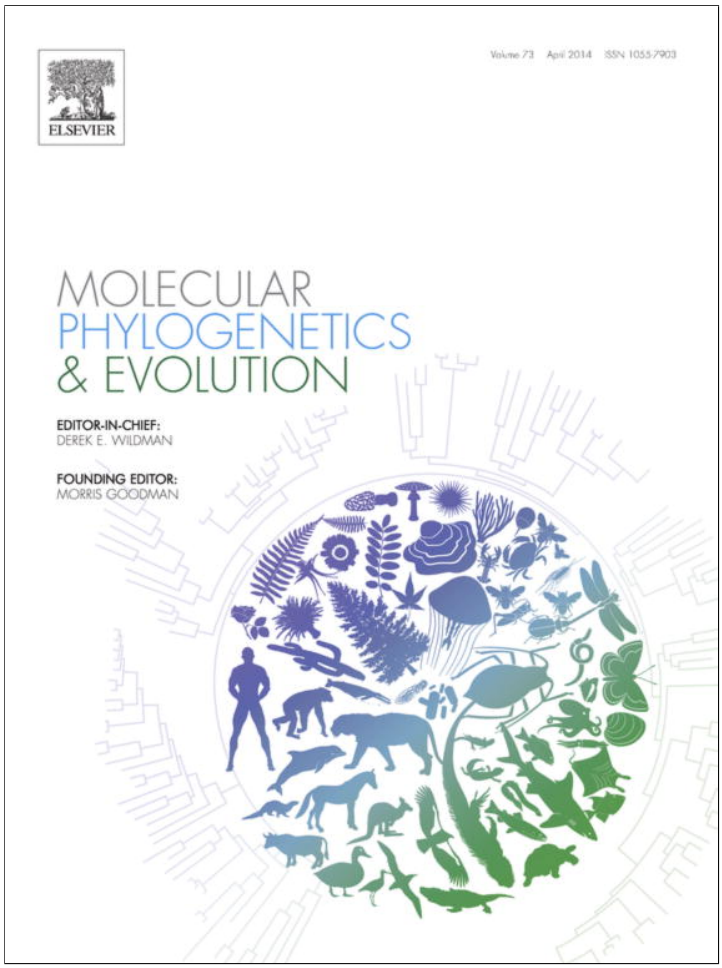

Molecular Phylogenetics and Evolution 73 (2014) 53–59

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Continental monophyly of cichlid fishes and the phylogenetic position of *Heterochromis multidens*

Benjamin P. Keck *, C. Darrin Hulsey

Department of Ecology and Evolutionary Biology, University of Tennessee, 569 Dabney Hall, Knoxville, TN 37996, USA

ARTICLE INFO

Article history:
Received 6 September 2013
Revised 15 January 2014
Accepted 18 January 2014
Available online 26 January 2014

Keywords:
Continental Drift
East African Rift Lakes
Gondwana
Tilapia

ABSTRACT

The incredibly species-rich cichlid fish faunas of both the Neotropics and Africa are generally thought to be reciprocally monophyletic. However, the phylogenetic affinity of the African cichlid *Heterochromis multidens* is ambiguous, and this distinct lineage could make African cichlids paraphyletic. In past studies, *Heterochromis* has been variously suggested to be one of the earliest diverging lineages within either the Neotropical or the African cichlid radiations, and it has even been hypothesized to be the sister lineage to a clade containing all Neotropical and African cichlids. We examined the phylogenetic relationships among a representative sample of cichlids with a dataset of 29 nuclear loci to assess the support for the different hypotheses of the phylogenetic position of *Heterochromis*. Although individual gene trees in some instances supported alternative relationships, a majority of gene trees, integration of genes into species trees, and hypothesis testing of putative topologies all supported *Heterochromis* as belonging to the clade of African cichlids.

1. Introduction

The phylogenetic placement of the cichlid *Heterochromis multidens* (Teleostei: Ovalentaria) endemic to the Congo River Basin in central Africa has been debated extensively. Despite a steady accumulation of morphological and molecular data for this evolutionarily distinctive cichlid species, debate remains as to whether *Heterochromis* is more closely related to either the over 1000 species of otherwise monophyletic "African" cichlids (Africa + eastern Mediterranean + Iran) or the monophyletic clade of more than 500 Neotropical cichlid species (Chakrabarty, 2004; Klett and Meyer, 2002). Several morphological and molecular studies have concluded that this lineage has greater evolutionary affinities with the Neotropical cichlid clade (Kullander, 1998; Lopez-Fernandez et al., 2010), thus making African cichlids paraphyletic. However, a number of other studies identify *Heterochromis* as belonging to a monophyletic clade with all other African cichlids (Farias et al., 2000, 1999, 2001; Friedman et al., 2013; Genner et al., 2007; McMahan et al., 2013; Smith et al., 2008). Several of these studies have also applied various weighting schemes to characters that resulted in variable phylogenetic placements of *Heterochromis* (Farias et al., 2001; Kullander, 1998; Stiassny, 1991). It has even been suggested that *Heterochromis* represents the sister lineage to a clade containing all African and Neotropical cichlid species that originated following the still earlier divergence of the basal cichlid groups found only in Madagascar and southern India (Lippitsch, 1995; Oliver, 1979, 1984). Despite its profound implications for cichlid biogeography, the phylogenetic placement of *Heterochromis* remains ambiguous.

The phylogenetic placement of *Heterochromis* could also have substantial influence on our understanding of the timeframe of cichlid diversification. Calibrations of molecular clocks for cichlids have often dated the node subtending the African and Neotropical clades with the geologic age of continental rifting between the African and South American landmasses (Azuma et al., 2008; Genner et al., 2007; Lopez-Fernandez et al., 2013). Additionally, other studies of temporal divergence have used the node subtending *Heterochromis* and all other African cichlids as a molecular clock calibration point (Genner et al., 2007; McMahan et al., 2013; Schwarzer et al., 2009). If African and Neotropical cichlids are reciprocally monophyletic, then the molecular divergence between any of the over 500 species of Neotropical cichlids and any of the over 1000 species of African cichlids could correctly estimate the divergence time of cichlids endemic to these two continents. Alternatively, if *Heterochromis* is more closely related to Neotropical cichlids, then only the node subtending these two groups should be utilized to provide estimates of temporal divergence between the African and the Neotropical cichlid radiations. Accurately dating the cichlid phylogeny has been an essential component of studies examining rates of cichlid speciation and

* Corresponding author.
E-mail addresses: bkeck@utk.edu (B.P. Keck), chulsey@utk.edu (C.D. Hulsey).

http://dx.doi.org/10.1016/j.ympev.2014.01.011
1055-7903/

phenotypic divergence (Arbour and Lopez-Fernandez, 2013; Chakrabarty, 2006; Genner et al., 2007; Gonzalez-Voyer and Kolm, 2011; Hoerner, 2011; Hulsey et al., 2010; Lopez-Fernandez et al., 2013, 2005, 2010; Martin and Bermingham, 1998; McMahan et al., 2013; Schwarzer et al., 2011, 2009; Sturmbauer et al., 2010). Therefore, any inferences dependent on a geographically calibrated temporal framework of cichlid evolution would be compromised if they were based on an inaccurate understanding of the phylogenetic placement of *Heterochromis*.

Reconstructing relationships among relatively anciently diverging groups like *Heterochromis* and other major clades of cichlids could be problematic if only isolated, individual loci with weak phylogenetic signal were examined. Furthermore, if these groups diverged relatively rapidly even in the distant past, many genes could show inconsistent topological relationships due to hybridization, variable rates of molecular evolution, or other stochastic sorting of ancestral polymorphism (Degnan and Rosenberg, 2006). There are now many methods that can reconstruct a robust species tree while accounting for the expected discordance among individual gene trees, and their use has become common practice (e.g., Near and Keck, 2013; Niemiller et al., 2013; Pyron et al., 2013; Reynolds et al., 2013; Sanders et al., 2013). These species tree methods estimate a containing tree from multiple gene trees while taking into account the possibility that not all gene trees within a species tree will be congruent (Degnan and Rosenberg, 2006; Maddison, 1997). There are several methods available to conduct species tree estimation that use either maximum likelihood, e.g., STEM (Kubatko et al., 2009) and MDC (Maddison and Knowles, 2006) or Bayesian inference, e.g., BEST (Edwards et al., 2007), ∗BEAST (Drummond and Rambaut, 2007; Heled and Drummond, 2010), and BUCKy (Larget et al., 2010) with the goal of avoiding misleading results associated with concatenation (Edwards et al., 2007; Kubatko and Degnan, 2007). However, in circumstances where gene tree estimation error is likely, for instance, in putatively rapidly radiating clades such as cichlids, species tree methods that make estimations under the multispecies coalescent or assume knowledge of the true gene trees can be less accurate than simple concatenation (McVay and Carstens, 2013b; Patel et al., 2013). Therefore, estimation of species trees using both a concatenated dataset as well as a framework such as that implemented in BUCKy (Larget et al., 2010) should provide a robust range of inferences concerning the phylogenetic position of *Heterochromis*.

To test among several alternative phylogenetic hypotheses (Fig. 1), we generated sequence data from 29 nuclear loci to resolve the phylogenetic position of *Heterochromis*. Using representative cichlid lineages from the Neotropics, Africa, and India, we examined the phylogenetic affinities of *Heterochromis* inferred from both individual gene trees and species trees. We also implemented several topology-based statistical tests to examine the support for alternative hypotheses. We used the level of congruence of results across these different methods to determine if *Heterochromis* should be considered (A) outside both the remaining African + Neotropical cichlids, (B) part of a three-way polytomy with the remaining African and Neotropical cichlids, (C) allied with the Neotropical cichlids, or (D) allied with the African cichlids.

2. Materials and methods

We examined the phylogenetic affinities of *Heterochromis* and a set of species that have been shown in previous analyses to represent evolutionarily disparate members of the major clades of cichlids (Farias et al., 2000; Friedman et al., 2013; Smith et al., 2008). Our exemplar Neotropical cichlids included *Herichthys cyanoguttatus*, *Cichla temensis*, and *Retroculus xinguensis*. These species were chosen to have a most recent common ancestor (MRCA)

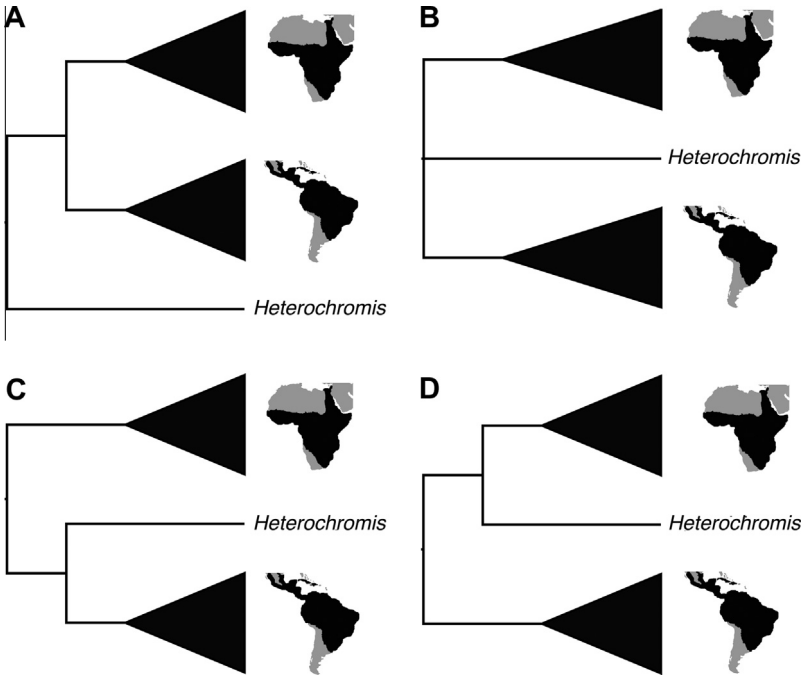

Fig. 1. Four hypothetical phylogenies depicting the possible relationship of *Heterochromis multidens* with clades representing all other "African" and all Neotropical cichlids. Importantly, we consider African cichlids herein to include those species in the Levant, or eastern Mediterranean, and Iran and to exclude species in Madagascar based on previous phylogenetic analyses (Klett and Meyer, 2002; Smith et al., 2008). In the hypothesized phylogenies, the branch leading to *Heterochromis* is labeled. The branches leading to the geographically defined clades are identified with black shading indicating the range of cichlids in those areas. The four topological relationships of *Heterochromis* suggested from previous studies are depicted. These topologies include: (A) African and Neotropical clades more closely related to each other than either is to *Heterochromis* (Oliver, 1979), (B) Polytomy among *Heterochromis* and the two continental clades (Oliver, 1984; Stiassny, 1991), (C) *Heterochromis* allied with the Neotropical clade (Kullander, 1998; Lopez-Fernandez et al., 2010), (D) *Heterochromis* allied with the African clade (Farias et al., 2000, 1999, 2001; Friedman et al., 2013; Genner et al., 2007; Smith et al., 2008).

representing the MRCA of all Neotropical cichlids. We included *Tylochromis mylodon*, *Hemichromis elongatus*, *Cyrtocara moorii*, and *Oreochromis niloticus* as representative African cichlids. These species were chosen to represent the MRCA of all African cichlids just prior to the potential divergence of this clade from *Heterochromis*. We sequenced *Etroplus maculatus* that is native to the Indian subcontinent as the outgroup for most genes. However, due to failed amplification for the genes *ctsh* and *sox11a*, we used sequences available on GenBank for *Lutjanus argentimaculatus* (FJ772425) and *Epinephalus coioides* (HM008643), respectively, in phylogenetic analyses of these two genes. Tissues were taken from both the wild and commercially obtained specimens, and H. Lopez-Fernandez provided aliquots of isolated genomic DNA from *Retroculus* and *Heterochromis*. We also analyzed sequences of *Oreochromis niloticus* from its Ensembl Genome database (Flicek et al., 2012).

To amplify the 29 loci utilized for this study, we used a combination of previously published primers for 17 genes (Terai et al., 2002; Lopez et al., 2004; Li et al., 2007, 2010a,b; Hulsey et al., 2011) and newly designed primer sets for 12 genes (Table 1). We

Table 1
List of primers. **New** indicates novel primers generated for this study.

Gene	Source	Primer name	Primer sequence
ak1	New	AK1For1	GATCATCTTTGTTGTGGGTGGGCCTGGCT
		AK1Rev1	ACTGGCTCRGTTGCTTTGTAATACA
acta1	New	ACTA1For1	CATGGTCGGTATGGGTCAGAAGGACT
		ACTA1Rev1	TCTTCATCAGGTAGTCGGTCAGATCG
bmp4	Terai et al. (2002)	BMP4FS1	CCCTTTCCTCTGGCAGATCATCT
		BMP4RS1	CATGTTCATCTAGGTAGAGCATGGAGA
cilp1	New	CILP1For1	CGCTACGATTACAACACAGTGCCATT
		CILP1Rev1	CCAGCATGGTGAACTCATTGGTGTCAT
ctsh	New	CTSHFor2	ACCAARGGGAACTACTTCAGCAGCAA
		CTSHRev2	TCKGTCATCAGYCCCTTGTTGTACA
dcn	New	DCNFor1	GGTCTACCTTCATGCCAACAAGATTGCT
		DCNRev1	TATCTASTACATTCTAGTATACCGGTTG
dlx3	New	DLX3For1	ATCCACCAGTGGACTTTCTTTATTCTG
		DLX3Rev1	GAGCCCGGAGGTGAAGCTTGCAGGTGA
dystB	Hulsey et al. (2011)	Dystbfor	GCGCATTGCAGACTTTGATCT
		Dystbrev	TGCTGCTGTTKCCAGATGCCTAAT
enc1	Li et al. (2007)	ENC1_F85	GACATGCTGGAGTTTCAGGA
		ENC1_R982	ACTTGTTRGCMACTGGGTCAAA
fgfr1	New	FGFR1For1	ATGCCAAATCATTAGTGGAAAGGCTT
		FGFR1Rev1	TGAACACCTGCTCCCTGTATTTGGTC
fz8a	New	FZ8AFlor1	TGGAGGTGCACCAGTTCTGGCCTCTGGT
		FZ8ARev1	GTGTTCACCGGCACCATGGGCTCCAA
		FZ8ARev2	TCGGTTCTGTTGTAGTCCATGCACAG
glyt	Li et al. (2007)	Glyt_F559	GGACTGTCMAAGATGACCACMT
		Glyt_R1562	CCCAAGAGGTTCTTGTTRAAGAT
nup155	Li et al. (2010b)	li1777E4for	AGGAGYTGGTGAACCAGAGCAAAGC
		li1777E4rev	AGATCRGCCTGAATSAGCCAGTT
myh	Li et al. (2007)	myh6_F459	CATMTTYTCCATCTCAGATAATGC
		myh6R1325	ATTCTCACCACCATCCAGTTGAA
opn4	New	OPN4For1	ACCTGTTTCCTACYGTGGATGTCCC
		OPN4For2	TCTGGTYATCTACGTGTTCTGTCGCAGT
		OPN4Rev1	GGTTGTGGATGGCAGARGCCTTGGCGAT
		OPN4Rev2	TTRYTGAYCTCRCCACAGTTCAGCTGCT
plagl2	Li et al. (2007)	plagl2_F9	CCACACACTCYCCACAGAA
		plagl2_R930	TTCTCAAGCAGGTATGAGGTAGA
ptr	Li et al. (2007)	Ptr_F458	AGAATGGATWACCAACACYTACG
		Ptr_R1248	TAAGGCACAGGATTGAGATGCT
rag1	Lopez et al. (2004)	RAG1F1	CTGAGCTGCAGTCAGTACCATAAGATGT
		RAG1R1	CTGAGTCCTTGTGAGCTTCCATRAAYTT
rho	New	RHOcicFor1	AGCTTATGCTGCTCTGGGTGCC
		RHOcicRev1	TCCAAATTCAGAGCCTTGATG
sidkey	Li et al. (2010a)	sidkey_F116	CGGATGARGYCTGCAGCAG
		sidkey_R1360	GCTGGGCYTTKGTCAGACTGT
sh3px3	Li et al. (2007)	SH3PX3_F461	GTATGGTSGGCAGGAACYTGAA
		SH3PX3_R1303	CAAACAKCTCYCCGATGTTCTC
sox11a	New	SOX11AFor1	ATGGAGCAGTCGCCGGACATGCACAACG
		SOX11ARev1	AATCCCCATCCGGTAGTCCTCYTCGTA
sreb2	Li et al. (2007)	sreb2_F10	ATGGCGAACTAYAGCCATGC
		sreb2_R1094	CTGGATTTTCTGCAGTASAGGAG
tbr1	Li et al. (2007)	tbr1_F1	TGTCTACACAGGCTGCGACAT
		tbr1_R820	GATGTCCTTRGWGCAGTTTTT
ube3a	Li et al. (2010a)	UBE3A_F354	GTCTACGACAGCCTATTGTCCAATGAGAARAT
		UBE3A_R1140	CGCTRTACATGCTGATCCKGTTGT
ube3a-like	Li et al. (2010a)	UBElike_F368	GCTGGCCATCGASCAGGTGGAG
		UBE3A_R1140	CGCTRTACATGCTGATCCKGTTGT
wnt7b	New	WNT7BFor1	GCAGCCACRGGGAGCGTGGGCACCCAGG
		WNT7BRev1	AGCAYAATCTCWATGTGACAGTAAAGTT
zic1	Li et al. (2007)	zic1_F9	GGACGCAGGACCGCARTAYC
		zic1_R967	CTGTGTGTGTCCTTTTGTGRATYTT
znf	Li et al. (2010a)	znf503_F12	GAAAAGTCCGYTGGCTCTKCT
		znf503_R1422	CGCCGAYGCTGTGGTSAGTCT

designed our new primer sets using pairwise alignments of various teleost protein coding regions (Supplementary Table 1). All primers were optimized to amplify at 55C using standard protocols for polymerase chain reaction of nuclear loci in teleost fishes (Li et al., 2007, 2010a). We used the medaka genome that has been fully sequenced and is extensively annotated (http://www.ensembl.org/index.html) in order to estimate the chromosomal location and putative genomic independence of the markers (Kasahara et al., 2007).

To infer the gene tree for each locus, we used only protein-coding regions except for the genes *fgfr1* and *dystB* that contained numerous small introns between the amplified exonic regions. After aligning each locus with MUSCLE 3.8 (Edgar, 2004), we used AIC in jModelTest 2.1.4 (Posada, 2008) to identify the best model of molecular evolution. Bayesian analyses for each gene were run in MrBayes 3.2 (Ronquist et al., 2012) for 10,000,000 generations in two simultaneous runs with four Markov chains in each run. Trees and parameters were sampled from the MCMC chain every 1000 generations. We examined trends in the likelihood vs. generation plots, values for the average standard deviation of the split frequencies, and effective sample sizes (ESS) in Tracer 1.5 (Rambaut and Drummond, 2003) to assess phylogenetic convergence. For all runs, we used the sumt command in MrBayes (Ronquist et al., 2012) to remove the first 20% of trees as burn-in and estimated the maximum clade credibility tree with the post-burn-in set of trees. We also used the same analytical scheme on a concatenated dataset of all 29 loci partitioned by gene with separate models of molecular evolution for each partition.

For each gene, we compared alternative phylogenetic hypotheses employing approximately unbiased (AU) tests, Kishino/Hasegawa (KH) tests, Shimodaira/Hasegawa (SH) tests, and the resampling of estimated log-likelihoods (RELL) to estimate bootstrap proportions (Hasegawa and Kishino, 1989; Kishino and Hasegawa, 1989; Kishino et al., 1990; Shimodaira, 2002; Shimodaira and Hasegawa, 1999), as implemented in the program CONSEL (Shimodaira and Hasegawa, 2001) and the baseml function of the program PAML 4.4 (Yang, 1997, 2007). Both the SH and KH tests compare the variance of the difference in log likelihood values between different topologies (Hasegawa and Kishino, 1989; Kishino and Hasegawa, 1989; Shimodaira and Hasegawa, 1999). Generally, the KH test is thought to be biased toward type I error and the SH test is thought to be much more conservative (Goldman et al., 2000; Shimodaira, 2002; Shimodaira and Hasegawa, 1999). For AU, Shimodaira (2002) developed a multi-scale bootstrap procedure extended from Efron et al. (1996) that calculates *p*-values from the change in bootstrap probabilities across replicates with differing sequence lengths. The RELL technique was developed as a less computationally expensive method than directly estimating log-likelihoods from resampled datasets (Kishino and Hasegawa, 1989; Kishino et al., 1990). RELL has been shown to accurately estimate bootstrap proportions assuming accurate model selection and adequate data, but may be biased toward type I error as is common with bootstrap methods (Felsenstein, 1985; Hasegawa and Kishino, 1994). We used the appropriate model of molecular evolution for each run in PAML. CONSEL requires specific output files of other programs as input files and we used the lnf file from PAML. We programmed CONSEL to make 1,000,000 bootstrap replicates for 10 sets. Using these tests for all 29 gene trees, we examined the support for four possible topologies (Fig. 1). These topologies postulated the phylogenetic position of *Heterochromis* as: (A) outside both the remaining African + Neotropical cichlids, (B) part of a three-way polytomy with the remaining African and Neotropical cichlids, (C) allied with the Neotropical cichlids, or (D) allied with the African cichlids.

We also used BUCKy (Larget et al., 2010) to estimate the containing species tree of all the species. BUCKy (Larget et al., 2010) uses the Bayesian concordance approach of Ane et al. (2007) to estimate a concordance factor (CF) for each clade, or the proportion of post-burn-in gene trees for all genes that include that clade, and produces a containing species tree of the clades with the highest CF. We ran BUCKy (Larget et al., 2010) with two datasets. The first included only the genes for which we had complete sampling ($n = 15$) and the second included all genes (Table 2). For both datasets, we set the analyses to run for 10,000,000 generations in two independent runs with four Markov chains in each run. We ran each of these analyses with different alphas (i.e., the relative expectation of incongruence among gene trees), including one, five, and ten. We also ran each set of these analyses three times to assess congruence among runs.

Table 2
List of nuclear loci. Name of each gene with source of primers, sequence length (bp = base pairs, bold indicates sequence for this locus was generated for all species), chromosomal position (chromosome number: megabase region) estimated from *Oryzias latipes* genome, and model of molecular evolution implemented.

Gene	bp	*Oryzias* chromosome	Model
ak1	342	12: 10.23 mb	SYM+I
acta1	**260**	3: 23.37 mb	TIM1ef+I
bmp4	**210**	22: 7.36 mb	K80
cilp1	409	3: 16.88 mb	TPM1+G
ctsh	165	21: 31.59 mb	K80
dcn	511	23: 1.68 mb	TPM2uf+G
dlx3	327	19: 2.76 mb	HKY+I
dystB	653	11: 16.02 mb	TIM3+I
enc1	**792**	12: 4.96 mb	TrNef+G
fgfr1	599	23: 3.00 mb	TPM3uf+I
fz8a	**315**	20: 6.87 mb	K80+I
glyt	**870**	16: 5.74 mb	TrNef+G
nup155	**181**	12: 12.52 mb	TrNef+G
myh	**729**	24: 11.81 mb	TrNef+I
opn4	793	17: 12.89 mb	K80+I
plagl2	**672**	7: 4.98 mb	K80+I+G
ptr	**705**	24: 11.00 mb	K80+I
rag1	1371	6: 17.34 mb	TrNef+G
rho	**677**	7: 17.09 mb	TVM+I+G
sh3px3	**705**	6: 5.08 mb	TrNef+I
sidkey	1027	24: 4.71 mb	TrNef+G
sox11a	380	22: 10.27 mb	SYM+I
sreb2	**987**	23: 4.68 mb	TrNef+I
tbr1	**639**	21: 23.07 mb	K80+I
ube3a	603	4: 29.77 mb	K80+I
ube3a-like	615	4: 29.77 mb	TVMef+G
wint7b	**220**	6: 6.04 mb	K80
zic1	**849**	20: 16.77 mb	K80+I
znf	1157	22: 9.54 mb	SYM+G

3. Results and discussion

We found the most support for *Heterochromis multidens* being allied with the remaining species of African cichlids. We made this inference of the monophyly of continental cichlid clades based not from the overwhelming support of any single method, but rather from the congruent inferences of all of our results. Results from tests of topology, e.g., AU, SH, KH, and RELL bootstrap proportions, and species tree methods were more conclusive than the phylogenies estimated from individual gene trees. This study highlights the recently identified importance and utility of employing multiple methods for the assessment of confidence in phylogenetic delimitation of monophyletic clades (Carstens et al., 2013; McVay and Carstens, 2013a).

We generated a dataset comprised of 29 non-anonymous molecular loci, more than any previous large-scale study of cichlid relationships (Farias et al., 2000, 1999, 2001; Friedman et al., 2013; Genner et al., 2007; Lopez-Fernandez et al., 2013, 2010; Smith et al., 2008). The 29 nuclear loci included 17,764 bp of sequence

data (Table 2). All new sequences are available on GenBank: KJ372530 – KJ372627. For 15 of the loci, we obtained sequences for all species and for the remaining 14 loci only one species was missing in our matrix (Table 2). It is unlikely that many of loci are highly non-independent due to physical linkage, because based on the positions in the *Orizias laticeps* genome, they are found on 14 different chromosomes with no more than three loci falling on a single chromosome. Additionally, loci on the same chromosome are minimally 0.3 megabases and up to 12.4 megabases distant from one another (Table 2). The increased availability of nuclear loci that can be sequenced across the most divergent lineages of this fish clade should enable the use of multiple nuclear gene datasets to address a substantial number of phylogenetic questions throughout the extensive radiation of cichlids.

The resolution and variation in topologies we observed for individual gene trees highlights the exceptional difficulty of determining the phylogenetic position of *Heterochromis*. The Neotropical and African (exclusive of *Heterochromis*) clades were monophyletic in the majority of gene trees. However, there was no resolution of the position of *Heterochromis* in the majority of gene trees (Table 3). *Heterochromis* grouped with the Neotropical clade in six gene trees, grouped with the African clade in eight gene trees, and in fifteen gene trees, was part of an unresolved polytomy (Table 3). The ambiguity of the position of *Heterochromis* contrasts strongly with our finding that the Neotropical species examined were monophyletic in 20 gene trees and the African species, excluding *Heterochromis*, were monophyletic in 22 gene trees (Table 3). The lack of resolution in the majority of gene trees, along with the incongruence among the gene trees, reinforces the need for our topological hypothesis testing and additional species tree approaches if the phylogenetic position of *Heterochromis* is to be resolved.

Hypothesis testing among the four topologies identified a greater number of loci supporting the monophyletic African clade, inclusive of *Heterochromis*. The best topology identified by the AU, SH, and KH tests run in CONSEL (Shimodaira and Hasegawa, 2001) and PAML (Yang, 1997, 2007) was variable among genes, but less ambiguous than the individual gene trees (Table 3). The topology with *Heterochromis* grouping with the African clade was selected 16 times, and *Heterochromis* grouping with the Neotropical clade was only selected seven times. The topology selected as best was the same for AU, SH, and KH tests, but only the AU test rejected hypotheses. For *tbr1*, the AU test rejected the hypotheses of a polytomy and of a Neotropical clade sister to an African clade with *Heterochromis* sister to this group, but did not reject the hypotheses of *Heterochromis* as sister to a monophyletic Neotropical clade. However, the AU test may not be appropriate when the many of the trees are nearly equally supported resulting in a breakdown of the theory underlying the test (Shimodaira, 2002). Nearly all of our runs in CONSEL resulted in warnings that the theory did not fit well. As suggested in Shimodaira (2002), the SH test should provide reliable results especially when there are few topological

Table 3
Phylogenetic estimation and hypothesis testing for each gene. Results of individual gene tree reconstructions are followed by summaries of results. The column labeled **Heterochromis** denotes the clade (either Neotropical or African) that *Heterochromis* was inferred to be most closely related to and posterior probabilities for these relationships are given. The **Neotropical** column indicates the posterior probability of the node representing the most recent common ancestor of the three species of Neotropical cichlids. If the posteriors are below 50, the monophyly of these groups is denoted as forming a polytomy. Similarly, **African** summarizes the posterior probability of the node representing the most recent common ancestor of the three species of African cichlids other than *Heterochromis*. The **Topological tests** column lists the best hypothesis determined using CONSEL and PAML, and within parentheses the hypotheses rejected (p-value $\leqslant 0.05$) and the test rejecting them. The **pRELL** values for the best hypothesis for each particular gene tree determined using PAML are also given. At the bottom of the table, we provide totals for the number of times *Heterochromis* was resolved as sister to the possible groupings African, Neotropical, Neotropical + African, or polytomy in the estimated phylogenies (numbers in parentheses indicate the number with posterior probabilities of $\geqslant 0.95$) and SH/KH tests (numbers in parentheses indicate the number with pRELL values $\geqslant 0.70$).

Gene	Heterochromis	Neotropical	African	Topological tests	pRELL
ak1	Neotropical 0.54	0.70	1.00	Neotropical	0.571
acta1	Polytomy	Polytomy	0.64	African	0.914
bmp4	Polytomy	Polytomy	0.73	Polytomy	0.494
cilp1	Polytomy	Polytomy	0.99	Neotropical	0.680
ctsh	Neotropical 0.71	0.71	0.99	Neotropical	0.554
dcn	Polytomy	1.00	1.00	African	0.440
dlx3	Polytomy	Polytomy	0.97	Neotropical/African	0.543
dystB	African 0.71	1.00	1.00	African	0.715
enc1	Polytomy	0.99	1.00	African	0.538
fgfr1	Polytomy	1.00	1.00	Neotropical/African	0.667
fz8a	Neotropical 0.99	0.76	Polytomy	Neotropical	0.889
glyt	African 0.98	1.00	1.00	African	0.863
nup155	African 0.83	Polytomy	0.83	African	0.741
myh	Polytomy	1.00	1.00	African	0.559
opn4	Neotropical 0.95	1.00	1.00	Neotropical	0.695
plagl2	African 0.72	0.80	1.00	African	0.563
ptr	African 0.92	0.94	1.00	African	0.840
rag1	Polytomy	1.00	1.00	Neotropical/African	0.490
rhodopsin	Polytomy	Polytomy	1.00	African	0.730
sh3px3	Neotropical 0.76	1.00	1.00	Neotropical	0.620
sidkey	African 0.81	0.99	1.00	African	0.546
sox11a	Polytomy	0.99	1.00	African	0.144
sreb2	Polytomy	0.58	0.99	African	0.577
tbr1	African 0.98	0.99	1.00	African (reject polytomy and Neotropical/African, AU)	0.824
ube3a	African 0.98	0.99	0.99	African	0.807
ube3a-like	Polytomy	0.99	1.00	African	0.522
wint7b	Neotropical 0.99	0.96	0.86	Neotropical	0.912
zic1	Polytomy	0.96	Polytomy	Polytomy	0.396
znf503	Polytomy	0.98	0.99	Neotropical/African	0.455
Totals	Neotropical = 6(3) African = 8(3) Polytomy = 15			Neotropical = 7(2) African = 16(9) Neotropical/African = 4(0) Polytomy = 2(0)	

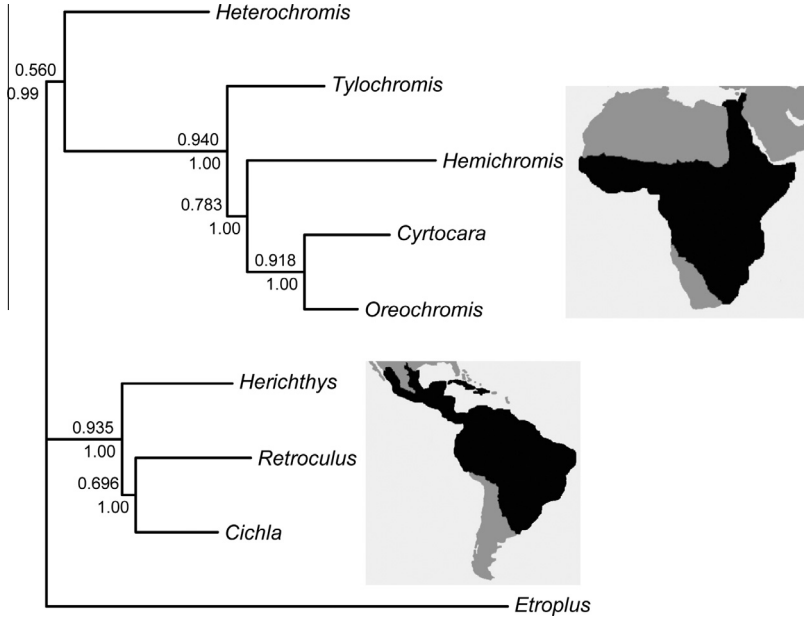

Fig. 2. Phylogeny estimated from the concatenated dataset. The Bayesian estimated phylogeny is shown annotated with node support given to the left of each node, with Concordance Factors from the BUCKy analysis (above) and Bayesian Posterior Probabilities from the MrBayes analysis (below).

hypotheses being tested, as in our study of *Heterochromis*. Eleven comparisons resulted in RELL bootstrap proportions of $\geqslant$0.7 (Table 3), of which seven supported the topology with *Heterochromis* sister to the African clade and only two supported the topology with *Heterochromis* sister to the Neotropical clade. The topology with *Heterochromis* sister to the African clade had more (n = 16) RELL bootstrap proportions that were the highest RELL bootstrap proportions for individual comparisons. This was followed by highest RELL bootstrap proportions supporting *Heterochromis* as sister to the Neotropical clade (n = 7), *Heterochromis* sister to the Neotropical plus African clades (n = 4), and only three with a polytomy among *Heterochromis,* the African clade, and the Neotropical clade.

In our phylogeny estimated from the concatenated dataset, the African and Neotropical species formed reciprocally monophyletic clades with posterior probabilities of 1.00, and *Heterochromis* was sister to the African clade with a posterior probability of 0.99 (Fig. 2). Additionally, our BUCKy (Larget et al., 2010) analyses of the species tree resulted in a topology with monophyletic African and Neotropical clades as well as *Heterochromis* recovered as sister to the African clade. The topology was the same across all runs, including those runs with different numbers of genes and different alphas. The CFs for all nodes were similar across all runs, generally ±0.05, and we reported the CFs from a run with all genes and an alpha of five (Fig. 2). The CF for the node subtending the Neotropical clade was 0.935 and for the African clade the CF was 0.940. The CF for the node uniting *Heterochromis* with the African clade was 0.560. Both our concatenated and BUCKY species tree estimations therefore inferred a topology of a monophyletic African clade that includes *Heterochromis*.

The inclusion of *Heterochromis* as part of a monophyletic group with African cichlids had more support than the alternative phylogenetic hypotheses among the individual gene trees, in the concatenated analyses, and in the BUCKy inferred species tree. However, it is important to note that the alternative topologies could not be rejected and the CF for this *Heterochromis* + African clade clearly had the least support of any node in the BUCKy inferred species tree (Table 3, Fig. 2). The difficulty of conclusively excluding the alternative topologies with 29 molecular loci and over 17,000 bp of sequence make it unsurprising that the phylogenetic position of *Heterochromis* has remained controversial (Chakrabarty, 2004; Farias et al., 2000, 1999, 2001; Kullander, 1998; Lippitsch, 1995; Oliver, 1979, 1984; Stiassny, 1991). However, our analyses did find the most support for a monophyletic African clade that contains *Heterochromis,* and our results validate the use of monophyletic, continental clades in comparative studies of cichlid diversification.

Acknowledgments

We thank J. Beaulieu and T Near for helpful discussions about this study. Two anonymous reviewers provided constructive comments that improved our manuscript. This work was partially funded through Grants from the National Geographic Society and National Science Foundation (NSF IOS-0919459).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2014.01.011.

References

Ane, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2007. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426.
Arbour, J.H., Lopez-Fernandez, H., 2013. Ecological variation in South American geophagine cichlids arose during an early burst of adaptive morphological and functional evolution. Proc. R. Soc. B-Biol. Sci. 280, 20130849.
Azuma, Y., Kumazawa, Y., Miya, M., Mabuchi, K., Nishida, M., 2008. Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. BMC Evol. Biol. 8, 215.
Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.
Chakrabarty, P., 2004. Cichlid biogeography: comment and review. Fish Fish. 5, 97–119.
Chakrabarty, P., 2006. Systematics and historical biogeography of Greater Antillean Cichlidae. Mol. Phylogenet. Evol. 39, 619–627.
Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2, 762–768.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.
Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA 104, 5936–5941.

Efron, B., Halloran, E., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. USA 93, 13429–13434.

Farias, I.P., Orti, G., Meyer, A., 2000. Total evidence: molecules, morphology, and the phylogenetics of cichlid fishes. J. Exp. Zool. 288, 76–92.

Farias, I.P., Orti, G., Sampaio, I., Schneider, H., Meyer, A., 1999. Mitochondrial DNA phylogeny of the family Cichlidae: monophyly and fast molecular evolution of the neotropical assemblage. J. Mol. Evol. 48, 703–711.

Farias, I.P., Orti, G., Sampaio, I., Schneider, H., Meyer, A., 2001. The cytochrome b gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes. J. Mol. Evol. 53, 89–103.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kahari, A.K., Keefe, D., Keenan, S., Kinsella, R., Komorowska, M., Koscielny, G., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Muffato, M., Overduin, B., Pignatelli, M., Pritchard, B., Riat, H.S., Ritchie, G.R.S., Ruffier, M., Schuster, M., Sobral, D., Tang, Y.A., Taylor, K., Trevanion, S., Vandrovcova, J., White, S., Wilson, M., Wilder, S.P., Aken, B.L., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernandez-Suarez, X.M., Harrow, J., Herrero, J., Hubbard, T.J.P., Parker, A., Proctor, G., Spudich, G., Vogel, J., Yates, A., Zadissa, A., Searle, S.M.J., 2012. Ensembl 2012. Nucleic Acids Res. 40, D84–D90.

Friedman, M., Keck, B.P., Dornburg, A., Eyton, R., Martin, C., Hulsey, C.D., Wainwright, P.C., Near, T.J., 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc. R. Soc. B-Biol. Sci. 280, 20131733.

Genner, M.J., Seehausen, O., Lunt, D.H., Joyce, D.A., Shaw, P.W., Carvalho, G.R., Turner, G.F., 2007. Age of cichlids: new dates for ancient lake fish radiations. Mol. Biol. Evol. 24, 1269–1282.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Gonzalez-Voyer, A., Kolm, N., 2011. Rates of phenotypic evolution of ecological characters and sexual traits during the Tanganyikan cichlid adaptive radiation. J. Evol. Biol. 24, 2378–2388.

Hasegawa, M., Kishino, H., 1989. Confidence limits on the maximum-likelihood estimate of the homonid tree from mitochondrial-DNA sequences. Evolution 43, 672–677.

Hasegawa, M., Kishino, H., 1994. Accuracies of the simple methods for estimating the bootstrap prabability of a maximum-likelihood tree. Mol. Biol. Evol. 11, 142–145.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580.

Hoerner, M.E., 2011. Testing for differences in rates of speciation, extinction, and morphological evolution in four tribes of cichlids endemic to lake Tanganyika, east Africa. Evolution 65, 3398–3412.

Hulsey, C.D., Keck, B.P., Hollingsworth, P.R., 2011. Species tree estimation and the historical biogeography of heroine cichlids. Mol. Phylogenet. Evol. 58, 124–131.

Hulsey, C.D., Mims, M.C., Parnell, N.F., Streelman, J.T., 2010. Comparative rates of lower jaw diversification in cichlid adaptive radiations. J. Evol. Biol. 23, 1456–1467.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, S.I., Yang, J., Lee, Y., Mastsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714–719.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum-likelihood inference of protein phylogeny and the origin of chloroplats. J. Mol. Evol. 31, 151–160.

Klett, V., Meyer, A., 2002. What, if anything, is a Tilapia? – Mitochondrial ND2 phylogeny of Tilapiines and the evolution of parental care systems in the African cichlid fishes. Mol. Biol. Evol. 19, 865–883.

Kubatko, L.S., Carstens, B.C., Knowles, L.L., 2009. STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25, 971–973.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24.

Kullander, S.O., 1998. A phylogeny and classification of the South American Cichlidae (Teleostei: Perciformes). In: Malabarba, L.R., Reis, R.E., Vari, R.P., Lucena, Z.M., Lucena, C.A.S. (Eds.), Phylogeny and Classification of Neotropical Fishes. EDIPUCRS, Porto Alegre, Brasil, pp. 461–498.

Larget, B.R., Kotha, S.K., Dewey, C.N., Ane, C., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26, 2910–2911.

Li, C.H., Orti, G., Zhang, G., Lu, G.Q., 2007. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol. Biol. 7, 44.

Li, C.H., Orti, G., Zhao, J.L., 2010a. The phylogenetic placement of sinipercid fishes ("Perciformes") revealed by 11 nuclear loci. Mol. Phylogenet. Evol. 56, 1096–1104.

Li, C.H., Riethoven, J.J.M., Ma, L.B., 2010b. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. BMC Evol. Biol. 10.

Lippitsch, E., 1995. Scale and squamation character polarity and phyletic assessment in the family Cichlidae. J. Fish Biol. 47, 91–106.

Lopez, J.A., Chen, W.J., Orti, G., 2004. Esociform phylogeny. Copeia 2004, 449–464.

Lopez-Fernandez, H., Arbour, J.H., Winemiller, K.O., Honeycutt, R.L., 2013. Testing for ancient adaptive radiations in neotropical cichlid fishes. Evolution 67, 1321–1337.

Lopez-Fernandez, H., Honeycutt, R.L., Winemiller, K.O., 2005. Molecular phylogeny and evidence for an adaptive radiation of geophagine cichlids from South America (Perciformes: Labroidei). Mol. Phylogenet. Evol. 34, 227–244.

Lopez-Fernandez, H., Winemiller, K.O., Honeycutt, R.L., 2010. Multilocus phylogeny and rapid radiations in Neotropical cichlid fishes (Perciformes: Cichlidae: Cichlinae). Mol. Phylogenet. Evol. 55, 1070–1086.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Maddison, W.P., Knowles, L.L., 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55, 21–30.

Martin, A.P., Bermingham, E., 1998. Systematics and evolution of lower Central American cichlids inferred from analysis of cytochrome b gene sequences. Mol. Phylogenet. Evol. 9, 192–203.

McMahan, C.D., Chakrabarty, P., Sparks, J.S., Smith, W.L., Davis, M.P., 2013. Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). PLoS ONE 8, e71162.

McVay, J.D., Carstens, B., 2013a. Testing monophyly without well-supported gene trees: evidence from multi-locus nuclear data conflicts with existing taxonomy in the snake tribe Thamnophiini. Mol. Phylogenet. Evol. 68, 425–431.

McVay, J.D., Carstens, B.C., 2013b. Phylogenetic model choice: justifying a species tree or concatenation analysis. Phylogenet. Evolut. Biol. 1, 114.

Near, T.J., Keck, B.P., 2013. Free from mitochondrial DNA: nuclear genes and the inference of species trees among closely related darter lineages (Teleostei: Percidae: Etheostomatinae). Mol. Phylogenet. Evol. 66, 868–876.

Niemiller, M.L., McCandless, J.R., Reynolds, R.G., Caddle, J., Near, T.J., Tillquist, C.R., Pearson, W.D., Fitzpatrick, B.M., 2013. Effects of climatic and geological processes during the Pleistocene on the evolutionray history of the Northern Cavefish, Amblyopsis spelaea (Teleostei: Amblyopsidae). Evolution 67, 1011–1025.

Oliver, M.K., 1979. *Heterochromis multidens*: the most plesiomorphic African cichlid fish? Am. Zool. 19, 892–892.

Oliver, M.K., 1984. Systematics of African Cichlid Fishes: Determination of the Most Primitive Taxon, and Studies on the Haplochromines of Lake Malawi. Ecology and Evolutionary Biology. Yale University, New Haven.

Patel, S., Kimball, R.T., Braun, E.L., 2013. Error in phylogenetic estimation for bushes in the Tree of Life. Phylogenet. Evolut. Biol. 1, 110.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Pyron, R.A., Burbrink, F.T., Wiens, J.J., 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13, 93.

Rambaut, A., Drummond, A.J., 2003. Tracer v1.4, MCMC Trace Analysis Package.

Reynolds, R.G., Niemiller, M.L., Hedges, S.B., Dornburg, A., Puente-Rolon, A.R., Revell, L.J., 2013. Molecular phylogeny and historical biogeography of West Indian boid snakes (Chilabothrus). Mol. Phylogenet. Evol. 68, 461–470.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Sanders, K.L., Rasmussen, A.R., Mumpuni, Elmberg, J., De Silva, A., Guinea, M.L., Lee, M.S.Y., 2013. Recent rapid speciation and ecomorph divergence in Indo-Australian sea snakes. Mol. Ecol. 22, 2742–2759.

Schwarzer, J., Misof, B., Ifuta, S.N., Schliewen, U.K., 2011. Time and origin of cichlid colonization of the lower Congo rapids. PLoS ONE 6, e22380.

Schwarzer, J., Misof, B., Tautz, D., Schliewen, U.K., 2009. The root of the East African cichlid radiations. BMC Evol. Biol. 9, 186.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Smith, W.L., Chakrabarty, P., Sparks, J.S., 2008. Phylogeny, taxonomy, and evolution of Neotropical cichlids (Teleostei: Cichlidae: Cichlinae). Cladistics 24, 625–641.

Stiassny, M.L.J., 1991. Phylogenetic intrarelationships of the family Cichlidae: an overview. In: Keenleyside, M.H.A. (Ed.), Cichlid Fishes: Behaviour, Ecology, and Evolution. Chapman & Hall, London, pp. 1–35.

Sturmbauer, C., Salzburger, W., Duftner, N., Schelly, R., Koblmuller, S., 2010. Evolutionary history of the Lake Tanganyika cichlid tribe Lamprologini (Teleostei: Perciformes) derived from mitochondrial and nuclear DNA data. Mol. Phylogenet. Evol. 57, 266–284.

Terai, Y., Morikawa, N., Okada, N., 2002. The evolution of the pro-domain of bone morphogenetic protein 4 (Bmp4) in an explosively speciated lineage of East African cichlid fishes. Mol. Biol. Evol. 19, 1628–1632.

Yang, Z.H., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z.H., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.