

# Pattern and Timing of Diversification in the African Freshwater Fish Genus *Distichodus* (Characiformes: Distichodontidae)

Jairo Arroyave (✉ [jarroyave@ib.unam.mx](mailto:jarroyave@ib.unam.mx))

Universidad Nacional Autonoma de Mexico <https://orcid.org/0000-0002-5389-1789>

John S. S. Denton

Florida Museum of Natural History

Melanie L. J. Stiassny

American Museum of Natural History

---

## Research article

**Keywords:** Distichodontidae, *Distichodus*, Congo Basin, molecular phylogeny, African fishes, geographic range evolution, molecular dating

**Posted Date:** April 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.2.17635/v4>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Evolutionary Biology on April 26th, 2020. See the published version at <https://doi.org/10.1186/s12862-020-01615-6>.

# Abstract

**Background :** *Distichodus* is a clade of tropical freshwater fishes currently comprising 25 named species distributed continent-wide throughout the Nilo-Sudan and most Sub-Saharan drainages. This study investigates the phylogenetic relationships, timing of diversification, and biogeographic history of the genus from a taxonomically comprehensive multilocus dataset analyzed using Maximum Likelihood and Bayesian methods of phylogenetic inference, coalescence-based species-tree estimation, divergence time estimation, and inference of geographic range evolution.

**Results:** Analyses of comparative DNA sequence data in a phylogenetic context reveal the existence of two major clades of similar species-level diversity and provide support for the monophyletic status of most sampled species. Biogeographic reconstruction on a time-scaled phylogeny suggest that the origins of the genus date back to the late Oligocene and that current geographic distributions are the result of a Congo Basin origin followed by dispersal and range expansion into adjacent ichthyofaunal provinces at different times during the evolutionary history of the group.

**Conclusions :** We present the most comprehensive phylogenetic, chronological, and biogeographic treatment yet conducted for the genus. The few instances of species paraphyly ( *D. teugelsi*, *D. fasciolatus* ) revealed by the resulting phylogenies are likely a consequence of post-divergence introgressive hybridization and/or incomplete lineage sorting due to recent speciation. Historical biogeographic findings are both in agreement and conflict with previous studies of other continent-wide African freshwater fish genera, suggesting a complex scenario for the assemblage of Africa's continental ichthyofaunal communities.

## Background

*Distichodus*, the type genus of the endemic African characiform family Distichodontidae, is a morphologically distinctive and moderately speciose lineage of endemic African freshwater fishes. *Distichodus* species are distributed across the continent, occurring throughout the freshwaters of most of sub-Saharan Africa and the river basins of the Nilo-Sudan, with representation in six of the nine ichthyofaunal provinces of continental Africa (Fig. 1). Although general aspects of the biology of the genus are poorly documented, a few studies indicate that most species are typically diurnally active and found primarily in lentic habitats shoaling in and around grasses along vegetated river banks and swamps [1]. Most species are primarily herbivorous, feeding almost entirely on periphyton, macrophytes, and detritus [2–4] [pers. obs.], although some, such as *D. lusosso*, have been characterized as dietary generalists feeding on a range of both plant and animal materials [3]. Besides playing an important role as a major constituent of the ecologically important herbivore/detrivore guilds in African freshwaters [5], *Distichodus* is also of considerable socio-economic importance, as many species constitute a highly valued, but increasingly over-exploited, component of artisanal and commercial fisheries across the continent [6], and due to their high fecundity and herbivorous diet are increasingly being cultured in fish farms and lentic water bodies, particularly in western Africa [1].

Currently, the genus contains 25 valid species [7–9], most of which are found in the Congo River basin with species diversity decreasing with distance from that central African center of diversity (Fig. 1). Although no morphological synapomorphies have yet been identified for *Distichodus*, the genus can be distinguished from all other distichodontid genera by the combination of: an upper jaw only slightly mobile with respect to the cranium; an edentulous maxilla not tightly applied posteriorly to the premaxilla; two rows (generally) of gracile, long stalked, bicuspid teeth in each jaw; a highly mobile joint between the angulo-articular and dentary (i.e., a *Distichodus*-type lower jaw [10]); a reduced dentary portion of the mandibular sensory canal; and a completely pored lateral line [10, 11].

Morphological variation within the genus includes notable differences in overall body size, spanning two orders of magnitude and ranging from over ~1 m in the largest species (*D. nefasch*, *D. langi*) to ~5 cm in the smallest (*D. decemmaculatus*, *D. teugelsi*), lateral line scale counts (large- vs. small-scaled), the position of the mouth (terminal vs. inferior), coloration (including presence and number of dark vertical bands and spots), tooth number in the oral jaws, and fin ray counts, among others [11–13] (Fig. 1).

The genus *Distichodus* was erected in the mid-19<sup>th</sup> century [14] and much of the currently recognized taxonomic diversity had been described by the early 20<sup>th</sup> century. As is typical of the taxonomic literature prior to the mid-20<sup>th</sup> century, these older descriptions are highly abbreviated, usually lacking anatomical or ecological detail, and often based on examination of little or no comparative material. In one of the earliest attempts at providing a classification scheme for *Distichodus*, Boulenger [12] divided the genus in two major groups based on the number of lateral line scales. Boulenger's classification scheme and the monophyletic status of the genus, however, were not tested until the cladistic study of Vari [10], in which the phylogenetic relationships of the Distichodontidae were investigated using comparative anatomical data. Although only five species of *Distichodus* were included in his study, Vari's findings failed to support the hypothesis of *Distichodus* monophyly, resolving some species more closely related to a clade formed by the diminutive distichodontid genera *Nannocharax* and *Hemigrammocharax*.

Contrary to Vari's work [10], the first molecular phylogenetic study focused on the Distichodontidae [15] found strong support for the monophyly of *Distichodus*, and while this study did not focus on the genus and sampling of *Distichodus* species was not exhaustive, it provided the first picture of *Distichodus* relationships. Despite this recent contribution to understanding of distichodontid relationships, taxonomic problems within *Distichodus* persist, and ongoing morphometric and morphological studies (Vreven, pers. comm.) indicate that considerable cryptic diversity remains unrecognized by current taxonomy [11, 13, 16]. Because the taxonomy of *Distichodus* has only been incidentally examined since the work of Boulenger [9, 10, 13, 15, 17], a comprehensive and focused phylogenetic treatment of the genus (including sampling of multiple individuals per species from a broad geographic range) is needed to test the current classification and to lay essential foundations for future investigations of this socio-economically important genus.

Therefore, to advance our understanding of the systematics and evolutionary history of *Distichodus*, in addition to providing insights into the processes generating fish diversity in freshwater environments of

continental Africa, this study investigates the phylogenetic, biogeographic, and chronological framework for the diversification of the genus based on multi-locus comparative DNA sequence data. The study provides a robust phylogenetic framework for testing the adequacy of the current *Distichodus* taxonomy, informing future revisionary studies and conservation actions, as well as addressing an array of questions about the evolutionary history of the genus. Furthermore, given its pan-African distribution, knowledge on the temporal and geographic context for the diversification of *Distichodus* holds considerable promise for shedding light on the very poorly understood biogeographic history of the continent's riverine networks.

## Results

### 2.1. Sequence data summary statistics, partitioning scheme and substitution models

The concatenated alignment of eight genes consisted of 6824 sites, of which 1581 were variable and 1339 parsimony-informative. The few instances of failed DNA amplification and/or sequencing resulted in <2% of missing data. The best partitioning scheme according to the PartitionFinder analysis comprise four partitions: 1) the entire mtDNA control region (*cr*), 2) 3<sup>rd</sup> codon positions of the protein-coding mitochondrial genes [*co1*, *cytb*, and *nd*], 3) 1<sup>st</sup> and 2<sup>nd</sup> codon positions of the nuclear genes [*enc1*, *glyt*, *myh6*, *shx3px3*] plus 2<sup>nd</sup> codon positions of the mitochondrial protein-coding genes, and 4) 3<sup>rd</sup> codon positions of nuclear genes plus 1<sup>st</sup> codon positions of mitochondrial protein-coding genes. The best-fit substitution models for these partitions were HKY+G+X, TrN+G+X, TrN+I+X, and TrNef+I+G, respectively. Models that include +X are those in which base frequencies are estimated using maximum likelihood rather than using the empirical frequency distributions.

For the BEAST2 analyses, all model parameter ESS values were greater than 200 and effective topological approximate ESS was always >570. All best-fit codon models for individual gene trees input to ASTRAL-III were Muse and Gaut's [18] (MG94) + M0 + F3x4 codon frequency models, with the exception of *myh6*, for which an MG94 + M3 + F3x4 model was inferred. Terminology for the number of omega ( $\omega$ ) classes follows Yang *et al.* [19].

### 2.2. *Distichodus* phylogeny

The phylogeny derived from ML analysis (RAxML tree) of the concatenated alignment of all eight markers is presented in Fig. 2. A summarized version of this phylogeny, highlighting interspecific relationships, is illustrated in Fig. 3. Single-locus phylogenies (*enc1*, *glyt*, *myh6*, *sh3px3*, mtDNA) are presented in Figs. 1S-5S, respectively. As expected, partially because of variation in substitution rates, single-locus phylogenies differed in the level of resolution and nodal support, with ncDNA markers resulting in less resolved and supported phylogenies when compared to the mtDNA locus.

Species-tree analyses (SVDquartets and ASTRAL-III) results are presented in Figs. 4 and 5, respectively. BEAST2 analyses yielded very similar topologies (Figs. 6 and 6S-19S), only differing slightly in resolution within one of the two main clades discovered. The RAxML, SVDquartets, and BEAST2 phylogenies exhibit largely congruent topologies with comparable nodal support, resolving the genus into two strongly supported major clades of roughly equivalent species diversity and with the same limits and composition (clades A and B in Figs. 2-4, 6, and 6S-19S). While these three different analytical methods revealed the same general pattern of relationships in clade A, with disagreement inside clade B (notably among *D. engycephalus*, *D. kasaiensis*, *D. lusosso*, and *D. atroventralis*), the ASTRAL-III analysis produced a considerably different topology (Fig. 5). The source of this disagreement with the other methods is unclear.

Regardless of inference method, and conforming to expectation, nodal support was greater at deeper divergences, while weaker (BS<75; PP< 0.75) at nodes corresponding to more recent divergences, likely reflecting intraspecific population-level structuring (when sampling multiple individuals per species). Nonetheless, for the most part, interspecific relationships are well supported, with the exception of a subclade of clade B.

### 2.3. Monophyly of *Distichodus* species

Sampling of multiple individuals per species allowed testing of the monophyletic status of most morphologically diagnosed *Distichodus* species, and the resulting total evidence phylogeny (Figs. 2 and 3) strongly supports the monophyly of most of the species for which multiple individuals were sampled. However, there are two notable exceptions: the species pairs *D. teugelsi* / *D. decemmaculatus*, and *D. fasciolatus* / *D. schenga*, each of whose members were resolved as paraphyletic with respect to the other. Specifically, the phylogenetic placement of all sampled individuals of morphologically determined *D. teugelsi* renders *D. decemmaculatus* paraphyletic, and similarly, the phylogenetic placement of the two sampled individuals of *D. schenga* renders *D. fasciolatus* paraphyletic (Fig. 2). Although based on considerably fewer comparative data, the mtDNA phylogeny agreed, for the most part, with the concatenated phylogeny in the monophyly of most sampled species. Most ncDNA single-locus phylogenies, on the contrary, exhibited lower degrees of resolution and support than the total evidence and mtDNA trees, failing to support the monophyletic status of several of the species evaluated.

### 2.4. Timescale of *Distichodus* diversification

The resultant chronograms from the BEAST2 analyses are presented in Figures 6 and 6S-19S, and a summary of the results including age estimates and associated HPD intervals of select nodes in Table 1. A number of findings are apparent regardless of calibration strategy, and therefore, of absolute times of divergence. Notable among these are that *Distichodus* (crown group) originated shortly after its

divergence from *Paradistichodus*, and that the two major components of the *Distichodus* radiation (clades A and B) started diversifying roughly concurrently. However, despite this initial chronological correspondence, a large subclade of clade B consisting of seven species (the MRCA of *D. kasaiensis* and *D. atroventralis* and all of its descendants) is, for the most part, of comparatively more recent origin.

Of the main variables defining calibration strategy (i.e., calibration node and  $P_{95}$  SMB), selection of calibration node appears to have the strongest effect on estimates of divergence times, with node  $D_{ne}+D_{ro}$  resulting in the oldest node age estimates (substantially older than those based on any of the other calibration nodes used), irrespective of  $P_{95}$  SMBs. However, node age estimates based on calibration node P+D did not differ considerably from those based on calibration node D, especially under equivalent  $P_{95}$  SMBs. This trend can be explained by the fact that the age difference between these nodes is relatively small, as previously mentioned. Unsurprisingly, older  $P_{95}$  SMBs resulted in older node age estimates, although perhaps not as much as anticipated.

According to the results of analysis 8 (Fig. 6), under what could be considered a “midway” calibration strategy, intermediate in terms of calibration node (D, crown) and  $P_{95}$  SMB (30 Ma), the origins of the *Distichodus* crown group date to the late Oligocene (24.1 Ma; 95% HPD=19.44-30.20). Conforming to expectation, this estimate is older (~7 Ma) than the only previously published estimate, inferred in the context of a time-scaled phylogeny of the Distichodontidae (17.22 Ma; 95% HPD=12-23) [15]. The results from analysis 8 also indicate that by the late Miocene/early Pliocene (~5 Ma) the bulk of species diversity in the genus was already present. Furthermore, this chronogram implies that stem lineages leading to the modern species *D. hypostomatus*, *D. maculatus*, *D. engycephalus* appeared around 21-18 Ma, while most remaining modern diversity likely originated during the late Miocene. Notably, the most recent divergences (~1 Ma) correspond to the seemingly paraphyletic species pairs *fasciolatus/shenga* and *teugelsi/decemmaculatus* mentioned above, an observation that supports the notion that each of these pairs may correspond to lineages at the early stages of differentiation and speciation.

## Table 1.

### 2.5. Geographic range evolution on the *Distichodus* phylogeny

Model comparison using AIC and AIC weights (Table 2) indicate support for the M1 model (CB-as-source) over the M2 model (CB-as-sink), while the unconstrained (M0) model received negligible support, regardless of absolute times of divergence (input chronogram). Likewise, the pattern of range shifts out of and expansions from the Congo Basin (the ancestral area) implied by the preferred model (M1) was equivalent across analyses, irrespective of absolute node ages and despite minor topological differences between input chronograms (particularly with respect to the relative placement of *D. engycephalus*). Specifically, the M1 model inferred six range shifts for *Distichodus* out of the Congo Basin (the ancestral

area) and three different range expansions from the Congo Basin to include adjacent ichthyofaunal provinces (Figs. 7, 20S, and 21S). Support/signal for model M1, however, appears to be stronger when based on older times of divergence (Table 2).

Table 2.

## Discussion

### 3.1. *Distichodus* relationships and taxonomic implications

Here we present the first comprehensive phylogenetic, temporal, and biogeographic framework for examination of the current taxonomy and evolutionary history of *Distichodus* and for future evolutionary studies of the genus. Regardless of analytical method (except for ASTRAL-III, but see below), our results, based on a dataset with considerably more inclusive taxon, character, and geographic sampling for *Distichodus* than previous works, strongly support the existence of two roughly equal-sized, and reciprocally monophyletic lineages within the genus, while corroborating the monophyletic status of most currently recognized species. RAxML, SVDquartets, and BEAST2 topologies are largely congruent, with some swapping of taxa inside the two-clade *Distichodus* structure that is also supported by morphology [10]. Only the ASTRAL-III analysis did not conform to this general picture of *Distichodus* relationships, but there are two confounding issues then at play. First, ASTRAL-III may be sensitive to gene tree estimation error. The codon model approach to gene tree inference used here should, in principle, be the most accurate method for gene tree inference (of those currently available reversible Markov models), given its hierarchical modeling structure [20]. However, it is still not immune to the requirements of large amounts of data [21]. Some close-to-zero-length branches in the individual gene trees may either be (1) a true artifact of ILS, or (2) a consequence of insufficient data (in terms of gene length, or in terms of evolutionary rate distribution). However, distinguishing differences between real ILS and insufficient data is not possible from the current analysis. We therefore focused our attention and present our conclusions based on the results with the largest overall congruence.

Conforming to expectation, nodal support tends to be higher at more basal nodes (deeper divergences), whereas more recent divergences are, on average, less strongly supported. Low nodal support and instances of conflict between analytical methods of phylogenetic inference are particularly evident for the clade consisting of the predominantly large-bodied species (*D. atroventralis*, *D. lusosso*, *D. kasaiensis*, *D. antonii*, and *D. sexfasciatus*), suggesting that additional data will be necessary to resolve interspecific relationships for this particular section of the tree. Disagreement between methods in this part of the *Distichodus* tree is an aspect worth revisiting in the future with genome-wide NGS-generated data, as larger amounts of DNA sequence data might be capable of better resolving and supporting these divergences. Besides the obvious reasons for wanting to unambiguously resolve this part of the *Distichodus* tree, such an endeavor is of special interest because the highly disparate trophic-related

morphologies displayed by members of this clade are undoubtedly an interesting character system from both evolutionary ecological and functional morphological perspectives.

Despite some of the disagreements between inference methods, our results offer a general working hypothesis of *Distichodus* relationships and, with few exceptions, are consistent with the current species-level morphology-based taxonomy of the group. Instances of questionable species monophyly and therefore in conflict with the current classification are discussed below.

### 3.1.1. Paraphyly of *D. teugelsi* with respect to *D. decemmaculatus*

Problems with the species recognition of the two dwarf species, *D. decemmaculatus* and *D. teugelsi*, have been noted by Verheyen et al. [16], and are confirmed here. Species identification has previously been based on the presence (*decemmaculatus*) or absence (*teugelsi*) of a series of dark spots or bars along the flanks, and 20 (*decemmaculatus*) versus 16 (*teugelsi*) scales around the caudal peduncle [8]. While our study finds strong support for a *teugelsi/decemmaculatus* clade, samples tentatively identified as *D. teugelsi* from the Kwilu River in the Kasai basin (with 16-17 scales around the caudal peduncle and variously marked spots or bars along the flanks), form a well-supported sister clade to the remaining samples. While samples of *D. teugelsi* from the type locality, the Lefini River, a right bank tributary of the Congo River upstream of Pool Malebo (lacking spots or bars on the flanks and with 16 scales around the caudal peduncle), form a clade sister to the *D. decemmaculatus* samples, all tentatively identified here as *D. decemmaculatus* or *D. cf. decemmaculatus*. Among these we record caudal peduncle scales counts ranging from 18-20, and flank pigmentation ranging from virtually absent to clearly marked and strongly spotted. While no taxonomic solution is proposed here, based on the molecular analysis presented and the observation of high variability in both pigmentation and scale counts in geographically disparate samples of both “species”, further study of the *teugelsi/decemmaculatus* clade, including representatives of populations across the range of each putative taxon, is needed. We note further that, as for the *fasciolatus/shenga* species pair discussed below, our estimation of the divergence time for the *teugelsi/decemmaculatus* pair (based on analysis 8; Fig. 6) is among the most recent (~1 Ma). The fact that this is a very recent divergence, might explain the resulting paraphyletic pattern. It is well known that genetic variation shared between closely related species can be due to retention of ancestral genetic polymorphisms resulting from incomplete lineage sorting (ILS) [22], a process that can confound phylogenetic inference and hinder robust tests of monophyly in recently diverged species pairs. Whereas mtDNA introgressive hybridization has been also recognized as one cause of misleading inferences of paraphyly, the overall congruence between the nc- and mtDNA signal involving the *teugelsi/decemmaculatus* pair supports ILS instead of introgression as a probable explanation for the observed pattern of paraphyly [23].

### 3.1.2. Paraphyly of *D. fasciolatus* with respect to *D. schenga*



Representatives of the widespread species *D. fasciolatus* are rendered paraphyletic by the placement of the two sampled individuals of the southern African species, *D. schenga*, a middle and lower Zambezi endemic, which are placed well nested within a strongly supported *D. fasciolatus* clade (Fig. 2). *Distichodus schenga* (type locality Tete, Zambesi River) was described by Peters in 1852 and *D. fasciolatus* by Boulenger in 1898 (type localities in the lower Congo River region), and the descriptions of both are minimal, not allowing for morphological species discrimination. Possibly because of this, Boulenger [12] did not include *D. schenga* in his key to *Distichodus*, and by implication did not recognize it as distinct from *D. fasciolatus*. Our molecular data clearly suggest that the synonymy of *D. fasciolatus* with *D. schenga* is in order, however ongoing morphometric and morphological study of the entire “*fasciolatus*-complex” is currently underway (Vreven, pers. comm.), and pending the results of that study we defer proposing a formal taxonomic synonymy based solely on our molecular data and minimal sampling of putative *D. schenga* from across the Zambezi basin.

We do note however, that the phylogenetic and chronological pattern revealed by our study (Figs. 2, 4, and 6), coupled with the allopatric distribution of these two taxa, suggest that populations currently recognized as *D. schenga* could have diverged from a lineage/population of *D. fasciolatus* that colonized the Upper Zambezi headwaters from the Kasai during the Pleistocene, when the two river systems shared a past connection [24–26]. This chronological and geographic dispersal scenario out of the Congo Basin is consistent with our estimated divergence time for this species pair (~1 Ma) (Fig. 6) and the inferred range shift involving *D. schenga* (Fig. 7), and has been hypothesized for various other fish taxa across the headwaters of the Congo-Zambezi watersheds [27–29].

In any case, a phylogenetic pattern of short, poorly supported branches is an indicator of recent species divergence that precluded mtDNA lineages from sorting to reciprocal monophyly [30]. Therefore, as for the *teugelsii/decemmaculatus* pair, we cannot rule out the possibility that the inferred paraphyly of *D. fasciolatus* with respect to *D. schenga* is an artifact of ILS issues. However, signal discordance between mtDNA and nuclear markers involving the *fasciolatus/schenga* pair strongly suggest that post-divergence introgressive hybridization could also explain the inferred paraphyly of *D. fasciolatus* with respect to *D. schenga*.

### 3.2. A spatiotemporal framework for *Distichodus* diversification

In the context of their time-scaled phylogeny of the suborder Citharinoidei, Arroyave et al. [15] were among the first authors to estimate an age for the origin of *Distichodus* and the timing of diversification within the genus. Their chronogram suggested that the *Distichodus* crown group appeared in the Miocene (~17 Ma), but that most of the species diversity likely originated during the past 5 Ma. These inferences were based on a molecular clock calibrated using ~7.5 Ma *Distichodus* fossilized dentition [31], which at the time was the oldest known fossil assignable to the genus. The recent discovery of a considerably older (18-19 Ma) *Distichodus* fossil [32], however, prompted our reexamination of the timescale of *Distichodus* diversification in the context of a larger dataset, both in terms of molecular markers (8 vs. 7

loci) and taxon sampling (20 vs. 16 spp.). This older fossil, however, presented us with the challenge of accurately assigning it to a node for the purpose of calibrating the molecular clock and estimating absolute times of divergence in the phylogeny of *Distichodus*.

Whereas the approach devised herein to address the uncertainties associated with the fossil-based calibration of the molecular clock resulted in multiple alternative chronograms, from our knowledge of the study subject we believe that some of these alternative calibration scenarios might be either overly conservative (e.g., D+P stem) or too liberal (e.g.,  $D_{ne}+D_{ro}$  crown), therefore possibly resulting in under- or overestimation of node ages, respectively. Nonetheless, because we have no means to empirically falsify any of these alternative calibration scenarios, we consider it important and valuable to offer the reader the possibility of choosing among alternative scenarios (including those we think too extreme) based on their own knowledge of the study subject and their personal beliefs regarding best practices for justifying fossil calibrations [33].

For the most part, our findings imply a temporal framework for the diversification of *Distichodus* older than previously reported [15], irrespective of calibration strategy. Only analyses 1 and 2, based on calibration node D+P (stem), resulted in younger divergence time estimates (Table 1). While at present we feel more comfortable grounding our discussion of the temporal and geographic context for the diversification of *Distichodus* in the results from analysis 8 (intermediate in terms of calibration node [D, crown] and  $P_{95}$  SMB [30 Ma]), we acknowledge that, should this calibration fossil be eventually confirmed as *D. nefasch*, *D. rostratus*, or their MRCA (a possibility due to fossil tooth shape, size, and geographic distribution), a reinterpretation of the biogeographic history will be necessary to reconcile the inferred patterns of geographic range evolution with a chronological framework more than twice as old as the one discussed below (Fig. 21S).

Despite the high ichthyofaunal diversity of Afrotropical continental waters, few studies have investigated the chronological and biogeographic context for the diversification of African freshwater fish clades, among which only a handful have focused on Pan-African riverine genus-level radiations, namely *Hydrocynus* [27], *Mastacembelus* [34, 35] and the species-rich *Synodontis* [29, 36]. Notably, a Miocene diversification for *Distichodus*, as implied by the chronogram resulting from analysis 8 (Fig. 6), broadly concurs with previous findings for both *Mastacembelus* and *Synodontis* [29, 35, 36]. Similarly, a concurrence of Miocene diversification among various lineages of fishes, frogs, and crabs has been pointed out by Daniels et al. [37], who suggest this likely reflects a shared response to mesic climatic shifts resulting in marked allopatric differentiation among each of these freshwater lineages during the mid- to late Miocene. While to our knowledge there are no empirical studies proving a causal relationship between particular paleohydrological events and diversification patterns in African freshwater fishes, some authors have suggested that Miocene tectonic and climatic upheaval may have influenced or even triggered diversification [28, 35]. The Miocene geological epoch was the setting for widespread epeirogenic uplift in Africa and global climate change that profoundly contributed to shaping the modern African hydrological landscape [38–40], which in turn, it is believed, promoted diversification in freshwater fishes as a consequence of river discharge shifts (due to climate change) and drainage

disruption and modification (due to rifting) [41]. Our findings about the timing of *Distichodus* diversification add to instances of Miocene continent-wide freshwater radiations, and therefore to a growing body of evidence in support for a “hydrogeological” hypothesis, that paleohydrological and paleoclimatic changes promote landscape evolution which in turn promotes cladogenesis in freshwater organisms [42, 43]. Further research, however, ideally in a multi-taxon comparative framework, is certainly needed to robustly test hypotheses of concerted responses to paleogeologic and paleoclimatic scenarios.

Analysis of geographic range evolution on the phylogeny of *Distichodus* favored a biogeographic model in which the Congo Basin (CB) is the center of origin (ancestral area) and source of the geographic diversity of the genus, irrespective of absolute times of divergence. In particular, the biogeographic reconstruction based on the chronogram resulting from analysis 8 (Fig. 7), implies that most cladogenetic events occurred in lineages still confined to the CB throughout most of the Miocene, but also multiple lineage range shifts out of and expansions from the CB into adjacent ichthyofaunal provinces at different times during the evolutionary history of the group. Only in the late Miocene (~9-7 Ma) are the first recorded instances of range shifts out of the CB and of cladogenesis occurring in other ichthyofaunal provinces, namely the Nilo-Sudan (NS) and Lower Guinea (LG). The remaining instances of range shifts and expansions are more recent, dating back to the Pliocene. While most ichthyofaunal provinces appear to have been colonized only once (or twice in the case of UG), our results indicate that LG was independently colonized by five different lineages, mostly during the Pliocene.

Our reconstruction of the biogeographic history of *Distichodus* suggests a central role of the CB in the distribution of the continent’s freshwater ichthyofauna during the late Cenozoic, offering support to the hypothesis that the CB is the source of the ichthyofauna of less diverse river basins throughout continental Africa [41]. While a CB origin has also been postulated for the African tigerfish *Hydrocynus* [27], other continent-wide African freshwater fish genera such as *Synodontis* [29] and *Mastacembelus* [35] do not conform to this pattern and suggest repeated independent colonization into the CB. Considering the vast geographic area under study, and that complex evolutionary histories of dispersal and vicariance are likely to exist among the different fish lineages, these conflicting biogeographic histories certainly suggest a complex scenario for the assemblage of the continent’s ichthyofaunal communities.

## Conclusions

The spatiotemporal framework for the diversification of African freshwater fish genus *Distichodus* presented herein provides a significant advance in our knowledge of the evolutionary history of this ecologically and socio-economically important group of fishes. With few exceptions, the resulting phylogeny is consistent with the current species-level taxonomy of the group, offering a working hypothesis of *Distichodus* relationships that will serve as phylogenetic framework for future evolutionary studies involving phenotypic and genomic systems. The few instances of species paraphyly (*D. teugelsi*, *D. fasciolatus*) revealed in our favored phylogeny are likely a consequence of introgression and/or incomplete lineage sorting due to recent speciation. Therefore, we refrain from making taxonomic/nomenclatural changes pending further morphological assessment based on a larger sample

of comparative material. While analysis of geographic range evolution favored a biogeographic scenario in which the Congo Basin is the source of geographic diversity of the genus, this finding is both in agreement and conflict with previous studies of other continent-wide African freshwater fish genera, suggesting a complex scenario for the assemblage of Africa's continental ichthyofaunal communities.

## Methods

### 5.1. Taxon Sampling

Ingroup sampling consisted of 133 specimens representing 20 of the 25 valid *Distichodus* species, thereby encompassing 80% of *Distichodus* currently recognized diversity (Table 3). *Distichodus brevipinnis*, *D. langi*, *D. mossambicus*, *D. rufigiensis*, and the newly described *D. ingae* [9], were not included in analyses due to unavailability of tissues. With the exception of *D. altus*, *D. nefasch*, *D. rostratus*, and *D. petersii*, for which only a single tissue sample was available, multiple individuals per species were sequenced to sample as large a portion of each species' range as possible (Table 3). In addition to increasing geographic coverage, inclusion of multiple individuals per species allowed for testing the monophyletic status—and therefore species limits under the phylogenetic species concept [44, 45]—of nominal species from which more than one individual was available for sequencing. Sampling of multiple individuals per species, however, was not aimed at making inferences about tokogenetic (intraspecific) relationships and/or phylogeographic patterns. *Paradistichodus dimidiatus* was included as outgroup based on the findings from a relatively recent molecular phylogenetic study that investigated relationships of the Distichodontidae [15], which resolved the monotypic genus *Paradistichodus* as the sister group of *Distichodus*. Similarly, *Nannocharax ansorgii* was included as additional and outermost outgroup for molecular dating and inference of geographic range evolution analyses.

Most tissue samples were obtained from specimens collected during recent expeditions in West and West-Central Africa by a research team from the American Museum of Natural History (AMNH) (led by co-author MLJS). Specimens were handled and euthanized prior to preservation in accordance with recommended guidelines for the use of fishes in research [46] and stress was ameliorated by minimizing handling and through the use of the anesthetic Tricaine mesylate (MS-222) for euthanasia. Tissue samples were taken in the field and immediately preserved in 95% ethanol. Voucher specimens were fixed in formalin and subsequently transferred to 70% ethanol for long-term storage. Data for specimens cataloged and stored in the ichthyology collection of the AMNH, are available online at <http://sci-web-001.amnh.org/db/emuwebamnh/index.php>.

Specimen collection was made in accordance with ethical and legal guidelines for international animal research approved by the AMNH Institutional Animal Care and Use Committee (IACUC) (approval #36/06). The AMNH IACUC has guidelines relating to studies involving its members in different countries, and this study conforms to those guidelines. Specimen collection and exportation of samples used in this study follow institutional and national ethical and legal guidelines of the Ministry of Fishery and Aquaculture, Republic of Guinea, No. 65/MPA/DGAGSP/11; the Ministry of Scientific Research and

Technical Innovation, Republic of Congo, No. 031/MRSIT/DGRST/GERBID.06.13; and the Ministry of Agriculture and Fisheries, Democratic Republic of Congo, No. 037/DP/SG/AGRIPEL/16.

Additional samples were obtained from colleagues at the Cornell University Museum of Vertebrates (CUMV), the Royal Museum for Central Africa (MRAC), and the South African Institute for Aquatic Biodiversity (SAIAB). Voucher specimens are deposited in the ichthyology collections of the AMNH, CUMV, MRAC, and SAIAB. Species identity of non-AMNH vouchers was confirmed either by direct examination of loaned specimens, photographs provided, or on taxonomic authority of the loaning institution. Voucher catalog numbers and GenBank accession numbers for the gene sequences generated and included in this study are listed in Table 3.

### Table 3.

## 5.2. Gene sampling and nucleotide data collection

Eight gene fragments, including the seven protein-coding loci sampled by Arroyave et al. [15] to address distichodontid interrelationships (*co1*, *cytb*, *enc1*, *glyt*, *myh6*, *nd2*, and *sh3px3*) were sequenced. Additionally, a faster-evolving mitochondrial non-coding marker, control region (*cr*), was added to address more recent divergences within the genus. DNA sequence data was generated from a total of 133 *Distichodus* individuals. General procedures for DNA extraction, amplification, and purification, along with primers and thermal profiles for sequencing the protein-coding genes used in this study follow Arroyave and Stiassny [47] and Arroyave et al. [15]. *Distichodus*-specific primers for *cr* (*cr\_Dist\_f*: 5'-AGCGCCGGTCTTGTAATCCG-3'; *cr\_Dist\_r*: 5'-TGCTTGTGGAACCTTCTAGGGTCCAT-3') were designed using the software Primer3 [48] from conserved flanking regions of aligned mtDNA control region sequences extracted from the two distichodontid complete mitochondrial genomes available in GenBank (*Distichodus sexfasciatus* AB070242 and *Ichthyborus* sp. AP011993). Amplification of *cr* via PCR was carried out using the following thermal profile: 5-min initial denaturation at 95°C, followed by 35 cycles of denaturation at 95°C for 60 s, annealing at 58°C for 60 s, and extension at 72°C for 120 s, followed by a 10-min final extension at 72°C.

## 5.3. Sequence editing and partitioning scheme/substitution model selection

Contig assembly and sequence editing was performed using Geneious v.11.0.2 [49]. IUPAC nucleotide ambiguity codes were used to represent heterozygous sites. The resulting sequences were trimmed to exclude primer regions and examined for appropriateness/homology using BLASTx [50]. Each gene was aligned using MUSCLE [51] under default parameters as implemented in Geneious, followed by

concatenation of individual alignments. All sequences were checked for stop codons and for miscalled amino acids by examining translation alignments.

Best-fit partitioning schemes and models of molecular evolution for the nucleotide data were determined using PartitionFinder2 [52] based on 22 pre-defined data blocks: the non-coding mtDNA control region (1 block) plus the 1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup> codon positions of the seven protein-coding genes (3 positions x 7 genes). The PartitionFinder2 greedy algorithm was employed to search for an optimal scheme under the assumption of independent model parameters and branch lengths for each partition. Selection of the partitioning scheme and models over the set of schemes and models produced during greedy search was accomplished using the Schwarz/Bayesian Information Criterion (BIC) [53].

#### **5.4. Phylogenetic, biogeographic, and chronological analyses**

Various analytical approaches were employed to infer phylogenetic relationships in *Distichodus* from the multilocus dataset generated in this study, one of which also simultaneously estimates absolute times of divergence in the resultant phylogeny. The results from the latter approach were subsequently used in analyses for testing historical biogeographic hypotheses of geographic range evolution in *Distichodus*.

##### **5.4.1. Maximum Likelihood (ML) estimation of phylogeny**

Phylogenetic analysis of the concatenated alignment of the eight sampled genes under a Total Evidence/Simultaneous Analysis [54, 55] approach was performed using the ML optimality criterion. Furthermore, to examine the degree of variation in topology, resolution, and clade support among the individual sampled loci, and to complement the inferences made from the simultaneous analysis of all markers, each of the nuclear genes (*enc1*, *glyt*, *myh6*, *sh3px3*) and a concatenated alignment of the mitochondrial genes (*co1*, *cr*, *cytb*, *nd2*; effectively inherited as a single locus), were independently analyzed, also using the ML optimality criterion. ML phylogenetic analyses were conducted with RAxML v.8 [56] through the CIPRES Science Gateway v.3.3 [57] as a single partition under the GTRGAMMA model with four rate classes using full ML optimization for the tree search and 1000 rapid bootstrap (BS) searches to assess nodal support [58].

##### **5.4.2. Species-tree approaches**

Although concatenation methods have been suggested to often perform well when incomplete lineage sorting (ILS) levels are low [21], the degree of ILS in *Distochodus* is unknown. To explore the outcomes of ILS-aware species-tree analyses relative to concatenation, both SVDquartets [59] and ASTRAL-III [60] were employed. SVDquartets has been suggested to perform well with low ILS and small

numbers of sites per gene, and ASTRAL methods have been suggested to perform well under high ILS conditions, but may be sensitive to small numbers of sites per gene [21]. SVDquartets analysis was conducted in PAUP\* v4.0a164 [61] sampling all ~8.6 million quartets under the multispecies coalescent on the full dataset, using the default QFM quartet assembly method. Bootstrap support values were assembled onto the SVDquartets tree using the *sumtrees* command in the DendroPy package [62]. Gene trees input to ASTRAL-III were estimated from best-fit codon models inferred in codonPhyML [63] under default search intensity, using custom R scripts written by the authors. Because the mitochondrial genome does not undergo recombination and is inherited as a single locus, the three protein-coding mitochondrial genes were fit with a single codon model and inferred gene tree. Gene trees for each autosomal locus were inferred separately.

### 5.4.3. Bayesian co-estimation of phylogeny and divergence times

Prior to co-estimation of phylogeny and divergence times, a new data matrix was created from the original multi-individual, multi-locus matrix, by including DNA sequence data from only a single individual per species, from or near the type locality whenever possible (for each sampled species, the first individual listed in Table 3). The resulting reduced matrix was analyzed in BEAST v.2.5.0 [64] under the optimal partitioning scheme and substitution models suggested by the PartitionFinder2 analysis. Node ages were estimated using a Bayesian relaxed-clock method [65] under the uncorrelated lognormal (UCLN) rate variation model, and assuming a birth-death process prior for topology and divergence times. By default, the prior on the mean parameter of the UCLN clock model (*uclMean.c*) is a uniform distribution on the interval (0,∞), which is an uninformative and improper prior (it does not integrate to 1). Although improper priors can sometimes lead to proper posterior distributions, they may also have undesired effects and cause problems with mixing and convergence [66]. Based on previous findings regarding substitution rates in *Distichodus* [15], we assumed a log-normally distributed prior for the clock rate (*uclMean.c*) with hyperparameters  $\mu=0.003$  and  $\sigma=0.5$ . On the other hand, the standard deviation parameter of the UCLN clock model (*uclStdev.c*) is by default assigned a gamma distribution prior. Variation in substitution rates among branches in *Distichodus*, however, appears to be low in general [15]. Accordingly, we assumed an exponential prior distribution with 95% of the probability density on values < 1 for the standard deviation of the UCLN (*uclStdev.c*).

The molecular clock was calibrated based on early Miocene (*ca.* 18 Ma) fossilized dentition attributable to *Distichodus* recovered from deposits of the Maradah Formation in Jabal Zaltan, Libya, by far the oldest fossil unambiguously assignable to the genus [32]. In fact, this fossil pushes back the first known appearance of *Distichodus* in the fossil record by 10 Ma with respect to the *Distichodus* calibration fossil used by Arroyave et al. [15] to infer a time-scaled phylogeny of citharinoid fishes. Although the Maradah fossil is unquestionably diagnostic of *Distichodus* (tall, slender necked tooth with a bifid apex bearing characteristically short and rounded lobes) and could potentially be ascribed to either *Distichodus nefasch* or *D. rostratus* on the basis of size and geographic distribution, its exact phylogenetic placement

is unknown. The absence of relevant comparative morphological data in a phylogenetic context to which to integrate the fossil taxon, coupled with its fragmentary nature, renders it difficult to confidently assign it to a particular node and to determine whether it should be used to constrain the age of the stem or the crown group of the calibration node. Because of this phylogenetic uncertainty, along with the challenge of objectively establishing a maximum age constraint to the calibration node, we conducted a series of analyses (Table 4) to assess the robustness of node ages to analytical ambiguity and to offer alternative output scenarios based on a variety of reasonable input parameters, particularly with respect to the phylogenetic placement of the calibration node and its maximum age constraint. Specifically, we used three alternative calibration nodes: 1) MRCA of *Distichodus* and *Paradistichodus* (D+P), 2) MRCA of *Distichodus* (D), and 3) MRCA of *D. nefasch* and *D. rostratus* ( $D_{ne}+D_{ro}$ ). The rationale behind this proposal is that, at the very least, the calibration fossil could be used to constrain the age of divergence between *Distichodus* and its sister group, *Paradistichodus*, but under more liberal phylogenetic designations, it could also be used to constrain the age of the entire genus or even the divergence between the species *D. nefasch* and *D. rostratus*. Furthermore, each calibration node was constrained both as stem and as crown group. Additionally, the temporal uncertainty of calibration nodes was modeled using log-normally distributed priors with a hard minimum bound set by the age of the fossil (18 Ma) and one of three alternative 95<sup>th</sup> percentile soft maximum bounds ( $P_{95}$  SMBs): 20, 30, and 40 Ma (Fig. 8; Table 4). The combinatorial exercise of choosing one of three alternative calibration nodes, constrained as stem or crown, and modeled by a log-normally distributed prior characterized by one of three alternative  $P_{95}$  SMBs, resulted in 18 different analyses (although effectively 15 since the node representing the MRCA of *Distichodus* as stem is equivalent to the node representing the MRCA of *Distichodus* and *Paradistichodus* as crown (Table 4). In each analysis, root age was indirectly constrained (as an implied prior) by the combined effects of the calibration prior on other internal node and the prior for topology and divergence times (birth-death process).

BEAST2 analyses were implemented using the Markov Chain Monte Carlo algorithm (MCMC) run for 50 million generations sampled every 1000 generations, under default proposal mechanisms and default priors for the parameters of the birth-death branching process used to provide the prior distribution for the non-calibration nodes (speciation and extinction rates) and the model of molecular evolution for each gene (substitution rates, base frequencies, gamma shape, and proportion of invariant sites). Convergence model parameter estimates were assessed via ESS values over 200, using Tracer v.1.7 [67]. Sufficient sampling of the estimate of the tree topology (ESS > 200) was determined by dividing the topological approximate ESS by the generation number of the approximate earliest stationary value in the topological autocorrelation plot, generated in the R package *rwt* [68]. Further assessment of MCMC convergence was undertaken by examination of the average standard deviation of split frequencies, with values << 0.01 taken as indicative of stationarity. All analyses used a 10% burn-in. A maximum clade credibility (MCC) topology was inferred using TreeAnnotator v.2.5 [64], resulting in a chronogram indicating posterior probabilities (PP) and mean ages of all nodes with their associated 95% highest posterior density (HPD) intervals.



## Table 4.

### 5.4.4. Inference of geographic range evolution

The evolution of geographic ranges in *Distichodus* was investigated using the null-range-excluded dispersal-extinction-cladogenesis model (DEC\*) [69], a modified version of the original likelihood-based dispersal-extinction-cladogenesis (DEC) model [70, 71]. The set of discrete geographic areas for the DEC\* analysis consisted of the six Afrotropical ichthyofaunal provinces of Roberts [72] (modified by Lévêque [73]) with presence of *Distichodus* species: Congo Basin (CB), Zambezi (Z), Nilo-Sudan (NS), Upper Guinea (UG), Lower Guinea (LG), and East Coast (EC) (Fig. 1). African ichthyofaunal provinces were delimited on the basis of current and historical patterns of drainage connectivity and the composition of the fish fauna, and therefore represent regions with a distinctive evolutionary history and a more or less characteristic biota at the species and higher taxonomic levels [72, 73]. To assess the relative fits of alternative models of faunal assemblage in the Congo Basin, three variants of the DEC\* model were fit to the data in the *BioGeoBEARS* R package [74], following the parameterization of dispersal multipliers from Day et al. [35]: **M0**, an unconstrained multiplier matrix allowing for dispersal to and from the Congo Basin; **M1**, an asymmetric multiplier matrix allowing only dispersal out of the Congo Basin (CB-as-source); **M2**, an asymmetric multiplier matrix allowing only dispersal into the Congo Basin (CB-as-sink). Tip-state ranges were assigned based on the presence of species in different ichthyofaunal provinces. In several cases, species spanned multiple provinces. The maximum range size was set to widespread (all six ichthyofaunal provinces). Given the high dimensionality of the transition matrix resulting from the combination of different provinces (areas) into ranges of sizes up to six, relative to the size of the dataset, 14 disjunct ranges of differing sizes were pruned from analysis, reducing the dimensionality of the matrix from 64 x 64 to 50 x 50. To assess the stability of numerical optimization, analysis was run five times from fresh R sessions. Model fits of the M0, M1, and M2 variants were compared using the Akaike information criterion [75] and supports were assessed using Akaike weights [76]. In an effort to take account of chronological uncertainty due to alternative molecular clock calibration scenarios, inference of geographic range evolution in *Distichodus* was conducted on three of the 15 time-scaled phylogenies previously inferred with BEAST2, namely the chronograms resulting from analyses based on each alternative calibration node constrained as crown and by a relatively moderate soft maximum bound ( $P_{95}$  SMB=30 Ma) (analyses 5, 8, and 14 in Table 4).

## Abbreviations

Ambrose Monell Cryo Collection (AMCC), American Museum of Natural History (AMNH), Bootstrap (BS), Congo Basin (CB), Cornell University Museum of Vertebrates (CUMV), Dispersal-extinction-cladogenesis (DEC), East Coast (EC), Effective Sample Size (ESS), Highest Posterior Density (HPD), Lower Guinea (LG),

Maximum Likelihood (ML), Markov chain Monte Carlo (MCMC), Maximum clade credibility (MCC), Nilo-Sudan (NS), Posterior probabilities (PP), Royal Museum for Central Africa (MRAC), Schwarz/Bayesian Information Criterion (BIC), South African Institute for Aquatic Biodiversity (SAIAB), Uncorrelated lognormal (UCLN), Upper Guinea (UG), Zambezi (Z).

## Declarations

### Ethics Approval and Consent to Participate

This research was conducted under the American Museum of Natural History (AMNH) Institutional Animal Care and Use Committee (IACUC) approval #36/06. Fishes were collected and euthanized prior to preservation in accordance with established guidelines for the use of fishes in research. Stress and suffering were ameliorated by minimizing handling and through the use of anesthetics prior to euthanasia. Voucher specimens examined in this study were loaned and used with permission from the loaning museums/institutions.

### Consent for publication

Not applicable.

### Availability of data and materials

The DNA sequence data supporting the results of this article are available in the GenBank<sup>®</sup> repository (<http://www.ncbi.nlm.nih.gov>) under accession numbers MT300523-MT301561 (see Table 3). Voucher specimens are deposited and readily available in their respective ichthyology collections.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This research was financially supported by AMNH Axelrod Postdoctoral Fellowships to JA and JSSD. Funding for supplies and field collections was provided by The Axelrod Research Curatorship. Our thanks also to Ms. Janine Luke for contributing supplementary funding through a generous donation. These

funding sources played no role in the design of the study, the collection, analysis, and/or interpretation of data, and the writing of the manuscript.

## Authors' Contributions

JA and MLJS conceived and designed the study. MLJS and JA collected voucher specimens in the field. MLJS conducted morphological study and taxonomic determination. JA and JSSD generated, processed, and analyzed comparative molecular data. JA drafted the initial version of the manuscript. All authors read, edited, enhanced, and approved the final version of the manuscript.

## Acknowledgements

We thank our colleagues at the Cornell University Museum of Vertebrates (CUMV), the Royal Museum for Central Africa (MRAC), and the South African Institute for Aquatic Biodiversity (SAIAB) for gifts of tissues and access to specimens under their care. Thanks also to our colleagues and collaborators in Africa, and the staff of the AMNH Ichthyology Department, for their assistance in the lab and in the field.

## Authors' information

Not applicable.

## References

1. Shinkafi BA, Salim AM, Yusuf MA. Some Aspects of the Biology of *Distichodus rostratus* (Günther, 1864) in River Rima, North-western Nigeria. *Greener Journal of Biological Sciences*. 2013;3:136–45.
2. Njoku DC, Keke IR, Allison ME. Food and feeding habits of *Distichodus engycephalus* (Boulenger, 1907) of lake oguta: search for a sustainable aquaculture in the New Millennium. *Zoologist (The)*. 2009;7. doi:10.4314/tzool.v7i1.52085.
3. Zebe VM, Micha JC, Moreau J, Bekeli MN. Age and growth of *Distichodus antonii* (Schilthuis, 1891) (Pisces, Teleostei, Distichontidae) in Pool Malebo, Congo River. *Afr J Env Sci & Tec*. 2010;4. <https://www.ajol.info/index.php/ajest/article/view/56360>. Accessed 28 Jan 2019.
4. Dietoa MY, Kone WM, Yao JK, Costa SD. Use of plant epidermis for determination of macrophytes consumed by *Distichodus rostratus* Günther, 1864 (Pisces: Distichodontidae), of Taabo artificial lake (Basin of Bandama, Côte d'Ivoire). *Agronomie Africaine*. 2014;26:103-112–112.
5. Lowe-McConnell RH, Lowe-McConnell RH. *Ecological studies in tropical fish communities*. Cambridge University Press; 1987.

6. Béné C, Heck S. Fish and food security in Africa. *Naga, Worldfish Center Quarterly*. 2005;28:8–13.
7. Fricke R, Eschmeyer W, Van der Laan R. Eschmeyer's Catalog Of Fishes: Genera, Species, References. Electronic version. 2019. <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>. Accessed 27 May 2019.
8. Mamonekene V, Vreven E. *Distichodus teugelsi* a new distichodontid from the middle Congo River basin, Africa (Characiformes: Distichodontidae). *Ichthyol Explor Freshw*. 2008;19:97–98.
9. Moelants T, Snoeks J, Vreven E. *Distichodus kasaiensis* and *ingae*, two new distichodontid species (Characiformes: Distichodontidae) from the Congo basin. *Ichthyol Explor Freshw*. 2018;28:177–192.
10. Vari RP. Anatomy, relationships, and classification of the families Citharinidae and Distichodontidae (Pisces, Characoidea). *Bull Br Mus Nat Hist (Zool)*. 1979;36:261–344.
11. Vari R. Distichodontidae. In: Stiassny MLJ, Teugels GG, Hopkins CD, editors. The fresh and brackish water fishes of Lower Guinea, West-Central Africa. Paris (France), Tervuren (Belgium): IRD, MNHN, MRAC; 2007. p. 413–65.
12. Boulenger GA. Catalogue of the fresh-water fishes of Africa in the British Museum (Natural history). 1909;4. <http://agris.fao.org/agris-search/search.do?recordID=XF2015027068>. Accessed 27 May 2019.
13. Moelants T, Zebe VM, Snoeks J, Vreven E. A review of the *Distichodus antonii* assemblage (Characiformes: Distichodontidae) from the Congo basin. *J Nat Hist*. 2014;48:1707–35.
14. Müller J, Troschel FH. *Horae Ichthyologicae: Beschreibung und Abbildung neuer Fische*. Berlin: Veit & Comp.,; 1845. <https://www.biodiversitylibrary.org/item/30200>.
15. Arroyave J, Denton JSS, Stiassny MLJ. Are characiform Fishes Gondwanan in Origin? Insights from a Time-Scaled Molecular Phylogeny of the Citharinoidei (Ostariophysi: Characiformes). *PLoS One*. 2013;8:e77269.
16. Verheyen E, Willems W, Martens K, Borges A, Darchambreau F, Lambert T, et al. Congo basin; From carbon to Fishes. COBAFISH Final Report. Brussels: Belgian Science Policy; 2017. [https://www.friscris.be/en/publications/congo-basin-from-carbon-to-fishes-cobafish-final-report\(64880bba-b66d-4944-9e24-82c1b372d5c2\).html](https://www.friscris.be/en/publications/congo-basin-from-carbon-to-fishes-cobafish-final-report(64880bba-b66d-4944-9e24-82c1b372d5c2).html). Accessed 27 May 2019.
17. Daget J. Note sur les *Distichodus* (Poissons Characiformes) de l'Ouest africain. *Bull IFAN*. 1959;21 A:1276–1303.
18. Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol*. 1994;11:715–24.
19. Yang Z, Nielsen R, Goldman N, Pedersen A-MK. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics*. 2000;155:431–49.
20. Ren F, Tanaka H, Yang Z. An Empirical Examination of the Utility of Codon-Substitution Models in Phylogeny Reconstruction. *Syst Biol*. 2005;54:808–18.

21. Chou J, Gupta A, Yaduvanshi S, Davidson R, Nute M, Mirarab S, et al. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*. 2015;16:S2.
22. Zhou Y, Duvaux L, Ren G, Zhang L, Savolainen O, Liu J. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity*. 2017;118:211–20.
23. Ballard JWO. When One Is Not Enough: Introgression of Mitochondrial DNA in *Drosophila*. *Mol Biol Evol*. 2000;17:1126–30.
24. Cotterill FPD. The evolutionary history and taxonomy of the *Kobus leche* species complex of South-Central Africa in the context of palaeo-drainage dynamics. Thesis. Stellenbosch Universit.; 2006. <https://scholar.sun.ac.za:443/handle/10019.1/21773>. Accessed 28 Jan 2019.
25. Stankiewicz J, de Wit MJ. A proposed drainage evolution model for Central Africa—Did the Congo flow east? *J of Afri Earth Sci*. 2006;44:75–84.
26. Moore AE, Cotterill FP, Main MP, Williams HB. The Zambezi River. *Large rivers: geomorphology and management*. 2007;311–332.
27. Goodier SAM, Cotterill FPD, O’Ryan C, Skelton PH, de Wit MJ. Cryptic diversity of African tigerfish (genus *Hydrocynus*) reveals palaeogeographic signatures of linked neogene geotectonic events. *PLoS One*. 2011;6:e28775.
28. Schwarzer J, Swartz ER, Vreven E, Snoeks J, Cotterill FPD, Misof B, et al. Repeated trans-watershed hybridization among haplochromine cichlids (Cichlidae) was triggered by Neogene landscape evolution *Proc of the Royal Soc of London B: Biol Sci*. 2012;279:4389–98.
29. Day JJ, Peart CR, Brown KJ, Friel JP, Bills R, Moritz T. Continental diversification of an African catfish radiation (Mochokidae: *Synodontis*). *Syst Biol*. 2013;62:351–65.
30. Zakharov EV, Lobo NF, Nowak C, Hellmann JJ. Introgression as a likely cause of mtDNA paraphyly in two allopatric skippers (Lepidoptera: Hesperidae). *Heredity*. 2009;102:590–9.
31. Stewart KM. Fossil fish remains from Mio-Pliocene deposits at Lothagam, Kenya. 2003.
32. Argyriou T, Cook TD, Muftah AM, Pavlakis P, Boaz NT, Murray AM. A fish assemblage from an early Miocene horizon from Jabal Zaltan, Libya. *J of Afr Ear Sci*. 2015;102:86–101.
33. Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, Holroyd PA, et al. Best Practices for Justifying Fossil Calibrations. *Syst Biol*. 2012;61:346–59.
34. Alter SE, Brown B, Stiassny MLJ. Molecular phylogenetics reveals convergent evolution in lower Congo River spiny eels. *BMC Evo Bio*. 2015;15:224.
35. Day JJ, Fages A, Brown KJ, Vreven EJ, Stiassny MLJ, Bills R, et al. Multiple independent colonizations into the Congo Basin during the continental radiation of African *Mastacembelus* spiny eels. *J of Biogeo*. 2017;44:2308–18.
36. Pinton A, Agnèsè J-F, Paugy D, Otero O. A large-scale phylogeny of *Synodontis* (Mochokidae, Siluriformes) reveals the influence of geological events on continental diversity during the Cenozoic.

- Mol Phyl Evo. 2013;66:1027–40.
37. Daniels SR, Phiri EE, Klaus S, Albrecht C, Cumberlidge N. Multilocus Phylogeny of the Afrotropical Freshwater Crab Fauna Reveals Historical Drainage Connectivity and Transoceanic Dispersal Since the Eocene. *Syst Biol.* 2015;64:549–67.
  38. Flower BP, Kennett JP. The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. *Palaeogeography, Palaeoclimatology, Palaeoecology.* 1994;108:537–55.
  39. Lavier LL, Steckler MS, Brigaud F. Climatic and tectonic control on the Cenozoic evolution of the West African margin. *Marine Geology.* 2001;178:63–80.
  40. Goudie AS. The drainage of Africa since the Cretaceous. *Geomorphology.* 2005;67:437–56.
  41. Livingstone DA, Rowland M, Bailey PE. On the Size of African Riverine Fish Faunas. *American Zoologist.* 1982;22:361–9.
  42. Lundberg JG. The temporal context for the diversification of Neotropical fishes. In: Malabarba LR, Reis RE, Vari RP, Lucena CAS de, Lucena Z de, editors. *Phylogeny and classification of Neotropical fishes.* EDIPUCRS; 1998. p. 49–68.
  43. Montoya-Burgos JI. Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. *Mol Ecol.* 2003;12:1855–67.
  44. Cracraft J. Species Concepts and Speciation Analysis. In: Johnston RF, editor. *Current Ornithology.* Boston, MA: Springer US; 1983. p. 159–87. doi:10.1007/978-1-4615-6781-3\_6.
  45. Donoghue MJ. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist.* 1985;88:172–181.
  46. Nickum J, Bart Jr HL, Bowser PR, Greer IE, Hubbs C, Jenkins JA, et al. Guidelines for the use of fishes in research. *FISHERIES-BETHESDA.* 2004;29:26–26.
  47. Arroyave J, Stiassny MLJ. Phylogenetic relationships and the temporal context for the diversification of African characins of the family Alestidae (Ostariophysi: Characiformes): Evidence from DNA sequence data. *Mol Phyl Evo.* 2011;60:385–97.
  48. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Rozen S, Skaletsky H, editors. *Bioinformatics methods and protocols.* Springer Verlag; 2000. p. 365–386.
  49. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28:1647–9.
  50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
  51. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:113.

52. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol.* 2017;34:772–3.
53. Schwarz G. Estimating the Dimension of a Model. *Ann Statist.* 1978;6:461–4.
54. Kluge AG. A Concern for Evidence and a Phylogenetic Hypothesis of Relationships among *Epicrates* (Boidae, Serpentes). *Syst Biol.* 1989;38:7–25.
55. Nixon KC, Carpenter JM. On Simultaneous Analysis. *Cladistics.* 1996;12:221–41.
56. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
57. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). 2010. p. 1–8.
58. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol.* 2008;57:758–71.
59. Chifman J, Kubatko L. Quartet inference from SNP data under the coalescent model. *Bioinformatics.* 2014;30:3317–24.
60. Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;19:153.
61. Swafford DL. PAUP\*. Phylogenetic analysis using parsimony (\* and other methods). Sunderland, MA: Sinauer Associates; 2003.
62. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics.* 2010;26:1569–71.
63. Gil M, Zanetti MS, Zoller S, Anisimova M. CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. *Mol Biol Evol.* 2013;30:1270–80.
64. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comp Biol.* 2019;15:e1006650.
65. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed Phylogenetics and Dating with Confidence. *PLoS Biol.* 2006;4:e88.
66. Heath TA. Divergence Time Estimation using BEAST v2.0 (Tutorial). In: Workshop on Applied Phylogenetics. Bodega Bay, California; 2015. p. 1–44. <http://treethinkers.org/divergence-time-estimation-using-beast/>.
67. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol.* 2018;67:901–4.
68. Warren DL, Geneva AJ, Lanfear R. RWTY (R We There Yet): An R Package for Examining Convergence of Bayesian Phylogenetic Analyses. *Mol Biol Evol.* 2017;34:1016–20.
69. Massana KA, Beaulieu JM, Matzke NJ, O’Meara BC. Non-null Effects of the Null Range in Biogeographic Models: Exploring Parameter Estimation in the DEC Model. *bioRxiv.* 2015;:026914.

70. Ree RH, Moore BR, Webb CO, Donoghue MJ. A Likelihood Framework for Inferring the Evolution of Geographic Range on Phylogenetic Trees. *Evolution*. 2005;59:2299–311.
71. Ree RH, Smith SA. Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Syst Biol*. 2008;57:4–14.
72. Roberts TR. Geographical distribution of African freshwater fishes. *Zool J Linna Soc*. 1975;57:249–319.
73. Lévêque C, Oberdorff T, Paugy D, Stiassny MLJ, Tedesco PA. Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*. 2008;595:545–67.
74. Matzke NJ. BioGeoBEARS: biogeography with Bayesian (and likelihood) evolutionary analysis in R scripts. R package, version 02. 2013;1:2013.
75. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19:716–23.
76. Akaike H. On the Likelihood of a Time Series Model *J of the Royal Stat Soc Ser D (The Statistician)*. 1978;27 3/4:217–35.
77. Food and Agriculture Organization of the United Nations (FAO). Hydrological basins in Africa (Derived from HydroSHEDS). 2000. <http://www.fao.org/geonetwork/srv/en/metadata.show?id=296&currTab=simple>.

## Tables

Due to technical limitations, tables are only available as a download in the supplemental files section

## Supplementary File Information

**Fig. 1S. *enc1 Distichodus* phylogeny as inferred by likelihood in RAxML.** Colored circles on nodes indicate degree of clade support as determined by bootstrap values (BS). The identity of leaves (terminals) not printed on the tree is specified by the species name (in bold) at the base of the most recent labeled ancestral node from which the sample descends. Names in bold black correspond to those species resolved as monophyletic (when multiple individuals were available), whereas those in bold green indicate that, while most of the sampled specimens fall into the clade subtended by that node, some samples fall outside the clade, and therefore the species is not resolved as monophyletic. Outgroup taxon (*Paradistichodus dimiatus*) not shown.

**Fig. 2S. *glyt Distichodus* phylogeny as inferred by likelihood in RAxML.** Same contextual information as in Fig. 1S.



**Fig. 3S.** *myh6* *Distichodus* phylogeny as inferred by likelihood in RAxML. Same contextual information as in Fig. 1S.

**Fig. 4S.** *sh3px3* *Distichodus* phylogeny as inferred by likelihood in RAxML. Same contextual information as in Fig. 1S.

**Fig. 5S.** mtDNA (*co1*, *cr*, *cytb*, *nd2*) *Distichodus* phylogeny as inferred by likelihood in RAxML. Same contextual information as in Fig. 1S.

**Fig. 6S.** A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 1. Same contextual information as in Fig. 6.

**Fig. 7S.** A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 2. Same contextual information as in Fig. 6.

**Fig. 8S.** A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 3. Same contextual information as in Fig. 6.

**Fig. 9S.** A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 4. Same contextual information as in Fig. 6.

**Fig. 10S.** A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 5. Same contextual information as in Fig. 6.

**Fig. 11S.** A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 6. Same contextual information as in Fig. 6.

**Fig. 12S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 7. Same contextual information as in Fig. 6.

**Fig. 13S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 9. Same contextual information as in Fig. 6.

**Fig. 14S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 10. Same contextual information as in Fig. 6.

**Fig. 15S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 11. Same contextual information as in Fig. 6.

**Fig. 16S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 12. Same contextual information as in Fig. 6.

**Fig. 17S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 13. Same contextual information as in Fig. 6.

**Fig. 18S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 14. Same contextual information as in Fig. 6.

**Fig. 19S. A time-scaled phylogeny of *Distichodus*.** Chronogram resulting from BEAST2 analysis 15. Same contextual information as in Fig. 6.

**Fig. 20S. A spatiotemporal reconstruction of *Distichodus* range evolution.** Based on the optimal DEC\* model (M1; CB-as-source) and input chronogram resultant from BEAST2 analysis 5. Ichthyofaunal provinces color-coded and abbreviated as in Fig. 1. Probabilities of ancestral areas at each node are presented in Table 2S.

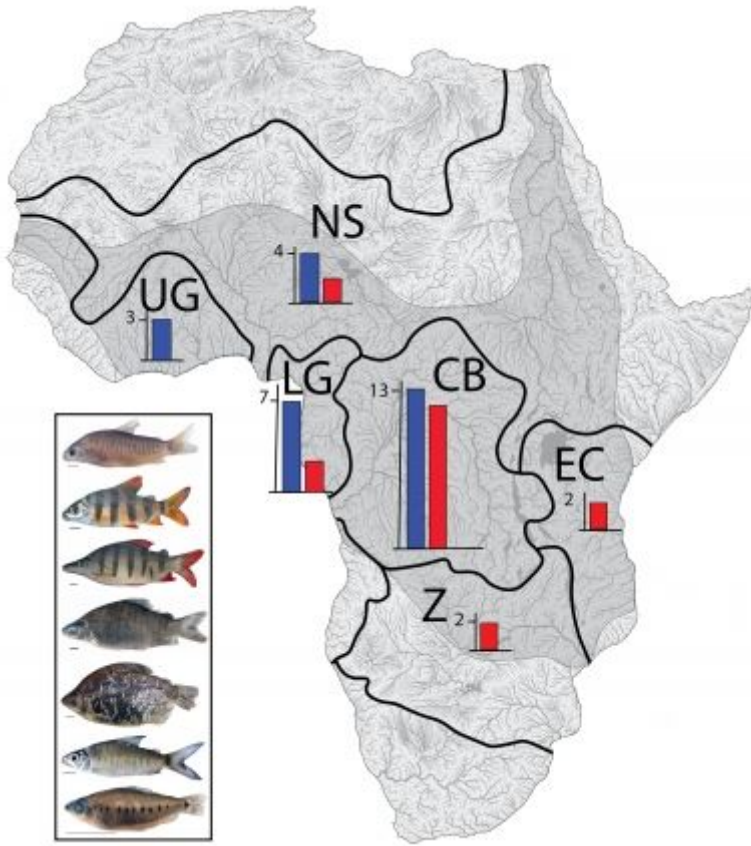
**Fig. 21S. A spatiotemporal reconstruction of *Distichodus* range evolution.** Based on the optimal DEC\* model (M1; CB-as-source) and input chronogram resultant from BEAST2 analysis 14. Ichthyofaunal provinces color-coded and abbreviated as in Fig. 1. Probabilities of ancestral areas at each node are presented in Table 3S.

**Table 1S. Probabilities of ancestral states/ranges at each node of the spatiotemporal reconstruction of *Distichodus* range evolution presented in Fig. 7.** Columns indicate ancestral areas, represented by all unique combinations for all possible group sizes for the six ichthyofaunal provinces. Rows indicate nodes, with numbering following the typical R phylo format, i.e., 1 is the first tip taxon/area, beginning at the bottom. After the last tip value, the numbering begins at the root, and moves tipward. Ichthyofaunal provinces abbreviated as in Fig. 1.

**Table 2S. Probabilities of ancestral states/ranges at each node of the spatiotemporal reconstruction of *Distichodus* range evolution presented in Fig. 20S.** Same contextual information as in Table 1S.

**Table 3S. Probabilities of ancestral states/ranges at each node of the spatiotemporal reconstruction of *Distichodus* range evolution presented in Fig. 21S.** Same contextual information as in Table 1S.

## Figures



**Figure 1**

Geographic distribution and variation in external morphology of *Distichodus* species diversity. Map of Africa divided into ichthyofaunal provinces (originally defined by Roberts [72], modified by Lévêque [73], and redrawn according to new hydrological basin mapping published by FAO [77]): Congo Basin (CB), East Africa (EA), Nilo-Sudan (NS), Lower Guinea (LG), South Africa (SA), and West Africa (WA). Shaded area represents *Distichodus* extent of occurrence. Inset bar charts indicate number of *Distichodus* species present in each ichthyofaunal province: endemic (red) and total (blue) (when more than endemics). Inset frame fish photographs illustrate the extent of variation in body shape, size, and coloration in *Distichodus* species (from top to bottom: *D. hypostomatus*, *D. sexfasciatus*, *D. lussoso*, *D. antonii*, *D. affinis*, *D. shenga*, *D. decemmaculatus*).

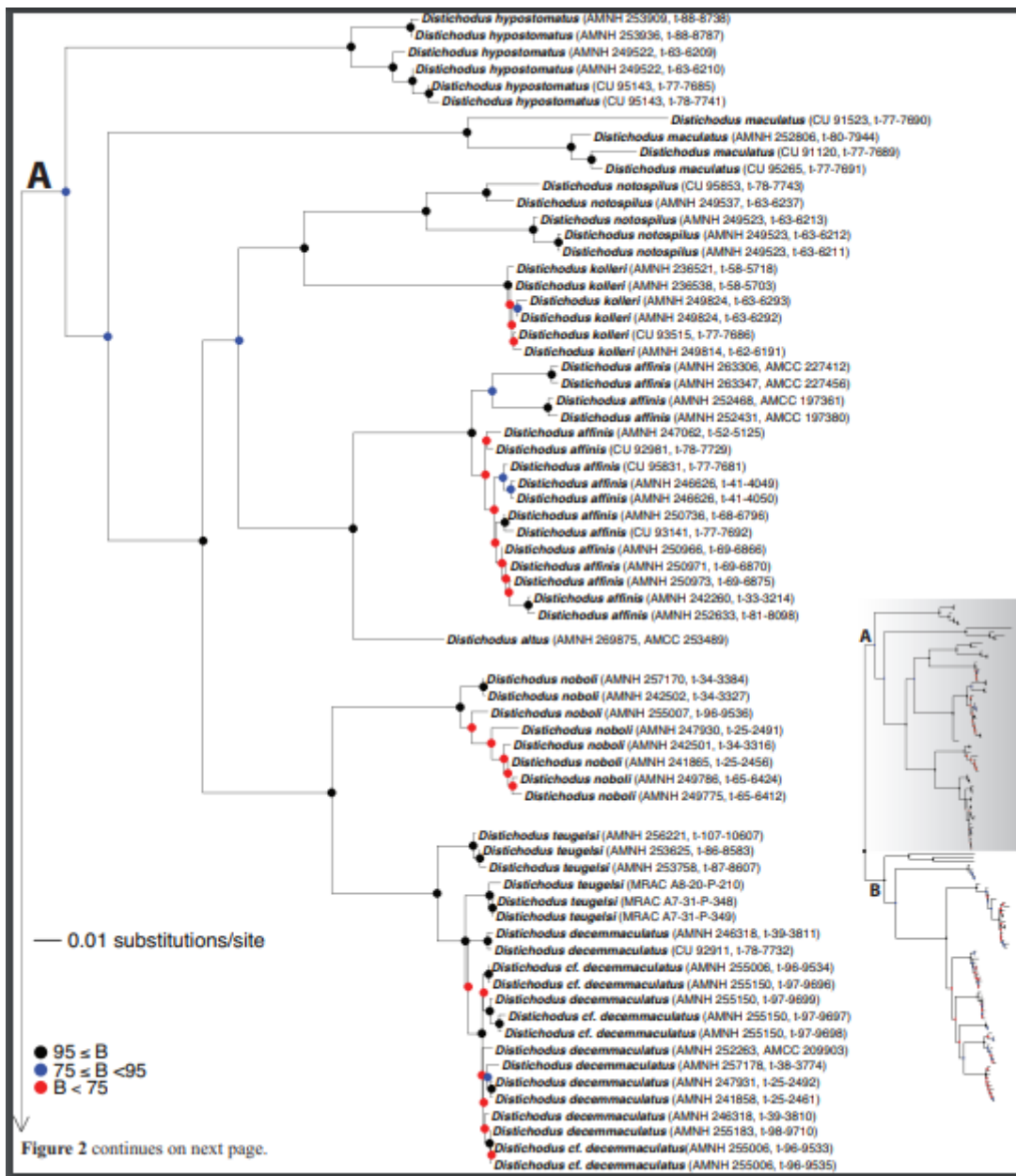


Figure 2

Total-evidence *Distichodus* phylogeny as inferred by likelihood in RAxML. Colored circles on nodes indicate degree of clade support as determined by bootstrap values (BS). Nodes labeled A and B represent the two main infrageneric clades. Outgroup taxon (*Paradistichodus dimiatus*) not shown.

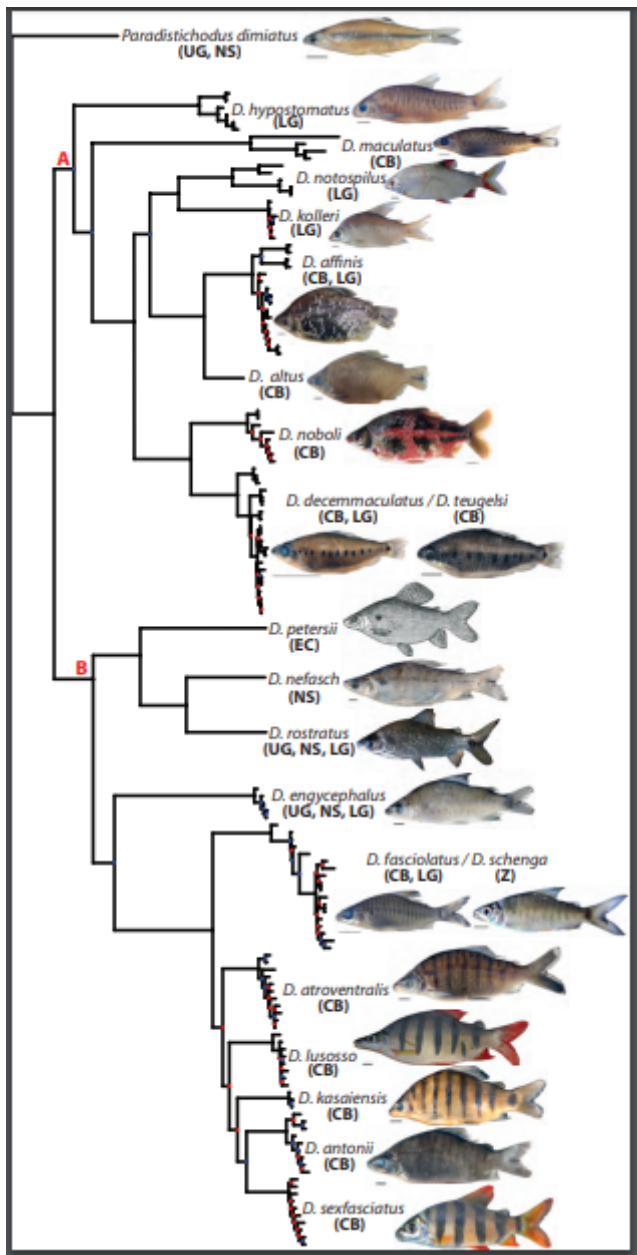


Figure 3

Summary tree of the RAxML *Distichodus* phylogeny highlighting interspecific relationships.

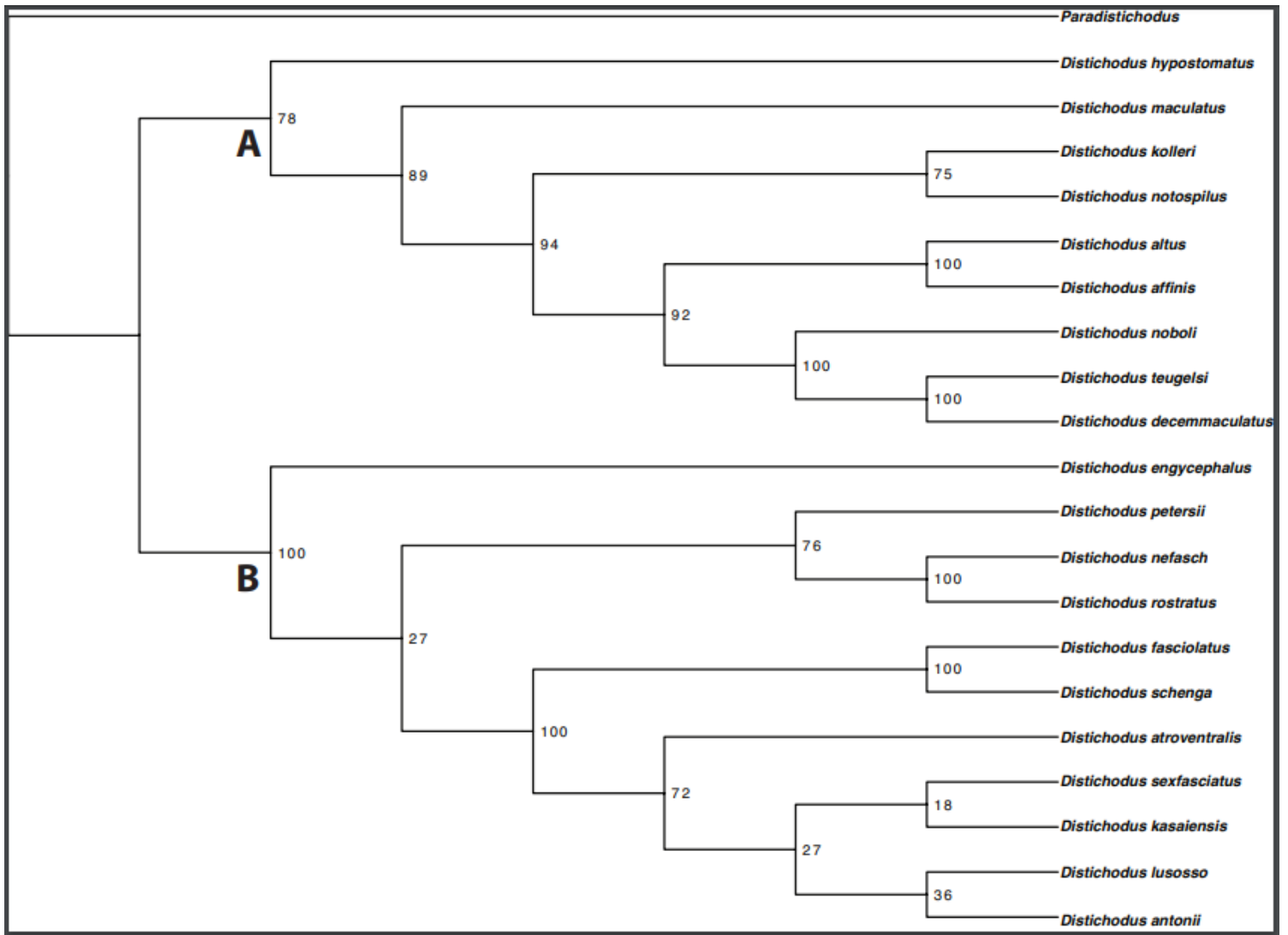


Figure 4

Distichodus species tree generated using the coalescence-based method SVDquartets.

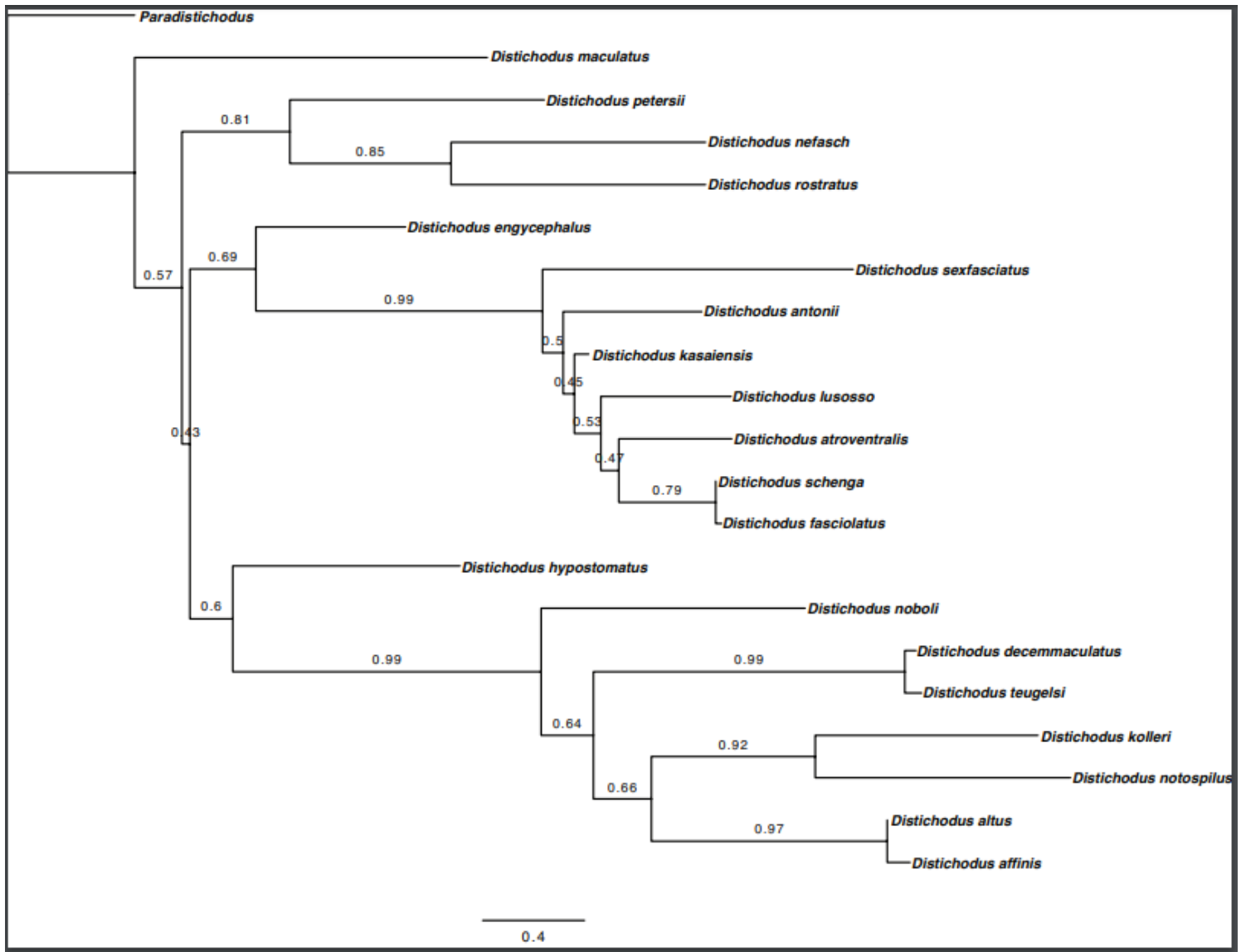
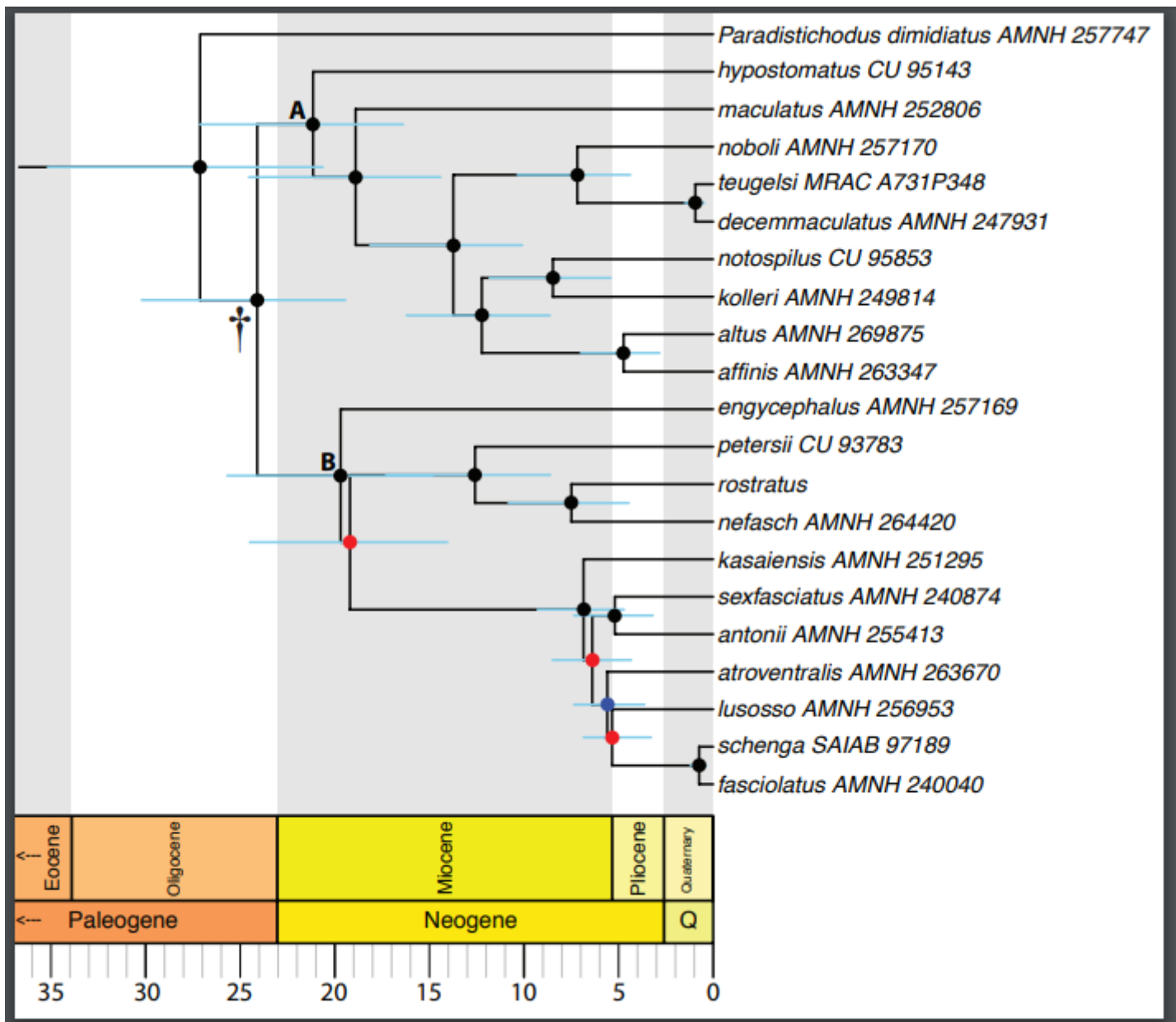


Figure 5

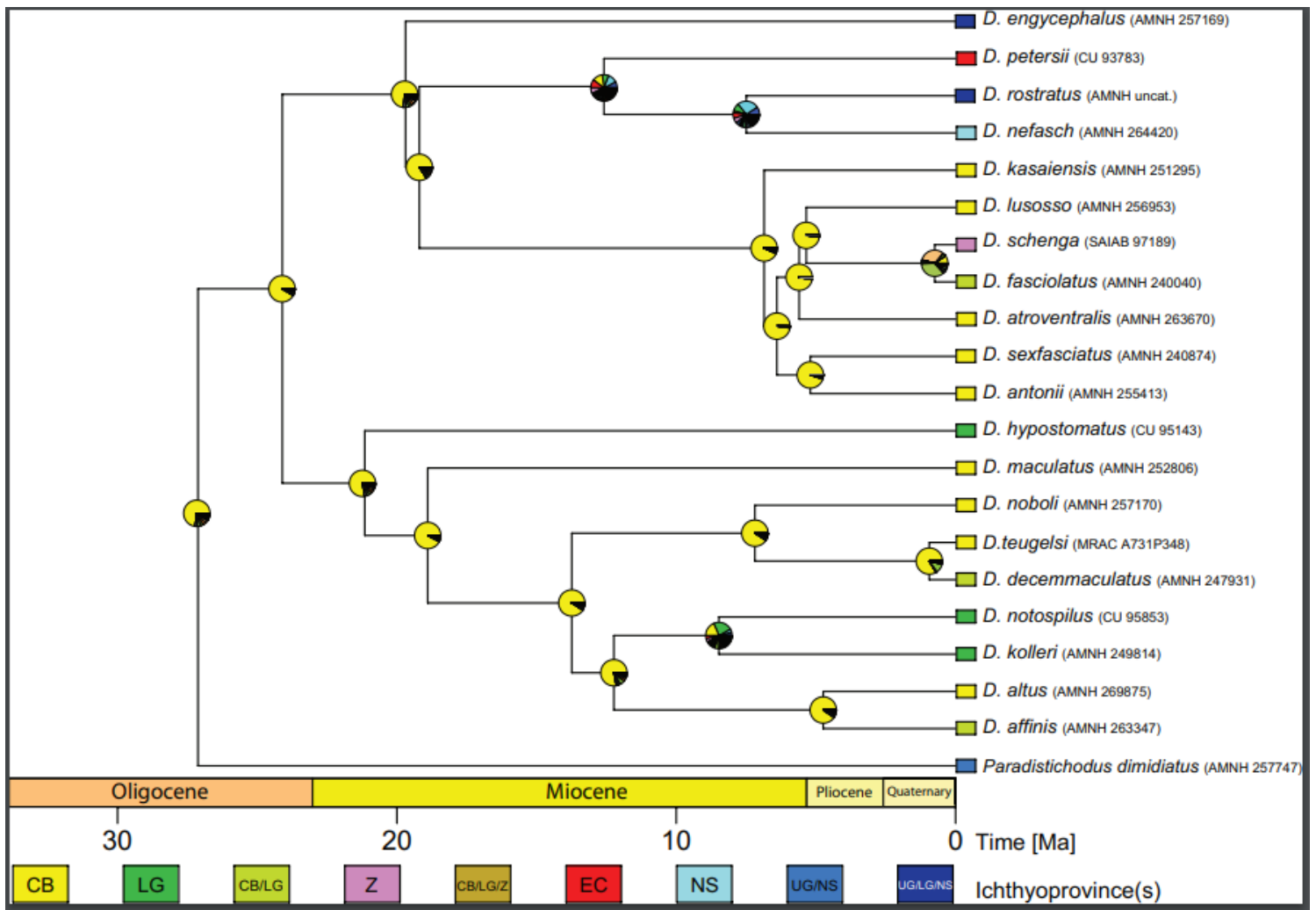
Distichodus species tree generated using the coalescence-based method ASTRAL-III.





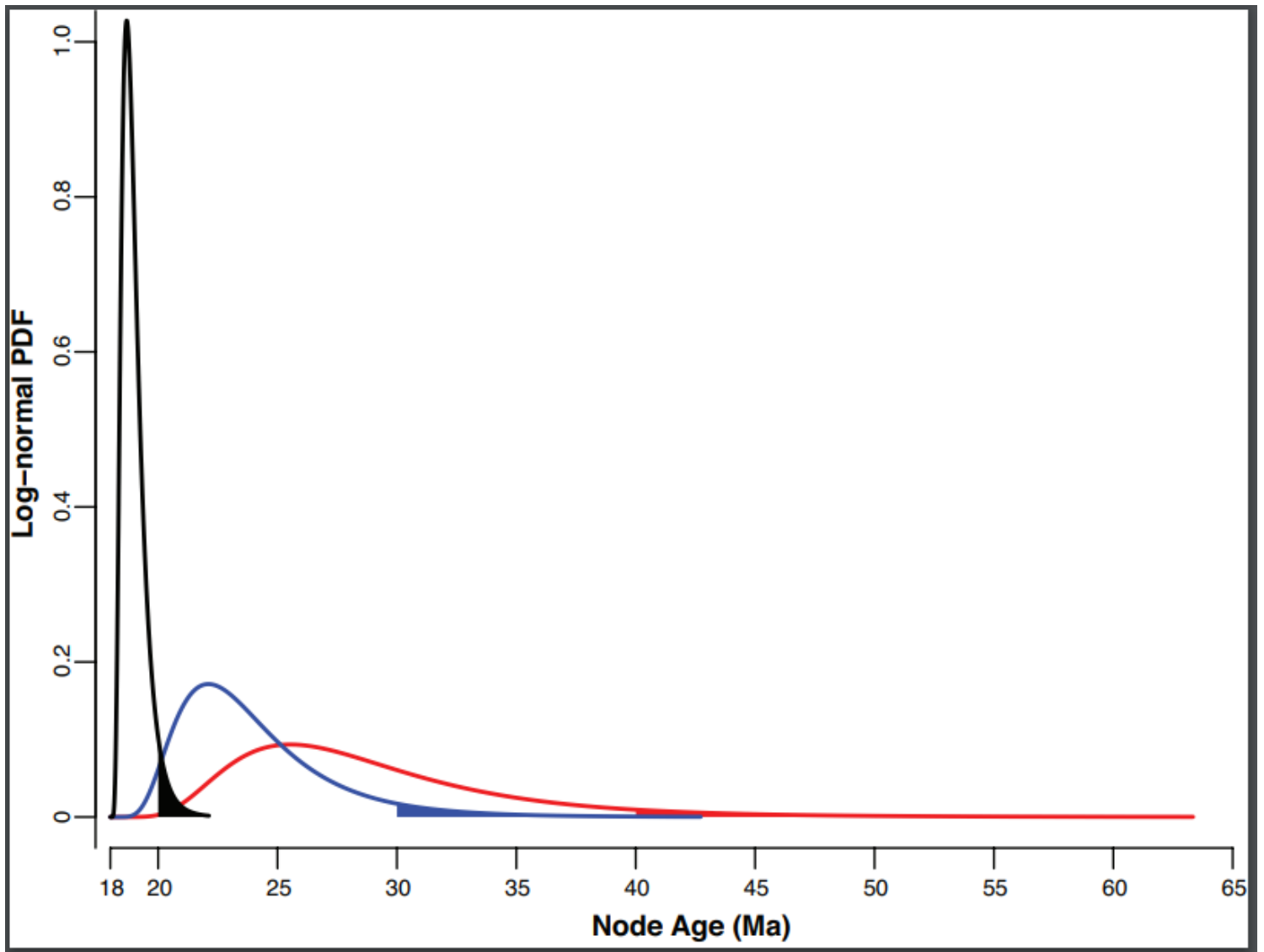
**Figure 6**

A time-scaled phylogeny of *Distichodus*. Chronogram resulting from BEAST2 analysis 8, intermediate in terms of calibration node (D, crown) and P95 SMB (30 Ma). Divergence-time estimates are represented by the mean ages of clades. Light red bars correspond to 95% highest posterior density (HPD) intervals of mean node ages. Calibration (fossil-based) node indicated by a dagger (†). Colored circles on nodes indicate degree of clade support as determined by posterior probabilities: black > 0.95, 0.95  $\square$  blue  $\square$  0.75, red < 0.75. Outgroup taxon (*Nannocharax ansorgii*) not shown.



**Figure 7**

A spatiotemporal reconstruction of *Distichodus* range evolution. Based on the optimal DEC\* model (M1; CB-as-source) and input chronogram resultant from BEAST2 analysis 8 (Fig. 7). Ichthyofaunal provinces color-coded and abbreviated as in Fig. 1. Probabilities of ancestral areas at each node are presented in Table 1S.



**Figure 8**

Alternative log-normally distributed priors used to account for temporal uncertainty of calibration nodes. Each prior probability density function (PDF) is characterized by a hard minimum bound of 18 Ma (set by the age of the calibration fossil), a standard deviation ( $\sigma$ ) of 0.5, and a variable mean ( $\mu$ ) (in real space) that probabilistically models the extent to which the node age spreads into the past:  $\mu=19$  (black),  $\mu=24$  (blue), and  $\mu=29$  (red). The lower limit of the x-axis interval defining the area shaded under each curve corresponds to its 95th percentile soft maximum bound (P95 SMB): 20 Ma (black), 30 Ma (blue), and 40 Ma (red).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig15S.pdf](#)
- [Fig8S.pdf](#)

- Fig14S.pdf
- Fig13S.pdf
- Fig12S.pdf
- Fig11S.pdf
- Fig18S.pdf
- Fig21S.pdf
- Fig20S.pdf
- Fig19S.pdf
- Fig17S.pdf
- Fig7S.pdf
- Fig6S.pdf
- Fig9S.pdf
- Fig1S.pdf
- Fig10S.pdf
- Table3.xlsx
- Table2.xlsx
- Fig2S.pdf
- Fig3S.pdf
- Fig4S.pdf
- Fig5S.pdf
- Table4.xlsx
- Table1.xlsx
- Table1S.xls
- Table3S.xls
- Table2S.xls
- Fig16S.pdf