, e.g. stickleback (0.023-0.079)(Catchen *et al.* 2013b). While differences in number of individuals across taxa could potentially have a large effect on some summary statistics (e.g., count-based statistics such as private alleles), we expect that large numbers of loci as obtained from genome-wide RADseq data can yield accurate estimates of overall nucleotide diversity and heterozygosity from even a small number of individuals, unless those individuals were unusually inbred (e.g., Nei 1978). Results for downstream population genomic analyses were qualitatively similar for output from the two pipelines, so we primarily present analyses of reference-aligned data herein, and analyses of *de novo* data in the supplementary material.

Phylogenetic analysis

The phylogenies constructed with Bayesian MCMC inference with different amounts of missing data were consistent in all cases. Our final analysis used 37,826 SNPs and shows maximum support for all current species designations (*T. monogramma*, *T. obamaorum*, *T. gracile*, *T. depressa* sensu stricto, *T. brichardi*) as well as for two morphologically cryptic lineages recognized herein as *T*. cf. *depressa* and *T*. cf. *brichardi* (Figure 2). The analysis also provides strong support for monophyly of a clade containing all middle and lower Congo species, indicating that the geographically isolated Kasai River endemic, *T. monogramma,* is sister to the remaining taxa. While the branching order of *T. gracile* and *T. obamaorum* is not strongly supported, all analyses (using different numbers of SNPs) produced the same topology. As noted previously, not all specimens originally designated as *T. depressa* form a monophyletic group; specifically, a group of individuals sampled from the upper section of the LCR (Kinsuka-Les Rapides to Foulakari) form a clade with the geographically proximate *T. brichardi* with strong support. Of the remaining individuals originally designated as *T. depressa*, those collected from Banda Nyenge and vicinity are reciprocally monophyletic with *T. depressa* from the Kinganga-Inga region. SVDquartets recovered the same topology as the SNAPP analysis, but with variable bootstrap support (Figure S1, Supporting Information).

From two mitochondrial loci (ND2 and COI), we recovered well-supported relationships across most clades in Bayesian analyses, but with a markedly different topology than that of the nuclear tree (Figure 3): unlike the genome wide nuclear signal, the mitochondrial signal places *T. gracile* within the *T. depressa + T.* cf. *depressa clade*, and *T. brichardi* falls outside the rest of the LCR clade (Figure 3). Divergence dating of mitochondrial clades indicates that the split between the *brichardi* mitochondrial lineage and the remainder of the LCR lineages occurred ~4.7 Myr (95% HPDI: 1.6-7.6 Myr) and that diversification of mitochondrial lineages in the remaining LCR clade (*T. gracile*, *T. depressa*, *T.* cf. *depressa*, *T.* cf. *brichardi*) occurred much more recently at ~1.4 Myr (95% HPDI: 0.4-2.5 Myr). The MRCA for the entire genus was dated as 5.3 Myr (95% HPDI: 1.9-8.8 Myr). It is important to note that, because of mitonuclear discordance, these dates correspond to the ages of mitochondrial lineages rather than species. In addition, molecular clock analyses are, in general, sensitive to selection and placement of calibration points, clock models selected, and taxa sampled, and thus the age estimates presented here should be considered tentative in light of ongoing debates about macroevolutionary timescales in cichlid radiations e.g. (Friedman *et al.* 2013; López-Fernández *et al.* 2013). However, the dates obtained are consistent with divergence time estimates for other LCR fishes, including mastacembelid eels (Alter *et al.* 2015) and two other cichlid clades (Schwarzer *et al.* 2011) within the last 5 Myr. These studies have hypothesized that similarly timed diversification events in independent clades have likely been driven by the same geological changes that resulted in the formation of the LCR's high-energy flow regime beginning roughly 5 Myr (Alter *et al.* 2015; Schwarzer *et al.* 2011).

Population structure and migration

We assessed population structure and migration using model-based clustering (ADMIXTURE), non-model based genetic PCA, and tree-based methods. We found that ADMIXTURE's CV error was minimized for K = 5 for reference-aligned SNPs and K = 6 for *de novo* SNPs (Figure S2, Supporting Information). Both analyses identified all taxa as separate evolutionary clusters with little to no admixture, with the exception of gene flow between *T. brichardi*-*T.* cf. *brichardi* (Figure 4a,b). For K=5, *T.brichardi*-*T.* cf. *brichardi* and *T.depressa*-*T.* cf. *depressa* are clustered together, consistent with phylogenetic results (Figure 4a). Tracy-Widom statistics were significant ($P < 0.01$) for the first eight PCs for the genetic PCA analysis, but most of the variation (67.5%) was captured by the first four PCs (Figure S3, Table S4, Supporting Information). PCA patterns were broadly similar to the clustering identified by ADMIXTURE. The first two PCs separated out all taxa except for *T. brichardi*-*T.* cf. *brichardi* and *T. depressa*-*T.* cf. *depressa* (Figure 4c). *T. depressa* and *T*. cf. *depressa* are clearly separated along PC 4; *T. brichardi* and *T.* cf. *brichardi* do not overlap but are close together (Figure 4d). These same patterns are reflected in pairwise F_{ST} values, with nearly all of the lowest values calculated for pairs of *T. brichardi, T.* cf. *brichardi*, *T. depressa*, and/or *T*. cf. *depressa* (F_{ST} = 0.137 – 0.397) and higher values for all other pairwise comparisons (F_{ST} = 0.363 – 0.605; Table S5).

We also applied ADMIXTURE to a restricted set of samples containing only *T. brichardi* and *T*. cf. *brichardi* individuals to determine population structure and gene flow between different sampling areas. While K = 1 had the lowest CV error, K = 2 had the second-lowest CV error and groups the three left bank sampling sites together (*T. brichardi* (Kinsuka), *T*. cf.

brichardi-Kinsuka, and Mbudi/Ngombe, Democratic Republic of Congo) and the two right bank sites together (Les Rapides, Foulakari, Republic of Congo), with some admixture shown between *T.* cf. *brichardi* at Mbudi/Ngombe and the right bank cluster (Figure S4). This pattern suggests that the massive, cross-bank rapids between left and right banks at Kinsuka and Les Rapides (the Regina or Kintamo Falls) likely played a role in restricting gene flow between these locations. In contrast, some 10 km downstream, in the region of Mbudi and Ngombe, surface rapids are less prevalent and the river channel narrows markedly, such that during periods of low water the distance between the two banks can be less than 0.25 km. Values of F_{ST} (Table S6, Supporting Information) range between 0.08-0.17 across all sampled sites, likewise indicating genetic structure across populations of *T*. cf. *brichardi* (though some pairwise F_{ST} values were not significant, perhaps due to limited sample sizes). This pattern differs somewhat from the results of the full-dataset ADMIXTURE analysis, in that some cf. *brichardi* populations cluster with *T. brichardi* rather than other cf. *brichardi* populations. We interpret this difference as resulting from applying ADMIXTURE's generative model, which assumes Hardy-Weinberg equilibrium between alleles and linkage equilibrium between loci, to a restricted dataset with a much higher extent of shared variation, and to populations for which gene flow has not been constant over time nor homogenous across the genome. Whereas in the full dataset, the model components corresponding to K=5 explain the majority of genetic structure underlying the data, the restricted dataset reveals additional shared genetic variation between *T. brichardi* and left-bank populations of *T.* cf. *brichardi*.

Results from TREEMIX (Figure 5) identified two migration events: 1) *T. obamaorum* --> *T. brichardi*; and 2) *T. gracile* --> *T. depressa*. However, neither event significantly improved the likelihood of the estimated tree model. From the three-population test of admixture, all f3 values were positive with associated Z-scores > 9.0 in all cases, thus indicating no evidence of admixture between any of the seven taxa (Table S7, Supporting Information).

Finally, we applied the NeighborNet algorithm (Bryant & Moulton 2004) to assess the presence of conflicting phylogenetic signal in the genome wide dataset, which can indicate a history of hybridization in the context of sexually reproducing organisms. The resulting splits graph (Figure S5, Supporting Information) indicates the most conflicting phylogenetic signal to be that between *T*. *brichardi* and *T*. cf. *brichardi*, with lower levels of conflicting splits/ambiguity between *T*. *depressa* and *T.* cf. *depressa*, and between the latter clade and *T*. *gracile*.

Morphology and Morphometrics

Morphometric analysis indicates high levels of intralineal variation across the genus (Figure 6). As expected, based on previous taxonomic descriptions of these species, most variation is associated with the degree of dorso-ventral depression (Fig. 6b), but with the exception of *T. obamaorum*, no significant separation of genetically recognized lineages was detected in this analysis. Notably, *T.* cf. *brichardi* appears to be the most morphologically variable taxon, spanning the range of shape variation exhibited by most other taxa, and also has the largest known geographic distribution within the genus; however, given that the sample size was largest for *T.* cf. *brichardi*, analysis of additional specimens of other *Teleogramma* taxa is needed to confirm this finding. Despite lack of clear separation between lineages, the overall morphometric pattern reflected in the PCA is broadly consistent with genetically determined taxonomic assignments of individuals.

In addition to body shape, various combinations of meristic features have also been employed to distinguish between *Teleogramma* species (Roberts & Stewart 1976; Stiassny & Alter 2015). While most of these meristic attributes (e.g. fin spine and ray numbers, vertebral numbers, scale counts etc.) are broadly overlapping in range across the genus, most noteworthy in the present context, is the presence in *T. monogramma*, *T. obamaorum* and, uniquely in *T. gracile* among LCR species, of a markedly reduced number of pored scales along the lateral line when compared to all other congeners. The low lateral line count of 32-36 large, pore bearing scales in these three species contrasts with a range of 46-60 markedly smaller scales in all of the other LCR lineages. Unfortunately, as *Teleogramma* is highly unusual among cichlids, and all other chromidotilapiines, in possessing a single lateral line (versus an interrupted lateral line series), polarity assessment of these alternative configurations is problematical.

Despite a lack of clearly diagnostic morphometric features, we note that genetic lineages are distinguishable based on differences of female fin patterning and coloration. Differences in patterning are evident in the posterior-most regions of the dorsal fin and over the dorsal field of the caudal fin, and these are illustrated photographically and schematically in Figure 7. With the exception of *T. obamaorum* (Figure 7G), in which the sexes are monomorphic, all female *Teleogramma* are characterized by the presence of a white flag (flushed red during reproductive activity) on the upper section of the caudal fin and extending, to varying degrees, along the dorsal fin. The extent and width of the white caudal flag, combined with the extent and width of a black margin on both the caudal and dorsal fin appears to characterize each lineage (Figure 7A-F). This seemingly species/lineage specific patterning is consistent across all specimens we have examined to date. As with similarity in lateral line counts, we note a strong similarity between the fin patterning of *T. gracile* (Figure 7E) and *T. monogramma* (Figure 7F). In both of these species patterning is characterized by a marked reduction of the width of the white flag and an expansion of the black margin over both the dorsal and caudal fins.

IV. Discussion

Riverine cichlids represent an important component of cichlid morphological and genetic diversity (Loh *et al.* 2013; Brawand *et al.* 2014; Stiassny and Alter 2015), but few studies have considered both genome-wide data and quantitative morphological data in African riverine lineages in order to investigate the factors potentially responsible for differentiation. Most previous studies of riverine cichlid diversification have revealed relatively large-scale allopatric divergence driven primarily by geological events resulting in watershed rearrangements, flow reversals, and headwater captures (e.g., Ready *et al.* 2006a; Ready *et al.* 2006b; Koblmüller *et al.* 2012; Egger *et al.* 2015). In contrast, few studies have identified such fine-scale, phylogeographic structure at highly localized, within-channel levels. Here, we present both densely sampled genome-wide data and morphological analysis for all known species in the riverine cichlid genus *Teleogramma*, indicating that microallopatric processes driven by the extreme hydrology of the LCR have likely shaped, and are continuing to shape, the evolutionary history of this group. These findings are of particular importance in light of the proposed development of a mega-dam (Grand Inga) that would massively impact the flow regime of the LCR (Showers 2011).

Species designations and cryptic lineages

Our SNP data strongly support most existing species designations, but also reveal the presence of two cryptic lineages within the LCR clade. Specifically, these genome-wide data indicate that *T. depressa* (sensu Roberts & Stewart 1976) is non-monophyletic and composed of three geographically and genetically distinct lineages. Individuals from the upper sections of the LCR (*T*. cf. *brichardi*) comprise a lineage that is resolved as sister to the geographically proximate Kinsuka endemic, *T. brichardi* (Figures 1, 2). Individuals collected around Banda Nyenge, at the boundary between the upper and middle sections of the LCR (*T.* cf. *depressa,* Figures 1, 2), are distinct from, yet likely reciprocally monophyletic with *T. depressa* sensu stricto, a species restricted herein to populations found in the lower section of the LCR in the region of Kinganga to Inga. These findings indicate that *T. depressa* has a much smaller range than previously thought, which is of particular concern given that its habitat overlaps nearly entirely with the area of the LCR proposed for mega-dam development in the Inga region (Showers 2011).

Although distinct based on the single morphological criterion of discernible differences in female tail patterning (comparing Figures 7B,C and 7A,D) no other morphometric (Figure 6A), meristic, or qualitative anatomical features have yet been found to differentiate these cryptic lineages from their putative sister species. This in conjunction with the lower genetic distance between each of these cryptic lineages and related congeners, as reflected in ADMIXTURE results and F_{ST} values (*T*. cf. *brichardi vs T. brichardi* F_{ST}=0.137, *T.* cf. *depressa vs T. depressa* F_{ST}=0.274, compared with described species, Table S5, Supporting Information) suggests that these lineages may not yet have fully differentiated genetically or morphologically. However, small sample sizes for several of the taxa preclude a definitive conclusion about the extent of differentiation. In light of these factors, and given the presence of hydrological barriers to dispersal, we hypothesize that these lineages may represent incipient species. Pending a detailed taxonomic treatment based on additional specimens (Stiassny & Alter in prep.), we recommend the adoption of the informal designations *T.* cf. *brichardi* and *T*. cf. *depressa* for these populations, thereby acknowledging their distinction from the related *T. brichardi* and *T. depressa*.

Previous studies have been unable to resolve phylogenetic relationships within *Teleogramma* using either nuclear genetic or morphological data (Stiassny & Alter 2015). The genome-wide data analyzed here indicate that the extreme bathymetry and hydrology of the LCR appear to have played a central role in structuring diversification, but as was found in a previous study of other LCR cichlid genera (Schwarzer *et al.* 2011), the simple null hypothesis of sequential downstream colonization is not supported. *T. monogramma* is resolved as the sister species of the middle-lower Congo clade, as expected based on its distant, non-contiguous range relative to congenerics, and the majority of LCR endemic species are supported as forming a monophyletic clade. However, the placement of *T. gracile* outside of this clade is interesting, given its downstream LCR location (Figure 1). To date, *T. gracile* has been found only in slower flowing, middle sections of the LCR, with a distribution bordered by, and partially overlapping with *T. cf. depressa* at the upstream limit of its range and possibly by *T. depressa* at the downstream limit of its range (Figure 1). A similar complex pattern was identified in the two other LCR cichlid genera that have been examined: *Steatocranus*, in which some older lineages (*S.* sp. aff. *casuarius* and *S.* sp. aff. *casuarius* "brown pearl") are found in the lower LCR; and *Nanochromis*, in which the older lineage *N. minor* is found in the middle LCR (Schwarzer et al. 2011). Divergence age estimates between these *Steatocranus* lineages and their sister taxa from central Congo tributaries (5.3 [4.1-6.7] Mya) are consistent with our estimated divergence between *Teleogramma* found in central Congo tributaries and the LCR endemics (5.3 [1.9-8.8] Mya), however, age estimates for central Congo *Nanochromis* species and *N. minor* (8.36 [6.5-10.4] Mya) are somewhat older.

Mitonuclear discordance and introgression.

When compared with the mitochondrial phylogeny, the nuclear tree shows several areas of agreement and a few areas of strong discord (Figure 2, 3). In addition to a differing placement for *T. gracile*, the mitochondrial tree places *T. brichardi* outside the LCR clade, and divergence dating indicates the split between the *brichardi* mitochondrial lineage and the remaining LCR lineages is old relative to subsequent mtDNA diversification events. Such incongruence between the mitochondrial and nuclear phylogenies has been interpreted as a result of hybridization events in other cichlid taxa (Koblmüller *et al.* 2007; Nevado *et al.* 2009; Schwarzer et al. 2011, 2012a,b). In the case of *Teleogramma*, mitonuclear incongruence strongly suggests post-divergence introgressive hybridization between *T. depressa* and *T. gracile*, and possibly between *T. brichardi* and an upstream lineage such as *T. obamaorum.* While the complexity of introgression is not as high as in another LCR cichlid examined, *Steatocranus* (Schwartzer et al. 2011), patterns in *Teleogramma* show some similarities. In particular, in both genera, mitonuclear discordance is evident between species found on either side of Pool Malebo.

An alternative explanation for conflicting mitochondrial and nuclear signals is incomplete lineage sorting (ILS), usually associated with rapid bursts of cladogenesis and short internode distances (eg. (Takahashi *et al.* 2001; Seehausen 2006). As such, ILS is expected primarily to shape patterns across recently diverged taxa, rather than older divergences, though ancient ILS has also been inferred from relatively old cichlid speciation bursts, such as those in Lake Tanganikya (Takahashi *et al.* 2001; Koblmüller *et al.* 2010). In addition, effects are expected to be distributed randomly across taxa (Maddison and Knowles 2006). As the radiation within *Teleogramma* shows little evidence of rapid cladogenesis with relatively long internal branches, and particularly as patterns of mitonuclear discordance are not spread randomly across the tree but rather align with geography, we interpret the likely source of this mitonuclear discordance as past hybridization in which mtDNA was transferred between lineages, rather than ILS.

Despite evidence of ancient hybridization from mitonuclear discordance, little trace of these events remains in the nuclear data based on ADMIXTURE and Treemix results. It is notable that results from Treemix did recover two episodes of migration (identified as maximum residuals across a full residual covariance matrix) between *depressa*-*gracile* and *brichardi*-*obamaorum* (Figure 5), though neither modeled event significantly improved the likelihood of the tree. However, the lineages identified as having undergone migration correspond to those for which we observe mitonuclear discordance, suggesting in this case there may remain a weak signal of introgression in the nuclear genome. These events are not apparent from ADMIXTURE results, which show no evidence of migration between these taxa. This may result from the limitations of ADMIXTURE to detect ancient hybridization since the method does not explicitly fit a historical model, and relies on simplified population genetic hypotheses such as no genetic drift and Hardy-Weinberg equilibrium in ancestral populations (Alexander *et al.* 2009). Similarly, the three-population test, while a formal test of admixture, can produce false negatives when one of the populations has undergone strong population-specific drift (Patterson *et al.* 2012). Despite these provisos, overall our analyses support a scenario of ancient, perhaps asymmetric, introgression followed by population splits with little ongoing gene flow. Such mitochondrial capture or replacement events accompanied by low amounts of, or no detectable, nuclear introgression have been reported in a number of other cichlids (Nevado *et al.* 2009; Schwarzer *et al.* 2012; Willis *et al.* 2014) and many other taxa (reviewed by Toews and Brelsford 2012). Notably, mitochondrial capture was detected in a clade of LCR endemic haplochromine cichlids ("*Haplochromis*" *demeusii* and "*H.*" *fasciatus*), which carry mitochondrial haplotypes closely related to those of East African haplochromines, although nuclear DNA (AFLPs) show no evidence of introgression (Schwarzer *et al.* 2012).

Hydrological complexity and microallopatric divergence

Previous studies of LCR cichlids have identified the hydrological complexity of the LCR as a major factor affecting the biogeography of species (Markert *et al.* 2010; Schwarzer *et al.* 2011), and hydrological features appear to be important in structuring both species and populations in *Teleogramma*. The majority of lineages (*T. brichardi*, *T. obamaorum*, *T. depressa*, *T.* cf. *depressa*, *T. gracile*) have narrow geographic distributions broadly corresponding to hydrological features or regions. While our sampling does not permit a comprehensive assessment of population structure within each species, we do find genetic structure between populations of *T.* cf. *brichardi* from the five sites sampled, also separated by a series of hydrological barriers. Values of F_{ST} (Table S6, Supporting Information) support the hypothesis that large rapids impede cross-channel gene flow between Kinsuka (left bank) and Les Rapides (right bank), whereas gene flow is higher at same-bank sites (Kinsuka/Mbudi-Ngombe) separated by a considerably greater distance, but without large intervening rapids, compared with cross-channel sites. Likewise, Markert *et al.* (2010) analyzed genetic differentiation between samples of *T. depressa* sensu stricto from three locales in the lower section of the LCR, and found a high degree of isolation between individuals sampled at Kinganga and the Inga region, located on either side of the large rapids at Isangila and Fwamalo. These rapids are notable due to the narrowness of the river channel at this point (0.3-0.8km), causing white water to extend across the entire channel at these locations.

History of colonization and diversification

The combination of genomic, morphological and biogeographic data suggests a scenario of diversification accompanied by ancient introgression in *Teleogramma*; however, the data do not support our original hypothesis of downstream sequential diversification for this genus. The phylogenetic pattern we observe from genome-wide SNPs suggests the possibility of two independent colonizations of the LCR by *Teleogramma* by an upstream ancestor: an initial colonization by *T. gracile*, which is the only LCR species that occurs in regions with no large surface rapids, followed by the subsequent arrival of the ancestor of the highly rapids associated *T. depressa*, *T.* cf. depressa, *T. brichardi* and *T.* cf. *brichardi*. This scenario is supported by the phylogenetic position of *T. gracile* outside the remaining LCR clade and its relatively old age based on branch length when compared with its LCR congeners (Figure 2). Additionally, *T. gracile* shares with *T. obamaorum* and *T. monogramma* a similar lateral line configuration, and with *T. monogramma* a strikingly similar fin patterning (Stiassny & Alter 2015). In these characteristics *T. gracile* is seemingly more similar to these upstream congeners than to its geographic neighbors in the LCR (*T. depressa*, *T.* cf. *depressa*). However, particularly given the relatively weak support for the node defining the phylogenetic position of *T. gracile*, we cannot yet rule out an alternative scenario in which *Teleogramma* arrived in the LCR once, diversified throughout the system, and experienced subsequent range shifts that led to the current biogeographic distribution. While both scenarios remain speculative, multiple independent colonization events have been supported in a number of other LCR fish clades, including two cichlid genera (*Steatocranus* and *Nanochromis*, (Schwarzer *et al.* 2011)) and mastacembelid eels (Alter *et al.* 2015), perhaps resulting from watershed rearrangements and onset of favorable ecological conditions that may have occurred at least twice over the last ~5 Myr. Notably, the median ages of endemic LCR taxa estimated for *Nanochromis* (1.6-2.67 Myr) and *Steatocranus* (0.94-4.48 Myr) (Schwarzer *et al.* 2011) are consistent with the dates we have estimated for *Teleogramma* from mitochondrial data, indicating the majority of these lineages diversified in the LCR over the last few million years.

Hybridization has been extensively documented in numerous cichlid lineages (Ford *et al.* 2015; Keller *et al.* 2013; Wagner *et al.* 2013; Willis *et al.* 2014), and based on the mitonuclear discordance documented in this study, has also occurred in the past in *Teleogramma*. It is therefore noteworthy that we find relatively little evidence for ongoing hybridization among contemporary species, even those living in close proximity and in the absence of obvious hydrological barriers. While most species are not strongly morphologically differentiated (Figure 6) nor, with the exceptions of *gracile* and *obamaorum*, ecologically, we do observe consistent differences in female fin patterning in all genetically-defined lineages (Figure 7). We speculate that these lineage-specific patterns may potentially act to reinforce boundaries between clades. As noted previously, *Teleogramma* are pair bonding, territorial, cave and crevice spawners and the females, rather than males, select spawning sites, and initiate courtship. Direct observation of reproduction in the Congo River has not been possible but Lamboj (2004) and our own observations confirm the following courtship sequence in aquarium held *Teleogramma* (*T. brichardi* and *T.* cf. *brichardi*). After an initial approach with fins held tightly against the body the female contorts into an "S" or "U" shape directing her intensely colored belly toward the male while at the same time erecting and depressing her dorsal fin. Unlike most other chromidotilapines, for which observations are available, in *Teleogramma* the female's caudal fin is fanned and prominently displayed in front of the male (see Figure 8 of courting *T. brichardi*, and video of courtship behavior: https://youtu.be/8rA4i60YY0I). Additional studies, particularly mate choice assays as well as in-field observations where conditions allow, are clearly needed. Nonetheless these preliminary behavioral observations, in combination with the apparently species/lineage specific female fin patterning documented here (Figure 7) hint that tail patterning may play a role in reinforcing species boundaries in *Teleogramma*.

Conclusions & future studies

The data presented here add to a growing number of studies exploring the population genomics and diversification patterns in river cichlids. A more thorough understanding of river cichlids can provide important insights into the complex evolutionary history of African cichlids as a whole, especially in light of recent studies indicating standing genetic variation from rivers as an important source of genetic polymorphisms for lacustrine radiations (Loh *et al.* 2013; Brawand *et al.* 2014). Our results, which indicate that hydrological barriers play an important role in shaping geographic structuring and patterns of diversification in *Teleogramma* even at microscales (<1.5 km), are consistent with previous studies (Markert *et al.* 2010; Schwarzer *et al.* 2011) and are important for understanding the potential effects of proposed mega-dam development on these lineages. These results also highlight numerous areas for future studies. As this genus includes divergences across a spectrum of ages, it presents an excellent opportunity to assess patterns of genomic differentiation along the speciation continuum. For example, is divergence between parapatric ecotypes limited to a few large genomic regions (islands), or evenly spread across the genome? Such studies would be facilitated by a full genome assembly of a chromidotilapiine cichlid, as the closest genomes available now are from evolutionarily distant tilapiines (*Oreochromis*, ~46 Myr). Another potential area for study is identification of genomic regions that are linked to particular phenotypes, in particular putatively adaptive features such as dorso-ventral depression and swim bladder loss. Such studies would aid our growing knowledge of diversification mechanisms in riverine cichlids, providing important context for understanding the sources of genetic variation driving continent-wide diversification in this model group of vertebrates.

Acknowledgments. For assistance with permits and logistical support during collecting trips to the Democratic Republic of Congo and the Republic of Congo we extend our appreciation to Julien Punga Kumanenga, Jean Malekani Mukulire and Koffi Mulumba (University of Kinshasa, and Ministry of Agriculture, D.R. Congo), and Valentin Serge Pangou (GERDIB, University of Marien Ngoubi, and the Ministry of Scientific Research and Technological Innovation, Republic of Congo). For companionship and assistance in the field we are grateful to our Congolese colleagues Victor Mamonekene and Armel Ibala (University of Marien Ngouabi), Raoul Monsembula Iyaba, Tobit Liyandja, José Justin Mbimbi Mayi Munene (University of Kinshasa) and former AMNH colleagues, Robert Schelly, Hugo Escobar, Jake Lowenstein, and Jairo Arroyave. Chris Martinez (UC Davis) provided much helpful input with geometric morphometric analyses, and Oliver Lucanus (Below Water, Montreal) kindly provided video footage and photographs of courtship in captive held *Teleogramma*. We thank George Amato and the staff of the Sackler Institute for Comparative Genomics (AMNH) for logistical and financial support of the project. Bianca Brown (York College/CUNY), Rebecca Hersch (SICB/AMNH) and Guerry Perez (York College/CUNY) provided valuable assistance in the lab. We are grateful to Dan Edwards and Gisella Cacconne (YIBS) for guidance in constructing ddRAD libraries. Helpful comments from four anonymous Axios reviewers greatly improved the manuscript. Funds from the Sackler Institute for Comparative Genomics, Axelrod Research Curatorship, PSC-CUNY, and a generous gift from Ms. Janine Luke provided financial support for this study.

Author contributions: MLJS and SEA conceived of and designed the study. SEA carried out genomic data collection and conducted phylogenomic analyses. JMS performed all data quality/filtering analyses and population genomic analyses. MLJS conducted field studies including specimen collection and all morphological data collection and analyses. All authors wrote and edited the manuscript.

References cited

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655-1664.

Alter SE, Brown B, Stiassny ML (2015) Molecular phylogenetics reveals convergent evolution in lower Congo River spiny eels. *BMC Evolutionary Biology* **15**, 224.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**, 289–300.

Bouckaert R, Heled J, Kühnert D*, et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS computational biology* **10**, e1003537.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, ... & Turner-Maier J (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375-381.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular biology and evolution*, mss086.

Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution* **21**, 255-265.

Burridge CP, Craw D, Jack DC, King TM, Waters JM (2008) Does fish ecology predict dispersal across a river drainage divide? *Evolution* **62**, 1484-1499.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013a) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-40.

Catchen J, Bassham S, Wilson T*, et al.* (2013b) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction ‐ site associated DNA ‐ sequencing. *Molecular ecology* **22**, 2864-2883.

Chifman J, Kubatko L (2014) Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30**, 3317-3324.

Darwall W, Smith K, Allen D*, et al.* (2008) *The diversity of life in african freshwaters: underwater, under threat. An analysis of the status and distribution of freshwater species throughout mainland Africa* Universidad de El Salvador, San Salvador (El Salvador).

Drummond A, Bouckaert R (2015) *Bayesian evolutionary analysis with BEAST* Cambridge University Press, Cambridge, UK.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792-1797.

Egger B, Klaefiger Y, Indermaur A, Koblmüller S, Theis A, Egger S, Näf T, Van Steenberge M, Sturmbauer C, Katongo C, Salzburger W (2015) Phylogeographic and phenotypic assessment of a basal haplochromine cichlid fish from Lake Chila, Zambia. *Hydrobiologia* **748**, 171-84.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**, 342-350.

Ford AG, Dasmahapatra KK, Rüber L*, et al.* (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular ecology*.

Friedman M, Keck BP, Dornburg A*, et al.* (2013) Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting **280**, 20131733.

Genner MJ, Ngatunga BP, Mzighani S, Smith A, Turner GF (2015) Geographical ancestry of Lake Malawi's cichlid fish diversity. *Biology letters* **11**, 20150232.

Griffiths D (2015) Connectivity and vagility determine spatial richness gradients and diversification of freshwater fish in North America and Europe. *Biological Journal of the Linnean Society* **116**, 773-786.

Heled J, Drummond AJ (2011) Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, syr087.

Ho SY, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology*, syp035.

Huson DH, Bryant D (2005) Estimating phylogenetic trees and networks using SplitsTree 4. *Manuscript in preparation, software available from www. splitstree. org*.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PD (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* **7**, 544-548.

Jackson P, Oberg K, Gardiner N, Shelton J (2009) Velocity mapping in the Lower Congo River: A first look at the unique bathymetry and hydrodynamics of Bulu Reach, West Central Africa.

Katongo C, Koblmüller S, Duftner N, Makasa L, Sturmbauer C (2005) Phylogeography and speciation in the Pseudocrenilabrus philander species complex in Zambian Rivers. In: *Aquatic Biodiversity II*, pp. 221-233. Springer.

Keller I, Wagner C, Greuter L*, et al.* (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular ecology* **22**, 2848-2863.

Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular ecology resources* **11**, 353-357.

Koblmüller S, Duftner N, Sefc KM*, et al.* (2007) Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika–the result of repeated introgressive hybridization. *BMC Evolutionary Biology* **7**, 7.

Koblmüller S, Egger B, Sturmbauer C, Sefc KM (2010) Rapid radiation, ancient incomplete lineage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. *Molecular Phylogenetics and Evolution* **55**, 318-34.

Koblmüller S, Katongo C, Phiri H, Sturmbauer C (2012) Past connection of the upper reaches of a Lake Tanganyika tributary with the upper Congo drainage suggested by genetic data of riverine cichlid fishes. *African Zoology* **47**, 182-186.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics* **5**, 288-298.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR, Lockwood SF (1995) Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. *Molecular phylogenetics and evolution* **4**, 420-432.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359.

Li C, Ortí G, Zhang G, Lu G (2007) A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology* **7**, 1.

Loh YHE, Bezault E, Muenzel FM, Roberts RB, Swofford R, Barluenga M, ... & Hey J (2013) Origins of shared genetic variation in African cichlids. *Molecular Biology and Evolution*, **30,** 906-917.

López - Fernández H, Arbour JH, Winemiller K, Honeycutt RL (2013) Testing for ancient adaptive radiations in neotropical cichlid fishes. *Evolution* **67**, 1321-1337.

Lowenstein JH, Osmundson TW, Becker S, Hanner R, Stiassny ML (2011) Incorporating DNA barcodes into a multi-year inventory of the fishes of the hyperdiverse Lower Congo River, with a multi-gene performance assessment of the genus Labeo as a case study. *Mitochondrial DNA* **22**, 52-70.

Markert JA, Schelly RC, Stiassny ML (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.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.

Nevado B, Koblmüller S, Sturmbauer C*, et al.* (2009) Complete mitochondrial DNA replacement in a Lake Tanganyika cichlid fish. *Molecular ecology* **18**, 4240-4255.

Nickum JG (2004) *Guidelines for the use of fishes in research* American Fisheries Society Bethesda.

Oberg K, Shelton JM, Gardiner N, Jackson PR (2009) Discharge and Other Hydraulic Measurements for Characterizing the Hydraulics of Lower Congo River, July 2008 **33**.

Patterson N, Moorjani P, Luo Y*, et al.* (2012) Ancient admixture in human history. *Genetics* **192**, 1065-1093.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* **2**, e190.

Pauers MJ, McKinnon JS (2012) Sexual selection on color and behavior within and between cichlid populations: Implications for speciation. *Curr. Zool* **58**, 472-480.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* **8**, e1002967.

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular biology and evolution* **25**, 1253-1256.

Rambaut A, Drummond A (2007) Tracer v1. 4.

Ready J, Ferreira E, Kullander S (2006a) Discus fishes: mitochondrial DNA evidence for a phylogeographic barrier in the Amazonian genus Symphysodon (Teleostei: Cichlidae). *Journal of fish biology* **69**, 200-211.

Ready J, Sampaio I, Schneider H*, et al.* (2006b) Colour forms of Amazonian cichlid fish represent reproductively isolated species. *Journal of Evolutionary Biology* **19**, 1139-1148.

Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* **461**, 489-494.

Robert M (1946) *Le Congo Physique* H. Vaillant Carmanne SA.

Roberts TR, Stewart DJ (1976) An ecological and systematic survey of fishes in the rapids of the lower Zaire or Congo River. *Bulletin of the Museum of Comparative Zoology* **147**, 239-317.

Rohlf F (2013) tpsDIG2.

Santos ME, Salzburger W (2012) How cichlids diversify. *Science* **338**, 619-621.

Schwarzer J, Lamboj A, Langen K, Misof B, Schliewen UK (2015) Phylogeny and age of chromidotilapiine cichlids (Teleostei: Cichlidae). *Hydrobiologia* **748**, 185-199.

Schwarzer J, Misof B, Ifuta SN, Schliewen UK (2011) Time and origin of cichlid colonization of the lower Congo rapids. *PLoS One* **6**, e22380.

Schwarzer J, Swartz ER, Vreven E*, et al.* (2012) Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution. *Proceedings of the Royal Society B: Biological Sciences* **279**, 4389-4398.

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society of London B: Biological Sciences* **273**, 1987-1998.

Showers KB (2011) Beyond mega on a mega continent: Grand Inga on Central Africa's Congo River. In: *Engineering Earth*, pp. 1651-1679. Springer.

Stiassny M, Brummett R, Harrison I, Monsembula R, Mamonekene V (2011) The status and distribution of freshwater fishes in central Africa. *Brooks, EGE, Allen, DJ & Darwell, WT (Compilers), The Status and Distribution of freshwater biodiversity in Central Africa. IUCN: Gland, Switzerland and Cambridge, UK*, 27-46.

Stiassny ML, Alter SE (2015) Phylogenetics of Teleogramma, a Riverine Clade of African Cichlid Fishes, with a Description of the Deepwater Molluskivore-Teleogramma obamaorum-from the Lower Reaches of the Middle Congo River. *American Museum Novitates*, 1-18.

Takahashi K, Terai Y, Nishida M, Okada N (2001) Phylogenetic Relationships and Ancient Incomplete Lineage Sorting Among Cichlid Fishes in Lake Tanganyika as Revealed by Analysis of the Insertion of Retroposons. *Molecular biology and evolution* **18**, 2057-2066.

Tedesco PA, Leprieur F, Hugueny B*, et al.* (2012) Patterns and processes of global riverine fish endemism. *Global Ecology and Biogeography* **21**, 977-987.

Thieme ML, Abell R, Burgess N*, et al.* (2005) *Freshwater ecoregions of Africa and Madagascar: a conservation assessment* Island Press.

Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**, 3907-3930.

Wagner CE, Keller I, Wittwer S*, et al.* (2013) Genome - wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* **22**, 787-798.

Wickler W (1959) Teleogramma brichardi Poll 1959. *Die Aquarien-und Terrarien-Zeitschrift (DATZ)* **12**, 228-230.

Willis SC, Farias IP, Ortí G (2014) Testing mitochondrial capture and deep coalescence in Amazonian cichlid fishes (Cichlidae: Cichla). *Evolution* **68**, 256-268.

Winemiller K, McIntyre P, Castello L*, et al.* (2016) Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. *Science* **351**, 128-129.

Data accessibility

DNA sequences: Genbank accession numbers KP714141-KP714177; NCBI SRA BioProject PRJNA351851
SNP data (VCF file); Morphometric data (PC scores, covariance matrix, procrustes coordinates): DRYAD entry doi:10.5061/dryad.c7c8f

Tables and Figures

Table 1. Summary genetic diversity statistics for reference-aligned data. These values were calculated by STACKS for nucleotide positions that were polymorphic in at least one population. *N* = average number of individuals genotyped at each locus; *Sites* = number of polymorphic nucleotide sites across the dataset; *%Poly* = percentage of polymorphic loci; H_{obs} = average observed heterozygosity per locus; π = average nucleotide diversity.

Taxon	**N**	**Sites**	**% Poly**	$\mathbf{H_{obs}}$	π	F_{IS}
T. brichardi	1.4	20,769	0.109	0.124	0.108	-0.024
T. cf. *brichardi*	12.5	24,207	0.275	0.155	0.131	-0.044
T. cf. *depressa*	2.5	22,567	0.144	0.114	0.102	-0.019
T. depressa	3.4	23,468	0.147	0.108	0.095	-0.019
T. gracile	7.5	22,054	0.198	0.132	0.110	-0.039
T. monogramma	4.7	20,673	0.109	0.098	0.067	-0.057
T. obamaorum	8.9	24,115	0.202	0.135	0.112	-0.041

Figure 1. (a) The lower Congo River (highlighted) and proximate Congo tributaries. Ranges of *T. monogramma* (brown stars), and *T. obamaorum* (inverted green triangles) inset. (b) distribution of species along the LCR color coded, and schematically represented by offset colored blocks. In light grey are regions where no sampling has been possible. No *Teleogramma* have been found in sampled regions below the shown range of *T. depressa*, or in Pool Malebo. In green, an elevational profile of the LCR from Pool Malebo (280m above sea level) to the Atlantic, with boundaries of three main hydrobiological regions delineated. Locations mentioned in the text are: 1. Kinsuka, 2. Les Rapides, 3. Mbudi, 4. Ngombe, 5. Foulakari, 6. Banda Nyenge, 7. Lenga Lenga, 8. Upstream Luozi, 9. Luozi, 10. Bulu, 11. Kinganga, 12. Inga.

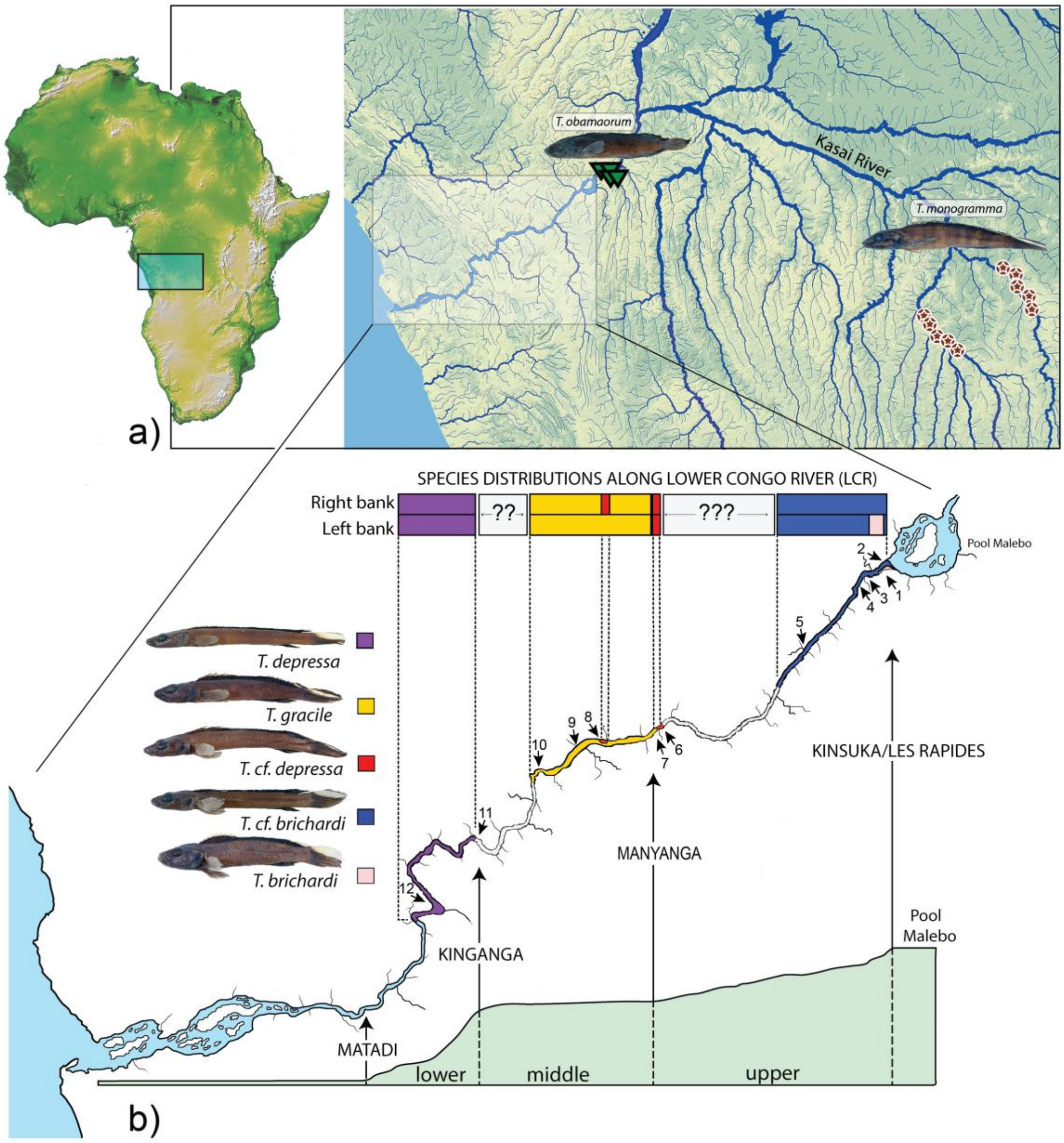


Figure 2. Phylogenetic reconstruction of relationships between *Teleogramma* lineages. Multispecies coalescent analysis using 37, 826 genome-wide SNPs. Maximum clade credibility tree inferred in SNAPP (Bryant *et al.* 2012) as implemented in BEAST v2.2 (Bouckaert *et al.* 2014) is shown in black right-angled tree with posterior probabilities indicated at nodes. Underlying colored tree cloud of the last 4500 trees (sampled every 1000 generations of 5,000,000 MCMC steps) visualized using DENSITREE (Bouckaert 2010) shows the range of topologies recovered.

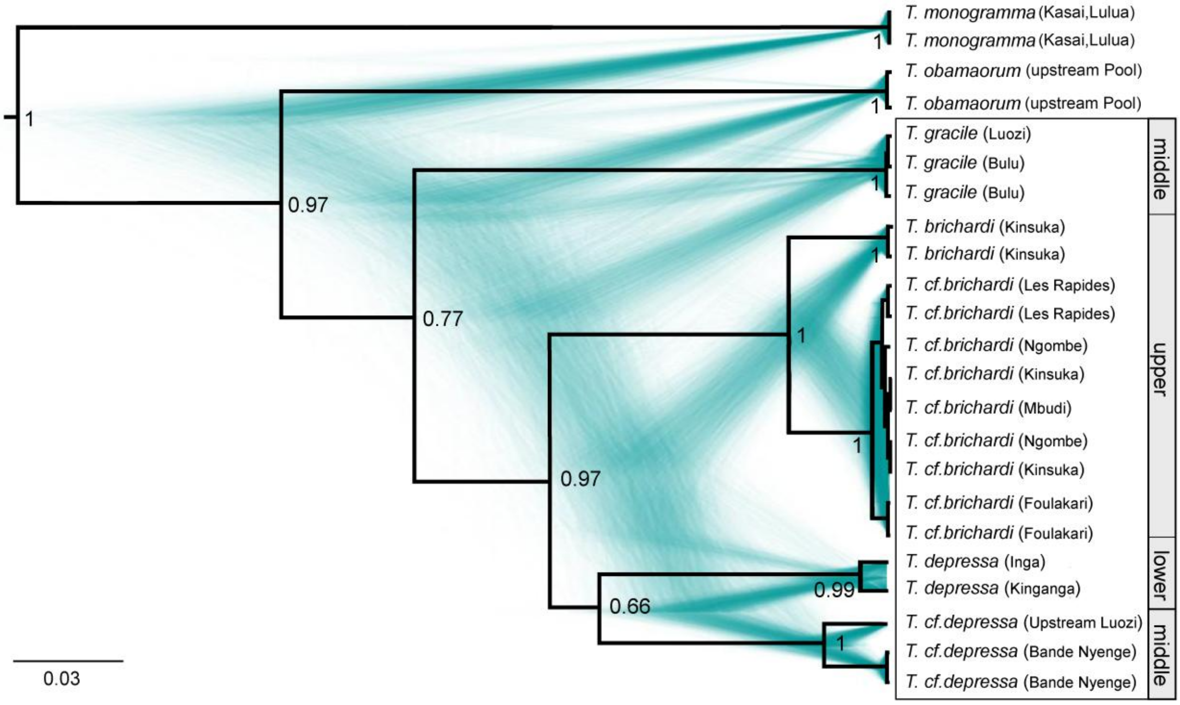

T. monogramma (Kasai,Lulua)
T. monogramma (Kasai,Lulua)
T. obamaorum (upstream Pool)
T. obamaorum (upstream Pool)
T. gracile (Luozi)
T. gracile (Bulu)
T. gracile (Bulu)
T. brichardi (Kinsuka)
T. brichardi (Kinsuka)
T. cf.brichardi (Les Rapides)
T. cf.brichardi (Les Rapides)
T. cf.brichardi (Ngombe)
T. cf.brichardi (Kinsuka)
T. cf.brichardi (Mbudi)
T. cf.brichardi (Ngombe)
T. cf.brichardi (Kinsuka)
T. cf.brichardi (Foulakari)
T. cf.brichardi (Foulakari)
T. depressa (Inga)
T. depressa (Kinganga)
T. cf.depressa (Upstream Luozi)
T. cf.depressa (Bande Nyenge)
T. cf.depressa (Bande Nyenge)
middle
upper
lower
middle
1
1
1
0.97
1
0.77
1
1
1
0.97
0.99
0.66
1
0.03

Figure 3. Time-calibrated phylogeny of *Teleogramma* taxa using concatenated mitochondrial genes (COI and ND2) inferred using Bayesian inference in BEAST v2.2 (Bouckaert *et al.* 2014) with a relaxed clock and using *Chromidotilapia kingsleyae* as an outgroup. Elevational profile of the LCR showing approximate locations of sampled sites inset at upper left.

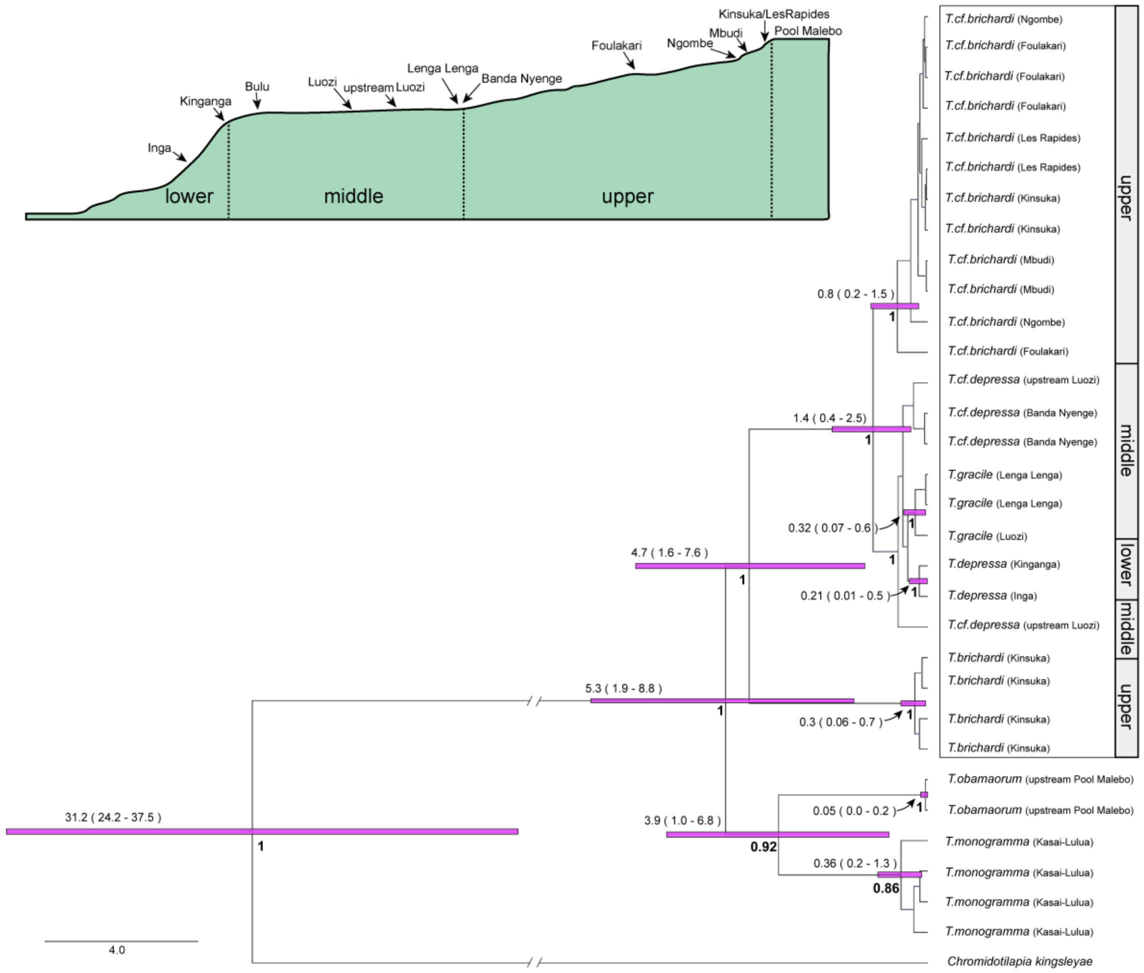


Figure 4. a) Ancestry proportions of *Teleogramma* specimens as determined by Maximum Likelihood estimation from reference-aligned SNP data for K = 5 in the program ADMIXTURE (Alexander *et al.* 2009), b) ADMIXTURE ancestry proportions from *de novo* SNPs for K = 6; c) scatterplot of PC1 vs. PC2 from a genetic principal component analysis (PCA) using reference-aligned SNPs; d) scatterplot of PC1 vs PC4 from a genetic PCA.

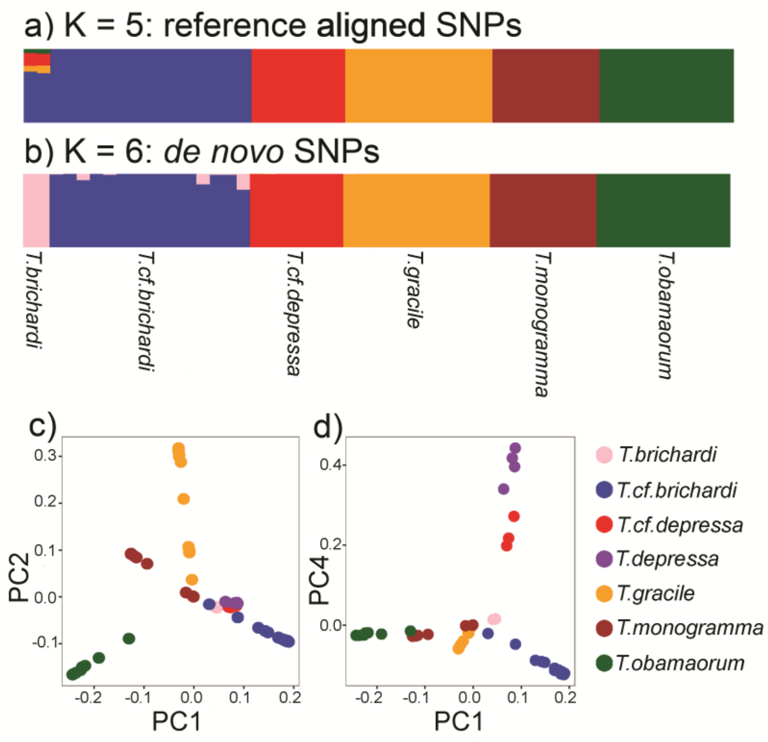


Figure 5. Graph of admixture between 22 individuals fitted using TreeMix (Pickrell & Pritchard 2012) with two migration edges. The drift parameter is proportional to 2Ne generations and migration weight indicates the proportion of ancestry deriving from the migration edge.

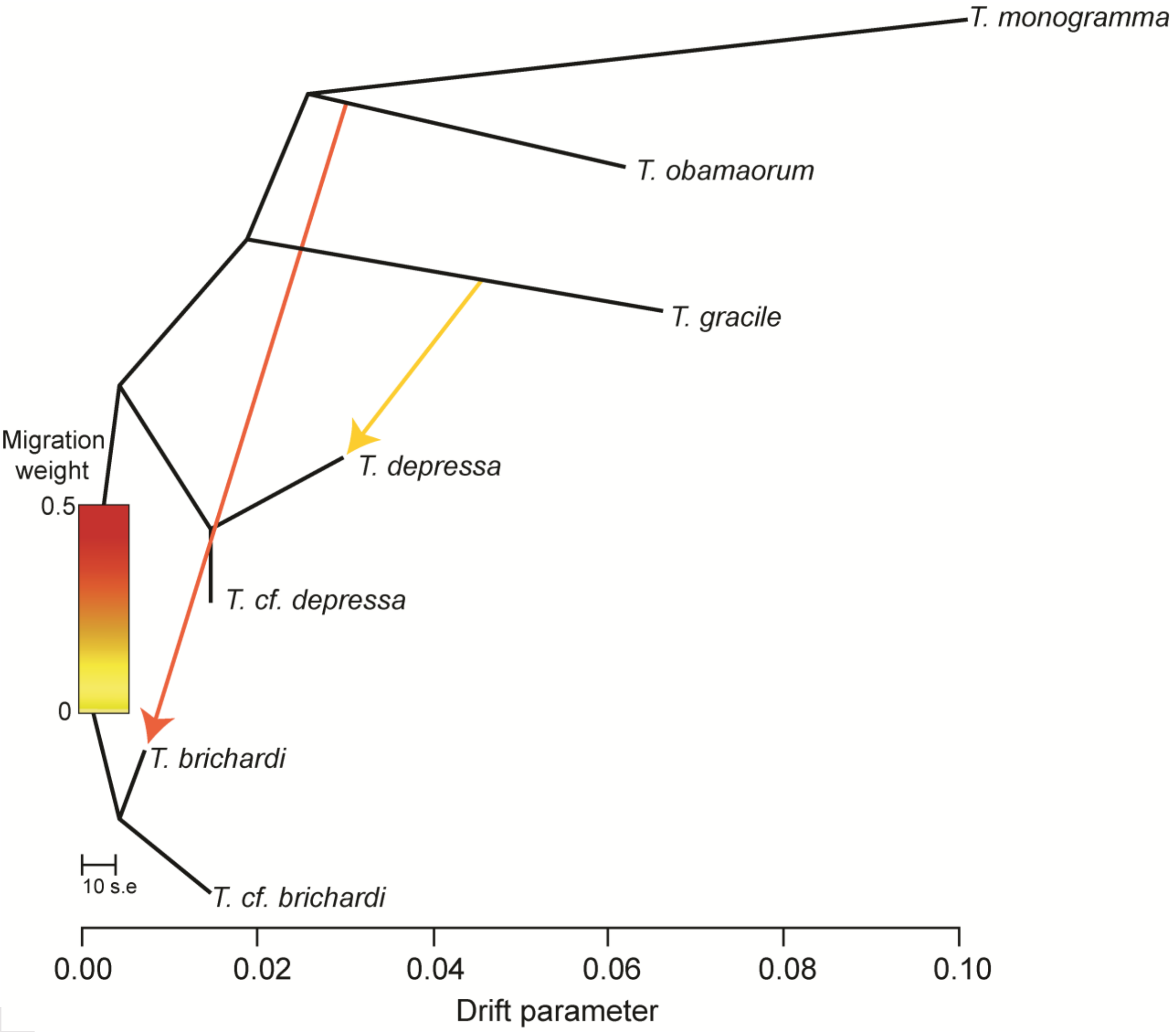


Figure 6. a) Principal Component Analysis of shape variance across 74 adult *Teleogramma* specimens sampled from across the generic range: b) location of homologous landmarks used in study, with deformations implied by PC1 scores indicated by black bars subtending landmarks c) Scree plot of percentage variance explained by each PC as a % of total variance; PC1 accounts for 49% of total shape variance, and PC2 16%.

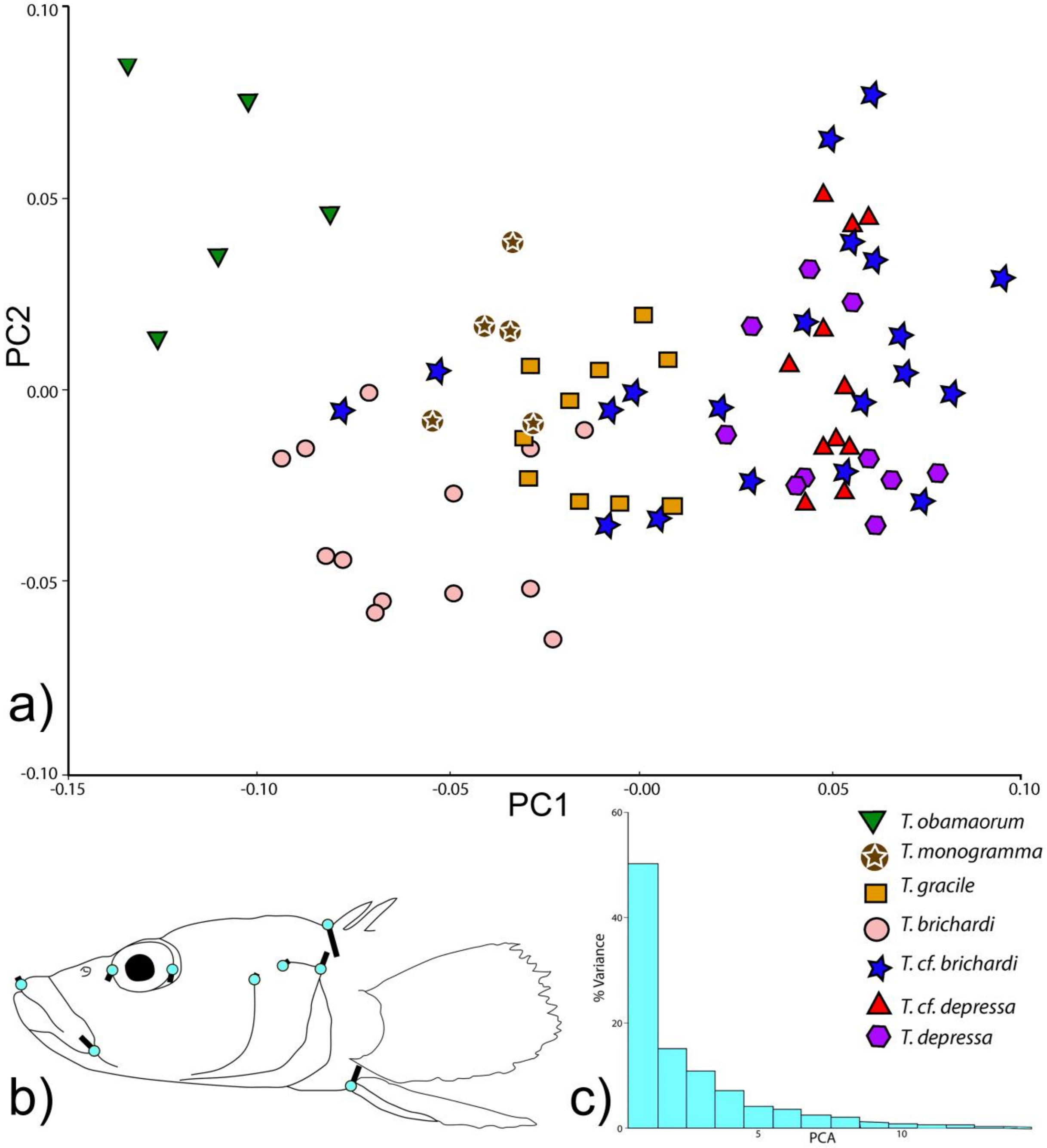


Figure 7. Variation in fin patterning across genetically-defined lineages. At top: aquarium-held female and male *Teleogramma brichardi*. In panel below (only females shown): a) *T. brichardi*, b) *T*. cf. *brichardi*, c) *T*. cf. *depressa*, d) *T. depressa*, e) *T. gracile*, f) *T. monogramma*, g) *T. obamaorum*.

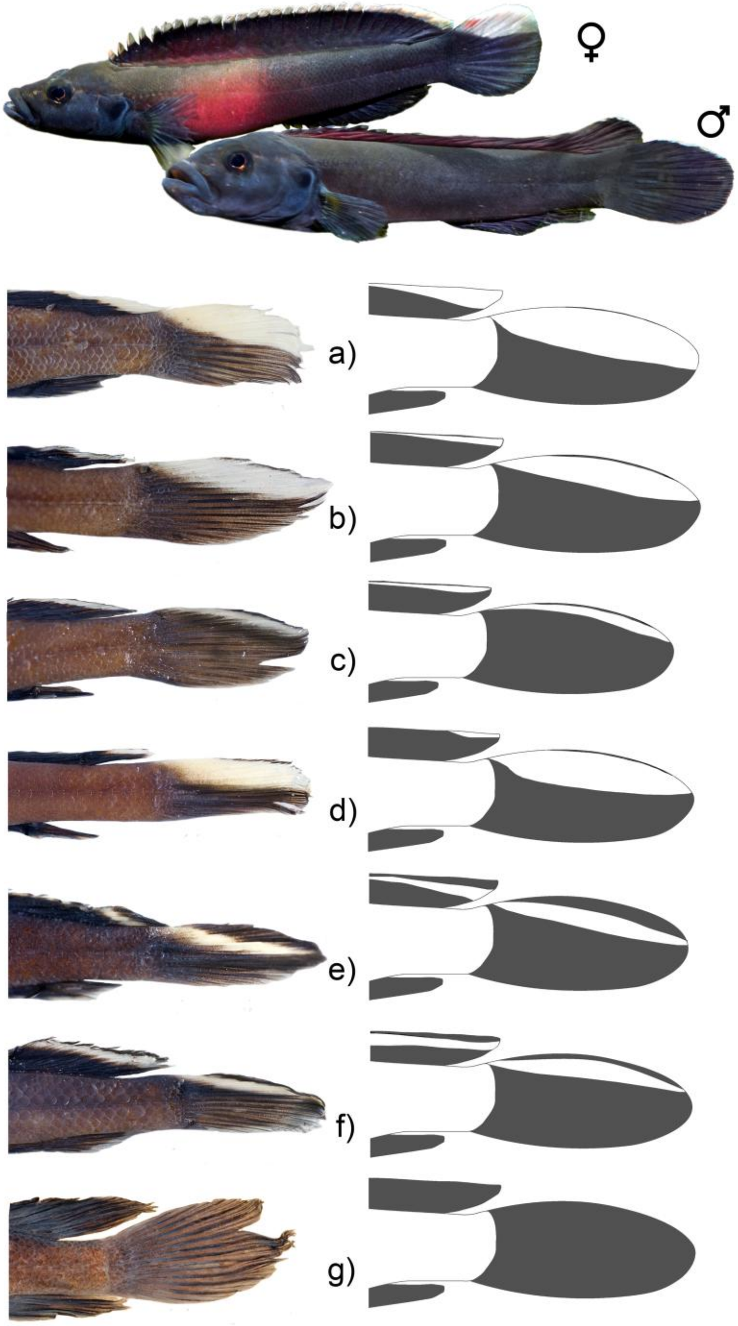
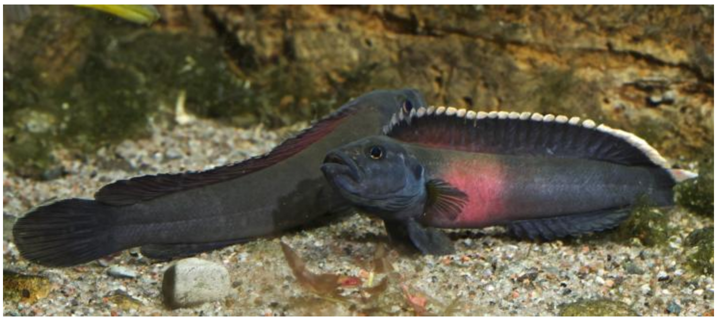


Figure 8. A pair of aquarium held *Telegramma brichardi* photographed during courtship, showing the characteristic display of the female presenting caudal fin to the male (courtesy of Oliver Lucanus).



Supporting Information:

Supporting Figures (see below)

Figure S1: Best supported topology of *Teleogramma* using the Single Root Decomposition SVDquartets method of phylogenetic analysis (Chifman and Kubatko 2014). Branch labels correspond to proportion of bootstrap values (500 bootstraps) supporting each node.

Figure S2. CV Error plots for ADMIXTURE analysis.

Figure S3: Percent variance in principal components using reference-aligned data

Figure S4: ADMIXTURE plot for *T. brichardi* and populations of *T.* cf. *brichardi*.

Figure S5. Unrooted phylogenetic analysis computed in SplitsTree4 using the NeighborNet algorithm (Huson and Bryant 2006).

Supporting Tables (see below)

Table S1: Specimens and accession numbers

Table S2: Number of reads, number of variable loci genotyped in each individual, percent of reads aligned to reference genome and mean coverage.

Table S3: Summary genetic diversity statistics for data analysed using the *de novo* pipeline in STACKS.

Table S4: Tracy-Widom statistics for PCA using reference-aligned data

Table S5: F_{ST} values calculated between taxa using the reference-aligned pipelines in STACKS. All values in bold were found to be significant after a false discovery rate correction for multiple tests (q=0.0476) (Benjamini and Hochberg 1995).

Table S6: F_{ST} values calculated between *T. brichardi* and populations of *T.* cf. *brichardi* using the reference-aligned pipelines in STACKS. All values in bold were found to be significant after a false discovery rate correction for multiple tests (q=0.0409)(Benjamini and Hochberg 1995).