Neotropical Ichthyology

Original article

https://doi.org/10.1590/1982-0224-2026-0014

New insights into the genetic diversity of *Microsternarchus* (Gymnotiformes: Hypopomidae) in the rio Negro using genomic markers

Correspondence:
Matthew A. Burnett
matthew.burnett@gwmail.gwu.edu

Matthew A. Burnett[1], Guillermo Ortí[1,2] and José Alves-Gomes[3]

Gymnotiformes comprise an order of fishes distributed throughout South and Central America, characterized by their electric organ discharges and electroreceptive capabilities. There are currently 281 valid species, but despite this high level of specific diversity, there is robust evidence for a wealth of undescribed diversity, especially within the genus *Microsternarchus*, which currently contains five species. Previous molecular work has revealed at least six genetically distinct lineages within the genus *Microsternarchus*, although this only relied on COI data. Here we infer the first phylogeny of *Microsternarchus* using double digest restriction enzyme associated DNA (ddRAD) markers. These multi-locus markers have shown to be useful in resolving phylogenetic relationships as they provide data across the whole genome and can be used to detect gene flow and introgression. Congruence between the ddRAD and COI phylogenies lends support to the hypothesis of undescribed diversity and lays the groundwork for future investigation into this group.

Keywords: COI, ddRAD, Delimitation, Electric fishes, Mito-nuclear discordance.

Submitted February 3, 2026
Accepted August 21, 2026
Epub ???, 2026

Associate Editor Priscila Camelier
Section Editor Izeni Farias
Editor-in-chief José Birindelli

1 Department of Biological Science, The George Washington University, 2029 G St. NW, 20052, Washington D.C., USA. (MAB) matthew.burnett@gwmail.gwu.edu (corresponding author), (GO) gorti@gwu.edu.

2 Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th St. & Constitution Ave. NW, 20560, Washington D.C., USA.

3 Laboratório de Fisiologia Comportamental e Evolução, Instituto Nacional de Pesquisas da Amazônia, Av. André Araújo, 2936, 69067-375, Manaus, AM, Brazil. (JAG) puraque@inpa.gov.br.

Online version ISSN 1982-0224
Print version ISSN 1679-6225

Neotrop. Ichthyol.
vol. 24, no. 4, 2026

Gymnotiformes é uma ordem de peixes distribuída por toda a América do Sul e Central, caracterizada pela capacidade eletrorreceptora e de gerar descargas de órgãos elétricos. Atualmente, existem 275 espécies válidas, mas apesar desse alto nível de diversidade específica, há fortes evidências de uma alta diversidade ainda não descrita, especialmente dentro do gênero *Microsternarchus*, que atualmente contém cinco espécies. Trabalhos moleculares anteriores revelaram pelo menos seis linhagens geneticamente distintas dentro do gênero *Microsternarchus*, ainda que estes estudos tenham se fundamentado somente em dados do gene COI. Neste trabalho, propomos a primeira filogenia de *Microsternarchus* utilizando marcadores associados à dupla digestão de DNA por enzimas de restrição (ddRAD). Esses marcadores "multi-locus" têm se mostrado úteis na resolução de relações filogenéticas, uma vez que fornecem dados de todo o genoma e podem ser usados para detectar fluxo gênico e introgressão. A congruência entre as filogenias obtidas com ddRAD e COI corrobora a hipótese de diversidade não descrita e estabelece as bases para futuras investigações sobre esse grupo.

Palavras-chave: COI, ddRAD, Delimitação, Discordância mito-nuclear, Peixes elétricos.

INTRODUCTION

The Amazonas basin, located in Northern South America, spans much of Brazil, French Guiana, Suriname, Guyana, Venezuela, Colombia, Ecuador, Peru, and Bolivia, and is known for its exceptional biodiversity. The first ichthyological survey of the region, conducted by Louis Agassiz in 1865, identified approximately 350 distinct fish species. Over the last 160 years, ongoing sampling efforts have dramatically increased the number of known species, with the total rising to an estimated 2,500 as of 2021 (https://www.amazon-fish.com/datavisualization/sampling-efforts-over-years). Increased exploration and better technology have aided in the efforts to catalogue this previously unknown or under-reported diversity. Among many others, a case study that follows this pattern is the genus *Microsternarchus* Fernández-Yépez, 1968.

Microsternarchus was originally described as a monotypic genus in the family Hypopomidae, with *Microsternarchus bilineatus* Fernández-Yépez, 1968 as the type species. While the type locality of *Microsternarchus* was given as the rio San José in the Orinoco basin, specimens of the genus have since been collected throughout the Orinoco, Amazonas, and Tocantins basins (Cox Fernandes, Williston, 2017). Following the original description, almost no systematic work was done on this poorly known genus of small gymnotiform fishes except for the validation of *M. bilineatus* by Mago-Leccia (1994) and the collection of molecular data for *M. bilineatus* to be included in an ordinal-level phylogeny of Gymnotiformes (Alves-Gomes *et al*., 1995). In 2012, the first taxonomic study focusing on this genus was conducted utilizing the mitochondrial gene cytochrome oxidase I (COI) (Maia, Alves-Gomes, 2012). Analyzing 415 specimens collected throughout the rio Negro (Amazonas basin), the largest blackwater river in the world, and one of the main tributaries of the rio Amazonas, the authors discovered

unexpected variation that they characterized as five genetically distinct lineages, initially labeled as lineages A to E. Further, significant genetic diversity was noted within lineages C, D, and E which were sub-divided into four, two, and two sub-lineages, respectively.

Specimens bearing COI lineage A have since been described as a new genus and species, *Procerusternarchus pixuna* Cox Fernandes, Nogueira & Alves-Gomes, 2014 and thus are excluded from analyses here, while the E-2 sub-lineage became the second described species within *Microsternarchus*, *M. brevis* Cox Fernandes, Nogueira, Williston & Alves-Gomes, 2015. A recently described species, *M. javieri* Pinilla, Cox Fernandes & Alves-Gomes, 2025, included individuals of Lineage C from Maia, Alves-Gomes (2012). Around the same time, two additional species, *M. longicaudatus* Sousa, Wosiacki, Muriel-Cunha, Prudente, Sousa & Peixoto, 2024 and *M. schonmanni* Cox Fernandes, Keeffe & Pinilla, 2024, were described from the upper Madeira basin and lower Amazonas basin of Pará, respectively, based solely on morphological characteristics. The initial COI survey by Maia, Alves-Gomes (2012) suggests that several more species of *Microsternarchus* may exist in the rio Negro basin alone and in the Amazonas basin at large.

Before the study by Maia, Alves-Gomes (2012), the diversity of *Microsternarchus* remained largely unrecognized. Much of this may be attributed to limited phenotypic scrutiny, which masked the cryptic lineages subsequently revealed by COI. This perceived phenotypic uniformity, combined with a lack of comprehensive taxonomic assessments across the genus' broad distribution, has hindered its taxonomic resolution (Fernández-Yépez, 1968; Cox Fernandes *et al*., 2015, 2024; Sousa *et al*., 2024; Pinilla *et al*., 2025).

In all recent descriptions of new *Microsternarchus* species there is no single diagnostic character in the external morphology that differentiates them. Instead, studies have relied primarily on a combination of anal fin rays, coloration, and length of caudal filament (Cox Fernandes *et al*., 2015, 2024; Cox Fernandes, Williston, 2017; Sousa *et al*., 2024; Pinilla *et al*., 2025). The combination of these characters has been sufficient to identify different species in this genus. Furthermore, discriminating between members of the identified mitochondrial lineages by eye is notoriously difficult, particularly across different ontogenetic stages (JAG, 2026, pers. obs.).

Additionally, the original type series of *M. bilineatus* has been lost, and redescriptions by Mago-Leccia (1994) and Cox Fernandes, Williston (2017), including designation of a neotype, have relied exclusively on morphology. While specimens for the redescription by Cox-Fernandes were collected from Río San Bartolo, roughly 110 km from the type locality, there are no DNA data for *M. bilineatus* from the type locality itself, and thus the identity of the "true" *M. bilineatus* remains elusive until detailed morphological study of all identified lineages can be completed.

This history underscores the limitations of using morphological data alone, which is why molecular tools have become increasingly central in understanding diversity, especially in species complexes and/or sibling species (Bickford *et al*., 2007; Schlick-Steiner *et al*., 2010; Crampton *et al*., 2016; Struck *et al*., 2018), which appears to be the case in *Microsternarchus*. The mitochondrial COI gene is widely considered "the DNA barcode" for vertebrates and other organisms and has been successfully used to quickly gain insights of genetic diversity and reveal putative cryptic species (Herbert *et al*., 2003). As evolutionary biology moves towards genome-scale datasets, however,

phylogenies generated from mitochondrial and nuclear genes have often been shown to be discordant.

Mitonuclear discordance poses significant challenges for species delimitation. While mitochondrial DNA is frequently utilized in systematics, its utility can be compromised due to processes such as incomplete lineage sorting, introgressive hybridization and sex-biased dispersal rates (Toews, Brelsford, 2012). Mitochondrial DNA is expected to track speciation events more rapidly due to the smaller effective population size of the mitochondrial genome compared to the nuclear genome. However, systems which have undergone rapid radiation may retain ancestral polymorphisms, leading to gene tree – species tree discordance (Funk, Omland, 2003). Furthermore, mitochondrial capture, a process through which the mitochondrial genome of one species is completely replaced by that of another through hybridization and backcrossing, can lead to polyphyletic or paraphyletic mitochondrial patterns that obscure stable nuclear boundaries (Harrison, Larson, 2014).

Since no study has looked for evidence of mito-nuclear discordance within *Microsternarchus,* this study utilizes genome-wide ddRADseq data (which have been used with success in previous studies (Hughes *et al*., 2020; Englmaier *et al*., 2024) to test the lineage predictions of Maia, Alves-Gomes (2012), comparing the results with COI from specimens throughout the rio Negro basin to explore concordance between these data types. Genome-wide ddRAD markers may be able to deliver new insights for more robust hypotheses and for the estimation of gene flow and detection of putative hybridization and/or introgression events, to allow us to advance the still incompletely described diversity within *Microsternarchus*. Further, it allows us to perform a quick comparison between the efficacy of the DNA barcode in relation to genomic markers for gymnotiform fish. Given the limited swimming capabilities of these fishes (JAG, 2024, pers. obs.), the vast distances between margins and high flow rate of rivers found in several regions of the Amazonas basin likely serve as significant barriers to gene flow among populations or putative species. Consequently, substantial discordance between mitochondrial and nuclear DNA variation might not be expected.

MATERIAL AND METHODS

Collection of specimens and DNA extraction. A total of 74 individuals (72 *Microsternarchus* and 2 *Racenisia* Mago-Leccia, 1994) were selected for this study from a much larger collection of thousands of specimens collected between 2002 and 2023. These individuals represent a targeted subsample, including representatives of each lineage described by Maia, Alves-Gomes (2012), as well as two new lineages (F and G). Collections were conducted throughout the rio Negro basin coordinated by the Laboratório de Fisiologia Comportamental e Evolução (LFCE) from the Instituto Nacional de Pesquisas da Amazonia (INPA), Manaus, Brazil (Tab. 1).

Gymnotiform electric fish were located in the field with the help of an “electric fish detector”, which consists of a home-made electronic circuit that captures the electric discharges of fishes using an underwater probe/electrode, amplifies the signal, and outputs it to a loudspeaker. Once located, the fish were captured primarily with dip or seine nets, depending on the specific conditions of the sampling location. After capture,

TABLE 1 | Specimens sequenced for ddRAD and COI. *Indicates COI specimen is from an individual from the same collecting event as the ddRAD specimen. **Indicates COI specimen is from an individual from the same site, but different collecting event. UF = University of Florida.

Individual	Species	ddRAD Museum Number	ddRAD Voucher Specimen	ddRAD SRA number	COI Museum number	COI Voucher specimen	COI GenBank number	
S101	*Microsternarchus* sp."B"	INPA-ICT 062565	0702111222_E_TIA_007	SRR38350309	INPA-ICT 62566	0811041400E_TIA_007	PZ357009	**
S140	*Microsternarchus* sp."B"	INPA-ICT 062597	0810311100_D_MRU_004	SRR38350308	INPA-ICT 62597	0810311100_D_MRU_004	PZ357010	
S133	*Microsternarchus* sp."B"	INPA-ICT 062597	0810311100_D_MRU_017	SRR38350297	INPA-ICT 62597	0810311100_D_MRU_017	PZ357011	
S124	*Microsternarchus* sp."B"	INPA-ICT 062567	0810311130_E_IAB_011	SRR38350286	INPA-ICT 62567	0810311130_E_IAB_011	PZ357012	
S166	*Microsternarchus* sp."C"	INPA-ICT 062574	1112081150_D_PRT_0020	SRR38350275	INPA-ICT 62574	1112081150_D_PRT_0020	PZ357020	
S182	*Microsternarchus* sp."C"	INPA-ICT 062575	1112081400_D_CAP_0003	SRR38350264	INPA-ICT 62575	1112081400_D_CAP_0003	PZ357021	
S174	*Microsternarchus* sp."C"	INPA-ICT 062576	1112091430_D_ABZ_016	SRR38350258	INPA-ICT 62576	1112091430_D_ABZ_016	PZ357022	
S103	*Microsternarchus* sp."C"	INPA-ICT 062576	1112091430_D_ABZ_026	SRR38350257	INPA-ICT 62576	1112091430_D_ABZ_022	PZ357023	*
S24	*Microsternarchus* sp."C"	INPA-ICT 062577	1211071610_D_CAM_004	SRR38350256	INPA-ICT 62577	1211071610_D_CAM_007	PZ357029	*
S121	*Microsternarchus* sp."C"	INPA-ICT 062581	1211081354_D_EXT_019	SRR38350255	INPA-ICT 62581	1211081354_D_EXT_0003	Pz357031	*
S72	*Microsternarchus* sp."C"	INPA-ICT 062578	1211120800_E_CAX_012	SRR38350307	INPA-ICT 62578	1211120800_E_CAX_012	PZ357033	
S34	*Microsternarchus* sp."C"	INPA-ICT 062585	2001071020_D_GNS_0005	SRR38350306	INPA-ICT 62585	2001071020_D_GNS_0012	PZ357007	*
S4	*Microsternarchus* sp."C"	INPA-ICT 062586	2004161000_D_TPR_005	SRR38350305	INPA-ICT 62586	2004161000_D_TPR_005	PZ357039	
S1	*Microsternarchus* sp."C"	INPA-ICT 062586	2004161000_D_TPR_008	SRR38350304	INPA-ICT 62586	2004161000_D_TPR_008	PZ357040	
S65	*Microsternarchus* sp."C"	INPA-ICT 062586	2004161000_D_TPR_017	SRR38350303	INPA-ICT 62586	2004161000_D_TPR_018	PZ357041	*
S57	*Microsternarchus* sp."C"	INPA-ICT 062586	2004161000_D_TPR_020	SRR38350302	INPA-ICT 62586	2004161000_D_TPR_024	PZ357042	*
S81	*Microsternarchus* sp."C"	INPA-ICT 062587	2006040945_D_TPR_009_2	SRR38350301	INPA-ICT 62587	2006040945_D_TPR_014	PZ357043	*
S41	*Microsternarchus* sp."C"	INPA-ICT 062588	2011251015_D_LIN_003	SRR38350300	INPA-ICT 62588	2011251015_D_LIN_018	PZ357044	*
S17	*Microsternarchus* sp."C"	INPA-ICT 062588	2011251015_D_LIN_022_2	SRR38350299	INPA-ICT 62588	2011251015_D_LIN_019	PZ357045	*
S20	*Microsternarchus* sp."C"	INPA-ICT 062589	2011251330_I_BVB_002	SRR38350298	INPA-ICT 62589	2011251330_I_BVB_002	PZ357046	
S60	*Microsternarchus* sp."C"	INPA-ICT 062589	2011251330_I_BVB_006	SRR38350296	INPA-ICT 62589	2011251330_I_BVB_006	PZ357047	
S83	*Microsternarchus* sp."C"	INPA-ICT 062589	2011251330_I_BVB_007	SRR38350295	INPA-ICT 62589	2011251330_I_BVB_009	PZ357048	*
S44	*Microsternarchus* sp."C"	INPA-ICT 062590	2011261430_D_TPR_001	SRR38350294	INPA-ICT 62590	2011261430_D_TPR_001	PZ357049	
S28	*Microsternarchus* sp."C"	INPA-ICT 062590	2011261430_D_TPR_014	SRR38350293	INPA-ICT 62590	2011261430_D_TPR_014	PZ357050	
S73	*Microsternarchus* sp."C"	INPA-ICT 062592	2106081115_D_TPR_002	SRR38350292	INPA-ICT 62592	2106081115_D_TPR_004	PZ357053	*
S42	*Microsternarchus* sp."C"	INPA-ICT 062592	2106081115_D_TPR_003	SRR38350291	INPA-ICT 62592	2106081115_D_TPR_006	PZ357054	*
S58	*Microsternarchus* sp."C"	INPA-ICT 062592	2106081115_D_TPR_011	SRR38350290	INPA-ICT 62592	2106081115_D_TPR_012	PZ357055	*
S67	*Microsternarchus* sp."C"	INPA-ICT 062594	2111290745_E_ZEZ_0006	SRR38350289	INPA-ICT 62594	2111290745_E_ZEZ_0006	PZ357056	
S6	*Microsternarchus* sp."C"	INPA-ICT 062595	2111300809_I_PAV_0002	SRR38350288	INPA-ICT 62595	2111300809_I_PAV_0002	PZ357057	
S94	*Microsternarchus* sp."C"	INPA-ICT 062596	2303120930_I_ARR_002	SRR38350287	INPA-ICT 62591	2009081000_I_ARR_0006	PZ357058	**
S35	*Microsternarchus* sp."C"	INPA-ICT 062596	2303120930_I_ARR_005	SRR38350285	INPA-ICT 62591	2009081000_I_ARR_0010	PZ357059	**
S106	*Microsternarchus* sp."D/C"	INPA-ICT 062569	0902171400_E_MUR_015	SRR38350284	INPA-ICT 62569	0902171400_E_MUR_015	PZ357013	
S148	*Microsternarchus* sp."D/C"	INPA-ICT 062570	0902171530_E_ARU_010	SRR38350283	INPA-ICT 62570	0902171530_E_ARU_010	PZ357014	
S169	*Microsternarchus* sp."D/C"	INPA-ICT 062571	0902181600_E_AZU_05	SRR38350282	INPA-ICT 62571	0902181600_E_AZU_05	PZ357015	
S183	*Microsternarchus* sp."D/C"	INPA-ICT 062572	0904230900_E_STC_003	SRR38350281	INPA-ICT 62572	0904230900_E_STC_003	PZ357016	
S153	*Microsternarchus* sp."D/C"	INPA-ICT 062572	0904230900_E_STC_017	SRR38350280	INPA-ICT 62572	0904230900_E_STC_017	PZ357017	
S164	*Microsternarchus* sp."D/C"	INPA-ICT 062572	0904230900_E_STC_023	SRR38350279	INPA-ICT 62572	0904230900_E_STC_023	PZ357018	
S180	*Microsternarchus* sp."D/C"	INPA-ICT 062573	0905261600_D_AUA_022	SRR38350278	INPA-ICT 62573	0905261600_D_AUA_022	PZ357019	
S187	*Microsternarchus* sp."D"	INPA-ICT 062564	0212061105_E_JAR_022	SRR38350277	INPA-ICT 62564	0212061105_E_JAR_022	PZ357008	

TABLE 1 | (Continued)

Individual	Species	ddRAD Museum Number	ddRAD Voucher Specimen	ddRAD SRA number	COI Museum number	COI Voucher specimen	COI GenBank number	
S115	*Microsternarchus* sp."D"	INPA-ICT 062582	1504101300_R_MCR_005	SRR38350276	INPA-ICT 62582	1504101300_R_MCR_005	PZ357035	
S181	*Microsternarchus* sp."D"	INPA-ICT 062582	1504101300_R_MCR_008	SRR38350274	INPA-ICT 62582	1504101300_R_MCR_008	PZ357036	
S165	*Microsternarchus* sp."D"	INPA-ICT 062582	1504101300_R_MCR_010	SRR38350273	INPA-ICT 62582	1504101300_R_MCR_010	PZ357037	
S141	*Microsternarchus* sp."D"	INPA-ICT 062583	1504101500_R_BAL_019	SRR38350272	INPA-ICT 62583	1504101500_R_BAL_019	PZ357038	
S173	*Microsternarchus* sp."E"	INPA-ICT 062579	1210240835_E_SBT_006	SRR38350271	INPA-ICT 62579	1210240835_E_SBT_006	PZ357024	
S108	*Microsternarchus* sp."E"	INPA-ICT 062579	1210240835_E_SBT_007	SRR38350270	INPA-ICT 62579	1210240835_E_SBT_007	PZ357025	
S114	*Microsternarchus* sp."E"	INPA-ICT 062579	1210240835_E_SBT_009	SRR38350269	INPA-ICT 62579	1210240835_E_SBT_005	PZ357026	*
S161	*Microsternarchus* sp."E"	INPA-ICT 062580	1210251730_E_JGR_032	SRR38350268	INPA-ICT 62580	1210251730_E_JGR_032	PZ357027	
S188	*Microsternarchus* sp."E"	INPA-ICT 062580	1210251730_E_JGR_034	SRR38350267	INPA-ICT 62580	1210251730_E_JGR_034	PZ357028	
S112	*Microsternarchus* sp."E"	INPA-ICT 062584	1504101000_R_CIR_001	SRR38350266	INPA-ICT 62584	1504101000_R_CIR_001	PZ357034	
S29	*Microsternarchus* sp."F"	INPA-ICT 053288	20151106_25%_P1	SRR38350265	INPA-ICT 53288	20151106_25%_P2	PZ357051	*
S111	*Microsternarchus* sp."F"	INPA-ICT 053288	20151106_25%_P3	SRR38350263	INPA-ICT 53288	20151106_25%_P3	PZ357052	
S32	*Microsternarchus* sp."G"	INPA-ICT 062581	1211081354_D_EXT_012	SRR38350262	INPA-ICT 62581	1211081354_D_EXT_012	PZ357030	
S80	*Microsternarchus* sp."G"	INPA-ICT 062581	1211081354_D_EXT_016	SRR38350261	INPA-ICT 62581	1211081354_D_EXT_0017	PZ357032	*
S36	*Racenisia fimbriipinna*	INPA-ICT 062590	2011261430_D_TPR_004	SRR38350260	UF 177352	2340	KF533338	**
S91	*Racenisia fimbriipinna*	INPA-ICT 062593	2106081030_D_MAA_0005	SRR38350259	UF 177352	2339	KF533337	**

fish were kept alive and carefully placed into 20–30 L plastic bags typically used for ornamental fish transport, filled to one-quarter of their volume with water from the collection site and atmospheric air. These bags were stored inside cooler boxes and transported to a laboratory facility, where the Electric Organ Discharges (EODs) were recorded and the fish were subsequently euthanized, measured, photographed, and preserved in 70% alcohol for DNA extraction. Physiochemical parameters of the water (temperature, conductivity, pH, and oxygen) were measured with a YSI (556 MPS) multi-probe device at each capture site. The geographic coordinates of the collection points are shown in Tab. 2.

For DNA extraction of samples collected prior to 2020, a small piece of muscle was removed from the left dorsal region of the body and digested following a slightly modified phenol-chloroform protocol (Sambrook, Russel, 2006). The DNA samples were dissolved in 0.1x TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and stored at -20°C. DNA quality was visualized on a 1% agarose gel and quantification was performed using a Nanodrop 2000 spectrophotometer. Once the concentrations were determined, the DNA sample was diluted to a working concentration of 30–50 ng/µl. For samples collected after 2020, we used a Qiagen DNeasy Blood and Tissue Kit, following manufacturer recommendations.

TABLE 2 | Latitude and longitude coordinates in decimal degrees for all *Microsternarchus* individuals.

Individual	sNMF Group	Latitude	Longitude
S101	B	0.213	-67.304
S124	B	1.130	-66.843
S133	B	1.128	-61.868
S140	B	1.128	-61.868
S1	C1	-2.727	-60.961
S28	C1	-2.727	-60.961
S4	C1	-2.727	-60.961
S42	C1	-2.728	-60.961
S44	C1	-2.727	-60.961
S57	C1	-2.727	-60.961
S58	C1	-2.728	-60.961
S65	C1	-2.727	-60.961
S73	C1	-2.728	-60.961
S81	C1	-2.727	-60.961
S17	C2_1	-2.557	-60.999
S20	C2_1	-2.531	-61.021
S34	C2_1	-2.656	-60.913
S35	C2_1	-2.624	-60.918
S41	C2_1	-2.557	-60.999
S6	C2_1	-2.599	-60.902
S60	C2_1	-2.531	-61.021
S67	C2_1	-2.410	-60.932
S83	C2_1	-2.531	-61.021
S94	C2_1	-2.624	-60.918
S103	C2_2	1.161	-61.342
S121	C2_2	1.432	-61.619
S166	C2_2	1.094	-61.949
S174	C2_2	1.162	-61.342
S182	C2_2	1.093	-61.916
S24	C2_2	0.811	-61.666
S72	C2_2	0.858	-62.015
S187	D	-0.318	-65.308
S106	D/C3	2.966	-60.796
S148	D/C3	3.067	-60.713
S153	D/C3	2.745	-60.668
S164	D/C3	2.745	-60.668
S169	D/C3	2.826	-60.594
S180	D/C3	2.939	-61.054
S183	D/C3	2.745	-60.668
S115	D3	-3.191	-60.713
S141	D3	-3.174	-60.710
S165	D3	-3.191	-60.713
S181	D3	-3.191	-60.713
S108	E1	-2.284	-64.678
S112	E1	-3.200	-60.685
S114	E1	-2.284	-64.678
S173	E1	-2.284	-64.678
S161	E2	-2.465	-64.811
S188	E2	-2.465	-64.811
S111	F	-2.966	-59.916
S29	F	-2.966	-59.916
S32	G	1.432	-61.619
S80	G	1.432	-61.619

Sequencing. The COI barcoding region was amplified for 53 *Microsternarchus* specimens using PCR primers, BOL-COIfishF1 (5' TCAACYAATCAYAAAGA-TATYGGCAC 3') and BOL-COIfishR1 (5' ACTCYGGGTGRCCRAARAATCA 3') following Ward *et al*. (2005). The purified PCR products were directly sequenced with the use of the ABI PRISM Big Dye Terminator Cycle Sequencing (Applied Biosystems) with the same primers used for amplification and resolved in an ABI 377 automated DNA sequencer (Perkin Elmer), following the protocol accompanying the sequencing kit. Sequences were checked, edited and aligned using the program SEQED v. 1.0.3.1992. This part of the study was conducted at the Molecular Biology core facility at the National Institute of Amazonian Research, Brazil. Sequences of COI for two *Racenisia* specimens were downloaded from NCBI (Tab. 1).

For ddRAD sequencing, we followed the protocol developed by Peterson *et al*. (2012), modified to use the enzymes *Mse*I and *Pst*I and a 350–550-bp size selection. Genomic libraries for 53 *Microsternarchus* samples and 2 *Racenisia* (Tab. 1) were prepared at the Smithsonian National Museum of Natural History Laboratory of Analytical Biology. Preliminary identifications were based on COI data, morphology and EOD waveforms. Sequencing was performed by Admera Health using the Illumina NovaSeq X Plus system. Libraries were pooled and loaded onto one lane of a 10B flow cell, then sequenced in a 2x150 bp paired-end format with a 10–15% PhiX spike-in to ensure high nucleotide diversity.

COI dataset assembly. Of the 53 *Microsternarchus* specimens sequenced for ddRAD, only 34 had COI sequenced for this project. To complete the COI dataset for the remaining 19 specimens, individuals were substituted from the same collecting event if available (16) or from the same locality, but different collecting event (3) (Tab. 2). The determination that these sequences from different individuals belong to the correct lineage was based on previous COI results (Maia, Alves-Gomes, 2012; Pinilla *et al*., 2025). The *Microsternarchus* sequences plus the *Racenisia* from NCBI were checked, edited, and aligned using SEQED v. 1.0.3.1992. The final COI matrix contained 55 individuals and 686 bp.

ddRAD dataset assembly. Demultiplexed raw sequence reads for the 53 *Microsternarchus* and 2 *Racenisia* specimens were filtered and assembled into ddRAD loci using ipyrad v. 0.9.93 (Eaton, Overcast, 2020). Sites with a quality score less than 20 were changed to N and reads with more than five low quality bases were discarded. Adapters were removed and trimmed reads shorter than 35 bp were discarded. Default parameters for *de novo* assembly were used, with a minimum sequence depth of 6 for base-calling and a cluster threshold of 0.85. Loci with reliable sequence data for at least 45 specimens (~85% of total) were retained. This dataset is referred to as the "tree dataset", and contains all 55 individuals and 2,307,154 bp. A second ipyrad assembly was made with all the same parameters, but excluding the 2 *Racenisia,* as outgroups can affect population structure analyses (Jombart *et al*., 2010; Frichot *et al*., 2014). This dataset is known as the "*Micro* dataset", and contains the 53 *Microsternarchus* individuals and 2,336,313 bp.

For the population structure analyses (*Micro* dataset), the assembly from ipyrad was put through more stringent filtering steps using vcftools v. 0.1.17 (Danecek *et al*.,

2011). First, SNPs with a minor allele frequency less than 5% were removed to reduce noise from rare variants that may represent sequencing error. SNPs were then filtered based on completeness, only retaining SNPs reliably scored in >90% of individuals. Lastly, as most population structure analyses assume SNPs to be biallelic and unlinked, non-biallelic SNPs were filtered out, and the effects of linkage were mitigated by keeping only the most complete SNP (defined by the highest number of genotyped individuals) for each RAD locus. For loci with multiple SNPs meeting this criterion, one was randomly selected. This procedure resulted in 3,503 unlinked SNPs for the 53 individuals. Completeness of individuals was then assessed, and all except one had less than 35% missing data. However, the one individual with high missing data was retained, as excluding it would have left only a single member of a proposed lineage, and this can cause problems for population structure algorithms (Puechmaille, 2016; Lawson *et al*., 2018; Prates *et al*., 2024).

Phylogenetic inference. Phylogenies for the COI dataset and the concatenated ddRAD loci from the "tree dataset" were estimated under maximum likelihood as implemented in IQ-TREE v. 2.3.6 (Minh *et al*., 2020). The best substitution model was determined using ModelFinder (Kalyaanamoorthy *et al*., 2017) and 1,000 UFBoot replicates were used to assess node support (Hoang *et al*., 2018).

A coalescent species tree was also estimated for the "tree dataset" in SVDQuartets (Chifman, Kubatko, 2014) implemented in PAUP* v. 4.0a build 168 (Swofford, 2003). SVDQuartets assumes unlinked SNPs. Ipyrad automatically outputs one random SNP per RAD locus in phylip format, for a total of 8,270 SNPs. This was used rather than the reduced 3,503 unlinked SNPs as it includes outgroups necessary for rooting and maximizes informative sites available for quartet evaluation, which is beneficial for coalescent-based species tree estimation (Chifman, Kubatko, 2014). For the 55 taxa in the "tree dataset" there were 341,055 possible quartets. All possible quartets were evaluated and support was assessed with 1,000 bootstrap replicates.

Species delimitation using COI. To infer groupings based on the COI dataset, we ran three different single locus delimitation tools: 1) Poisson Tree Process (PTP) model in both a maximum likelihood and Bayesian framework (Zhang *et al*., 2013), 2) Assemble Species by Automatic Partitioning (ASAP; Puillandre *et al*., 2021), and 3) Automatic Barcode Gap Discovery (ABGD; Puillandre *et al*., 2012). PTP was implemented through the webserver (https://species.h-its.org/ptp/) which gives the ML estimate and best fit Bayesian scheme. ASAP and ABGD were both downloaded from iTaxoTools (Vences *et al*., 2021) and run assuming K2P distances. ASAP reports multiple schemes, so the one with the lowest ASAP-score and a significant p-value was chosen. ABGD was run recursively and reported multiple schemes. The best delimitation model was determined by recurrence of the barcode gap distance and congruence of species numbers with other methods. Outgroups were removed from all single-locus delimitation analyses.

Population structure analyses. To determine if the groupings observed in the COI delimitation analyses were mirrored in the ddRAD dataset, we used the "*Micro* dataset" to run population structure analyses. *Microsternarchus* individuals were first coded by their previously assigned COI lineages. Then, using the R package adegenet v. 2.1.10

(Jombart, 2008; Jombart, Ahmed, 2011) we ran a PCA to visualize the distribution of nucleotide diversity in relation to these *a priori* assignments (Fig. **S1**).

Then, Discriminant Analysis of Principal Components (Jombart *et al*., 2010) or DAPC was used to identify statistical clusters within the data. First, individuals were re-assigned to populations using the find.clusters() function. The "best" number of populations (K) was determined automatically using the "diffNgroup" criterion which splits the difference between successive BIC values into two groups with Ward's clustering method. The K value that is retained is the one before the first group switch. K values were tested from 1 through 20, as 20 is slightly more than double the identified COI lineages to allow for further division in the ddRAD data. Each run of find.clusters() used 50 randomly chosen starting centroids and ran for 10,000 iterations, and the maximum number of principal components were retained.

Next, this clustering information was put through an initial DAPC run with 20 principal components, and the optim.a.score() was used to determine the optimal number of principal components to retain. DAPC was then re-ran using the optimal number of principal components.

Following this, we used the sNMF algorithm (Frichot *et al*., 2014), implemented in the R package LEA v. 3.8.0 (Frichot, Francois, 2015) to estimate group assignments and ancestry coefficients for each specimen. This allowed us to detect potential genomic admixture resulting from past hybridization. We initially used an exploratory sNMF run to identify the optimal K (again tested from 1 through 20) and the optimal alpha (regularization parameter) for each value of K. Alpha values of 1, 5, 10, 50, 100, 500, and 1000 were used to test across several orders of magnitude (Pyron *et al*., 2023). Twenty exploratory iterations were run for each alpha and K combination. To identify the optimal K value, we calculated a combined z-score from the lowest cross-entropy value and highest adjusted rand index to account for both model fit and clustering stability. The optimal alpha for each K was determined by minimizing the median cross-entropy across the exploratory runs. Once the optimal values were selected, sNMF was re-run for 100 iterations.

All results were visualized with ggplot2 v. 3.4.4 (Wickham *et al*., 2016). All analyses using R were conducted in R studio v. 2024.12.0+467 using R v. 4.4.3 (R Development Core Team, 2025).

Genetic distances. Genetic distances were calculated to test the distinctiveness of groups identified by the various species delimitation and population structure analyses. For the COI data, K2P distances were calculated, while uncorrected p-distances were calculated for the ddRAD data. The decision was made to use p-distance as it is directly comparable to K2P distance, unlike a chord distance such as Cavalli-Sforza and Edwards. Distances were calculated between and within the groups identified by the COI species delimitation methods and the groups identified by the ddRAD population structure analyses. K2P distances for the COI data were calculated with the dist.dna function implemented in the R package ape v. 5.8 (Paradis, Schliep, 2019) and the meandist function implemented in the R package vegan v. 2.7–1 (Oksanen *et al*., 2025). P-distances for the ddRAD data were calculated using the R package hierfstat v. 0.5–11 (Goudet *et al*., 2022).

RESULTS

Phylogenetic inference. Maximum likelihood phylogenies estimated from the COI and concatenated ddRAD loci were compared for concordance in topology and the composition of monophyletic groups. All phylogenies were rooted on *Racenisia*. For the COI matrix, the best-fit model determined by Modelfinder using BIC was HKY+F+G4. Six discrete monophyletic lineages within *Microsternarchus* were identified, five of which are well-supported (ultrafast bootstrap values >90%) and a sixth lineage received moderate support (ultrafast bootstrap value = 71%) (Figs. 1A, **S2A**). These six lineages are designated as B (n = 4), C (n = 27), D (n = 12), E (n = 6), F (n=2), and G (n =2). The terminology for Lineages B to E follows that of Maia, Alves-Gomes (2012), while Lineages F and G were absent in the mentioned paper and identified here. Lineage B is placed as the sister group to the rest of *Microsternarchus*, with the remaining lineages split into two clades: Lineage E + Lineage G and Lineage F + (Lineage C + Lineage D) (Fig. 1A). These results largely align with those proposed by Maia, Alves-Gomes, except for the two new lineages identified here.

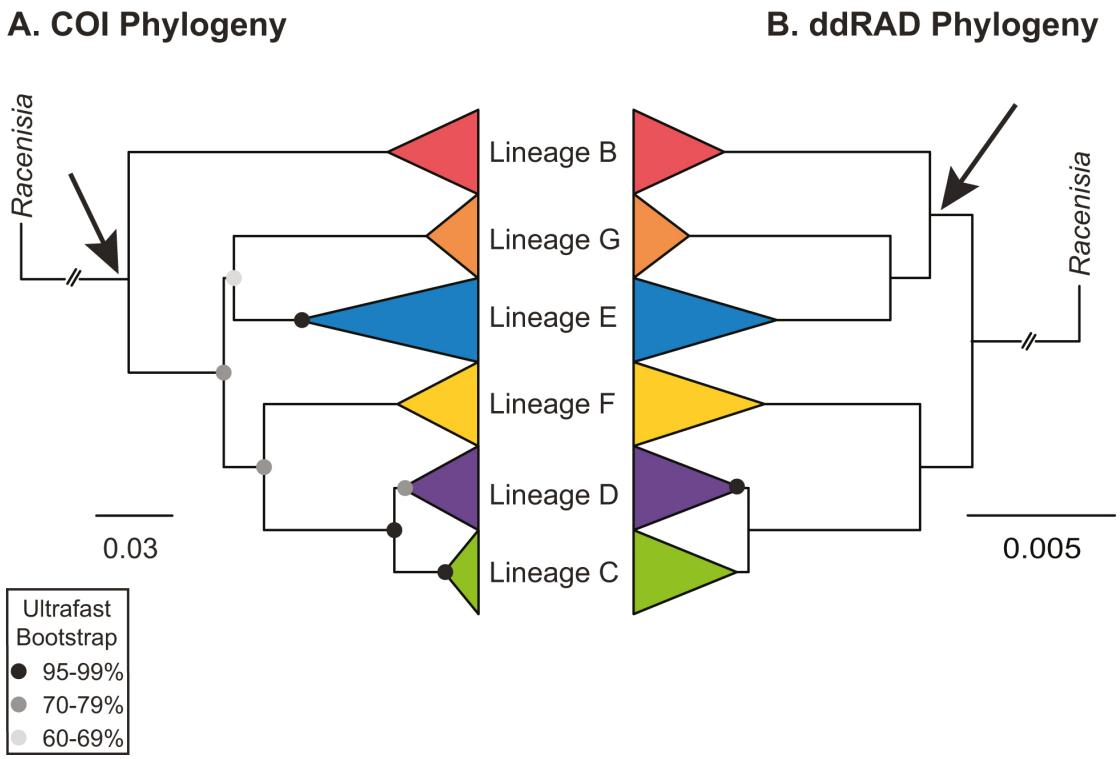

FIGURE 1 | Topological comparisons between COI and ddRAD data. In both the COI and ddRAD phylogenies there are six distinct clades of *Microsternarchus* (Lineages B – G). The only topological difference, indicated by the black arrow, is the placement of Lineage B as either sister to Lineage G + Lineage E (ddRAD) or sister to all *Microsternarchus* lineages (COI). (**A**) Maximum likelihood phylogeny of COI data for 53 in-group taxa and 2 out-group taxa. Branch lengths are shown in the number of substitutions per site below the tree. (**B**) Maximum likelihood phylogeny of ddRAD data for 53 in-group taxa, 2 outgroup taxa. All loci had all least 45 individuals (85%), resulting in a 2,307,154 bp alignment. Branch lengths are shown in the number of substitutions per site below the tree. Circles on nodes represent ultrafast bootstrap values between 60 and 99, while all unlabeled nodes had ultrafast bootstrap values of 100. The terminology for Lineages B to E follows that of Maia, Alves-Gomes (2012), while Lineages F and G were absent in the mentioned paper and identified here. Although not included here, *M. brevis* would be a member of Lineage E and *M. javieri* would be a member of Lineage C.

The maximum likelihood phylogeny inferred for the ddRAD "tree dataset" had a best-fit model of TIM3+F+I+G4. The same six discrete monophyletic *Microsternarchus* lineages were identified, but in contrast to COI, all six are well-supported (ultrafast bootstrap values >98%) (Figs. 1B, **S2B**). The relationships between the lineages are largely consistent with those from the COI analysis, with only one difference: Lineage B is now the sister group of the Lineage E + Lineage G clade, rather than of the rest of *Microsternarchus* (Fig. 1B). The composition of the lineages also changes slightly between ddRAD and COI. The COI clade D3 is nested within Lineage C with ddRAD (referred to as D/C3 hereafter) (Fig. 2). As a result, Lineage C now has 34 individuals, and D has five.

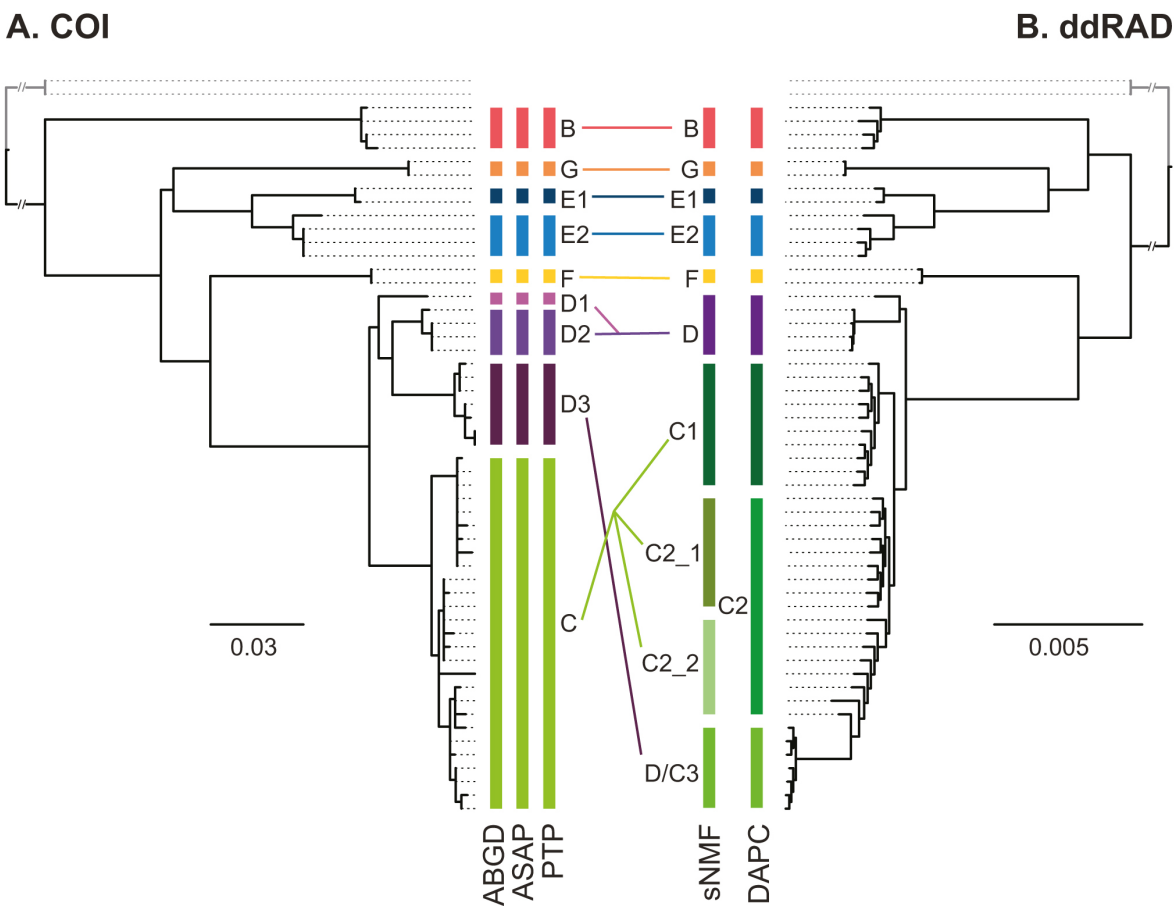

FIGURE 2 | Tanglegram of clustering results between COI and ddRAD data. **A** (COI data): the three single-locus delimitation tools for COI (ABGD, ASAP, PTP) support the same nine groups of *Microsternarchus*, all of which are monophyletic. Lineages B, G, F, and C are equivalent to single clusters, while Lineage E is split into two clusters and Lineage D into three clusters. **B** (ddRAD data): the two population structure analyses (DAPC and sNMF) give similar results, both to each other and to COI, supporting identical B, G, E1, E2, and F clusters. Lineage D3 from COI is also resolved as monophyletic but nested within Lineage C in the ddRAD phylogeny, and is referred to as D/C3. Lineage C from COI is split into two (DAPC) or three (sNMF) clades in the ddRAD tree, but without complete monophyly. We combine D1 and D2 from COI into a single Lineage D.

The phylogeny inferred under the multispecies coalescent model for the "tree dataset" also identified the same six lineages, and the topology is consistent with the ddRAD maximum likelihood tree. However, support for Lineages C and D was notably lower (Fig. **S3**).

COI delimitation. All outgroups were removed prior to the following analyses. The maximum likelihood framework of PTP estimated nine groups, four of which correspond directly to lineages: B, C, G, F. The remaining two lineages, D and E, were divided into three and two groups, respectively (Fig. 2). The Bayesian version, bPTP, estimated between 6 and 32 groups, with the best supported model identifying 19. Due to the higher count in bPTP compared to other methods, these results were excluded from further analyses (see Fig. **S4**). ASAP estimated between 5 and 37 groups, with the scheme resulting in the best score and significant p-value (<0.05) resolving the same nine groups as PTP. ABGD detected two recurring barcode gaps (0.439% and 1.03% K2P distances). The 0.439% gap corresponded to partitions that resolved either nine or ten groups. Because the composition of the nine-group solution was congruent with the results from PTP and ASAP, we selected nine groups as the best model for ABGD as well.

ddRAD population structure. The optimal number of K clusters in the *Micro* dataset" resolved by DAPC was nine (Figs. 2–3), based on the diffNgroup criterion (see methods). In the DAPC analysis, Lineages B, F, and G were identified as distinct groups, while lineage E was split into two groups. Lineage C was divided into three groups, and Lineage D formed a single group except for one individual. This individual was sister to Lineage D in the ddRAD phylogeny but clustered with one of the sub-groups of C. The posterior probability of group assignment was nearly 1.0 for all 53 individuals (Fig. **S5**).

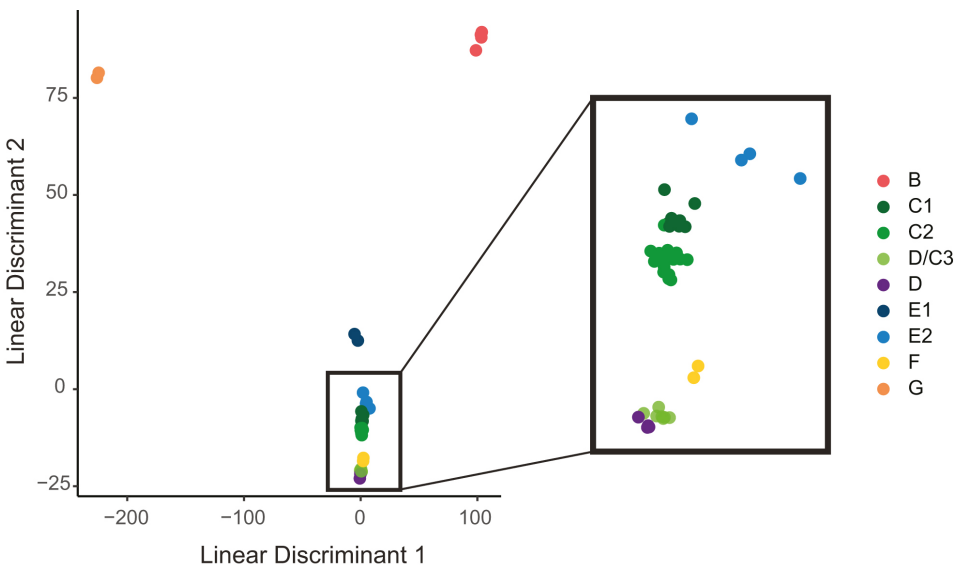

FIGURE 3 | Discriminant Analysis of Principal Components for lineages of *Microsternarchus*. A scatterplot (left) showing the first two linear discriminants, which accounts for the majority of the genetic variation between the nine identified groups. B and G are the most differentiated from the rest. The zoomed-in view (outlined in the black square on the right) visualizes the separation between E2, C, D, and F. Colors correspond to lineages.

The optimal number of K clusters resolved by sNMF was ten, as determined by our combined z-score method (Fig. **S6**; Tab. **S7**). Although both K = 1 and K = 2 had better combined z-scores, these were influenced by a mean ARI of 1.0. This is the expected result for K=1, where all individuals are grouped together, and for K = 2 which reflects the split between the two major clades (B,(E,G)) and (F,(C,D)). Therefore, we discarded these two hypotheses. The K = 10 hypothesis closely matches that of DAPC, with the only difference being that lineage C is subdivided into four rather than three groups (Figs. 2, 4). The levels of admixture between lineages were negligible, with only 13 of the 53 *Microsternarchus* individuals having ancestry coefficients less than 0.9 for their primary assigned group (Tabs. 3, **S8**).

Comparisons. All three COI-based delimitation analyses, along with both ddRAD clustering methods, identified the same structure: Lineage B, Lineage G, Lineage F, and two E sub-lineages, all of which are monophyletic in both phylogenies (Fig. 2). This structure is further supported by genetic distance measures. These lineages each exhibit 0.1 or greater K2P divergence (COI data) from any other lineage and 0.12 or greater p-distance divergence (ddRAD data) (Tabs. 4–5). The two E sub-lineages, as expected, were more similar to each other than to any of the other lineages, showing 0.052 divergence in COI and 0.074 different in ddRAD (Tabs. 4–5).

Results for Lineages C and D varied slightly between data types. The COI data clustered all individuals from Lineage C into a single group, whereas DAPC and sNMF split this clade two and three groups respectively. These clades are designated as C1 and C2 for DAPC (Figs. 2B, 3) and C1, C2_1, and C2_2 for sNMF (Figs. 2B, 4). Sub-lineage names used here or after do not match those from Maia, Alves-Gomes (2012). The COI data supported a subdivision of Lineage D into three groups, one of a single individual (D1), one of five individuals (D2), and one of seven individuals (D3), all of which are monophyletic in the phylogeny. DAPC and sNMF both identified D2 and D3 as well, but their placement in the ddRAD phylogeny is different. Rather than sister clades, D3 is nested within C2 (DAPC)/C2_2 (sNMF) and is referred to as D/C3.

The single individual of D1 was clustered within C2 by DAPC and C2_1 by sNMF. The sNMF admixture percentages for this individual were 60% C2_1 and 20% D, making it the only individual with more than 20% of its alleles from multiple lineages. There were two other individuals with 20% of alleles from multiple clusters, however both clusters are within the same lineage, either the two E sub-lineages or C2_1 and C2_2 (Tab. 3).

The genetic distances reflect these findings. K2P distances show very little "within-cluster" genetic variation (0.012) within the COI Lineage C compared to p-distance (0.037) (Tab. 4). However, when compared to the distances between the clusters identified by sNMF, which sub-divided Lineage C the most, K2P distances ranged from 0.009 (C1 to C2_2) to 0.06 (C1 to D/C3) (Tab. 5). The p-distances on the other hand ranged from 0.039 (C2_1 to C2_2) 0.077 (C1 to D/C3) (Tab. 5). The conflicting placement of D/C3 between Lineages D and C depending on the data type is also reflected here: K2P distance showed 0.032 divergence from the rest of Lineage D *versus* 0.059 from Lineage C (Tab. 4). With p-distance, the group was 0.09 different from Lineage D and 0.058–0.077 divergent from the other C sub-lineages (Tab. 5).

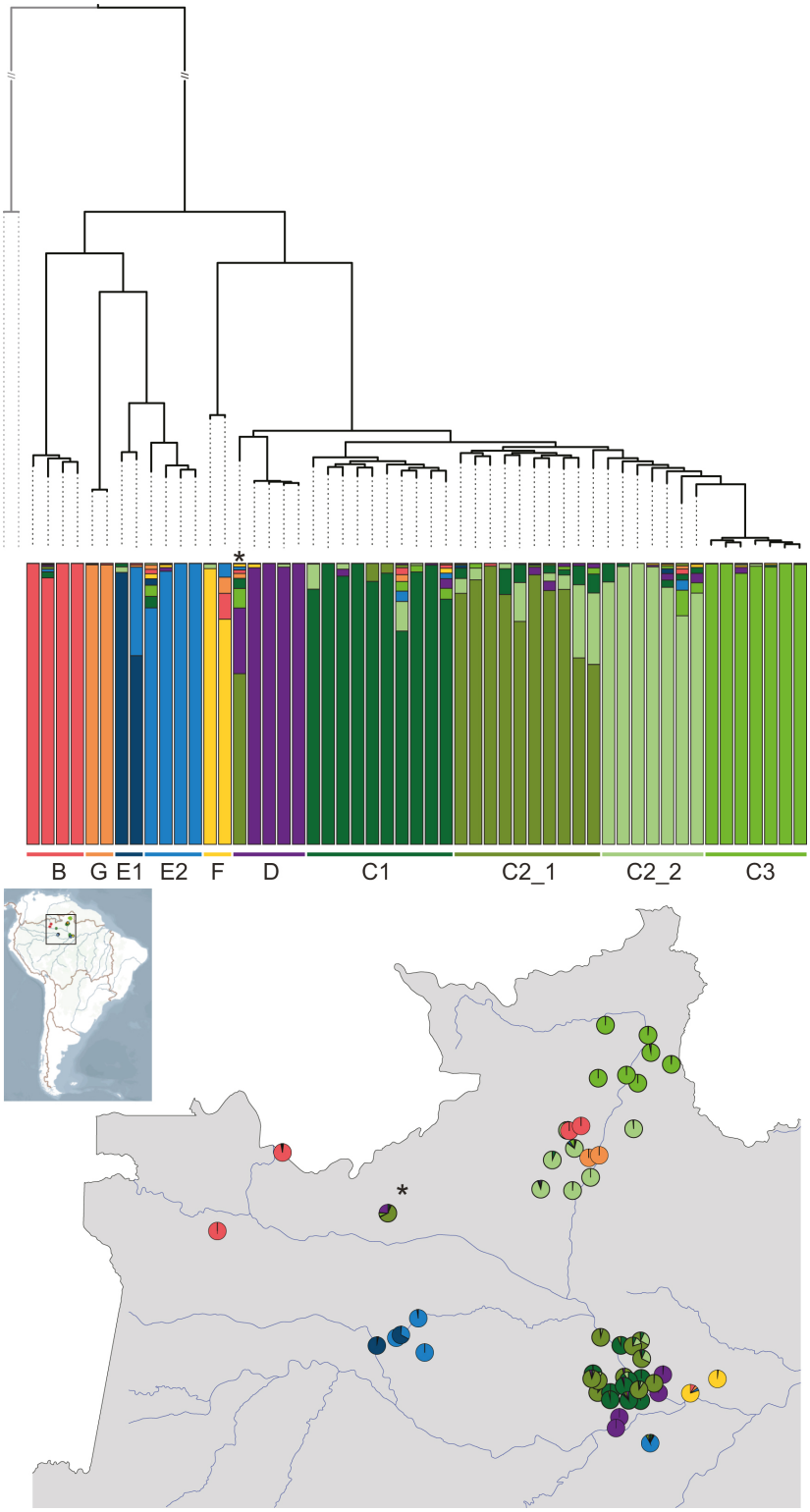

FIGURE 4 | sNMF results show clear separation between ten unique groups of *Microsternarchus*. Top shows the ddRAD phylogeny mapped onto the admixture bar plot for each individual. The majority of individuals have at least 90% of their alleles consistent with their designated group. Bottom shows admixture pie charts plotted on a map to show geographic distribution of the sNMF groups. All groups are restricted to a single locality except B, E2, and D. Points are randomly jittered by a tenth of a decimal degree due to overlap of points. The only individual with at least 20% of its alleles from multiple lineages (indicated with a star) was grouped with D rather than C based on the mixed ancestry, phylogenetic placement, and geographic location.

TABLE 3 | sNMF results. Primary group assignment refers to the sNMF cluster to which the individual was assigned. The two groups with the highest ancestry coefficients are reported. For individuals in which the primary group was 0.999, no secondary group was given. N/A = not applicable. Used when no secondary group is given.

Individual	Primary assignment	Primary ancestry coefficient	Secondary assignment	Secondary ancestry coefficient
S101	B	0.999	N/A	N/A
S133	B	0.999	N/A	N/A
S140	B	0.999	N/A	N/A
S124	B	0.948	C1	0.022
S28	C1	0.999	N/A	N/A
S4	C1	0.999	N/A	N/A
S1	C1	0.993	D	0.006
S57	C1	0.969	C3	0.023
S65	C1	0.965	C2_1	0.034
S73	C1	0.954	D	0.025
S42	C1	0.936	C2_1	0.064
S58	C1	0.908	C2_2	0.089
S81	C1	0.872	C3	0.038
S44	C1	0.759	C2_2	0.105
S83	C2_1	0.990	B	0.009
S17	C2_1	0.959	D	0.027
S60	C2_1	0.941	C2_2	0.042
S34	C2_1	0.907	C2_2	0.051
S35	C2_1	0.903	D	0.033
S94	C2_1	0.893	C2_2	0.053
S20	C2_1	0.889	C1	0.092
S6	C2_1	0.793	C2_2	0.139
S41	C2_1	0.663	C2_2	0.261
S67	C2_1	0.640	C2_2	0.254
S187	C2_1	0.607	D	0.234
S166	C2_2	0.999	N/A	N/A
S24	C2_2	0.989	C1	0.011
S174	C2_2	0.987	D	0.006
S182	C2_2	0.934	C1	0.066
S72	C2_2	0.915	C1	0.025
S103	C2_2	0.894	C3	0.037
S121	C2_2	0.813	C3	0.091
S164	C3	0.999	F	0.000
S106	C3	0.998	E1	0.001
S180	C3	0.998	G	0.002
S183	C3	0.996	E2	0.003
S169	C3	0.990	C2_2	0.008
S153	C3	0.987	C1	0.007
S148	C3	0.964	D	0.021
S115	D	0.999	N/A	N/A
S181	D	0.999	N/A	N/A
S165	D	0.986	C2_2	0.011
S141	D	0.985	F	0.013
S114	E1	0.999	N/A	N/A
S173	E1	0.999	N/A	N/A
S108	E1	0.972	D	0.013
S112	E1	0.841	C1	0.041
S161	E2	0.968	C2_2	0.020
S188	E2	0.671	E1	0.315
S111	F	0.981	C2_2	0.016
S29	F	0.801	B	0.092
S32	G	0.995	C1	0.002
S80	G	0.994	B	0.005

TABLE 4 | Genetic distances between clusters within *Microsternarchus* (B–G) identified by COI delimitation methods. Bolded values on the diagonals represent divergence within the groups. Values below the diagonal represent COI K2P distances, while values above represent ddRAD p-distances.

	B	C	D1	D2	D3	E1	E2	F	G
B	**0.005 / 0.024**	0.357	0.348	0.375	0.379	0.293	0.250	0.162	0.268
C	0.163	**0.012 / 0.037**	0.045	0.063	0.072	0.351	0.307	0.125	0.327
D1	0.142	0.050	**NA**	**0.054**	**0.075**	**0.345**	**0.303**	**0.122**	**0.322**
D2	0.149	0.045	0.035	**0 / 0.006**	0.090	0.368	0.322	0.132	0.342
D3	0.139	0.059	0.048	0.032	**0.006 / 0.014**	0.373	0.328	0.133	0.344
E1	0.131	0.123	0.139	0.117	0.125	**0.007 / 0.023**	0.074	0.169	0.230
E2	0.135	0.123	0.137	0.120	0.138	0.052	**0.002 / 0.012**	0.149	0.202
F	0.133	0.112	0.100	0.108	0.104	0.110	0.102	**0 / 0.001**	0.159
G	0.144	0.141	0.148	0.139	0.143	0.108	0.099	0.123	**0 / 0.001**

TABLE 5 | Genetic distances between clusters within *Microsternarchus* (B–G) identified by sNMF. Bolded values on the diagonals represent divergence within the groups. Values below the diagonal represent COI K2P distances, while values above represent ddRAD p-distances.

	B	C1	C2_1	C2_2	C3	D	E1	E2	F	G
B	**0.005/0.024**	0.356	0.365	0.343	0.379	0.375	0.293	0.250	0.162	0.268
C1	0.165	**0.003/0.025**	0.041	0.041	0.077	0.062	0.351	0.306	0.126	0.327
C2_1	0.160	0.022	**0.014/0.037**	0.039	0.075	0.062	0.360	0.314	0.127	0.334
C2_2	0.160	0.009	0.019	**0.002/0.031**	0.058	0.063	0.338	0.296	0.120	0.315
C3	0.139	0.060	0.058	0.057	**0.006/0.014**	0.090	0.373	0.328	0.133	0.344
D	0.149	0.046	0.045	0.044	0.032	**0/0.006**	0.368	0.322	0.132	0.342
E1	0.131	0.125	0.126	0.118	0.125	0.117	**0.007/0.023**	0.074	0.169	0.230
E2	0.135	0.127	0.124	0.119	0.138	0.120	0.052	**0.002/0.012**	0.149	0.202
F	0.133	0.114	0.112	0.106	0.104	0.108	0.110	0.102	**0/0.001**	0.159
G	0.144	0.137	0.149	0.135	0.143	0.139	0.108	0.099	0.123	**0/0.001**

DISCUSSION

Most of our current understanding of *Microsternarchus* diversity has relied primarily on morphology (Fernández-Yépez, 1968; Mago-Leccia, 1994; Cox Fernandes, Williston, 2017; Cox Fernandes *et al*., 2024; Sousa *et al*., 2024), but a few studies using molecular data have changed the pace of discovery (Maia, Alves-Gomes, 2012; Cox Fernandes *et al*., 2015; Pinilla *et al*., 2025). Our analysis of genome-wide ddRAD markers show strong support for the pattern of lineage diversity originally revealed by COI, with a few exceptions, establishing well-supported groups within *Microsternarchus*. This is contrary to what has been observed in many groups of vertebrates, including bony fishes (Willis *et al*., 2014; Hughes *et al*., 2020; Englmaier *et al*., 2024), amphibians (Firneno *et al*., 2020), reptiles (Salles *et al*., 2025), birds (DeRaad *et al*., 2023) and mammals (Good *et al*., 2015).

There are at minimum six unique lineages supported with both datatypes: B, C, D, E, F, G. Lineages C, D, and E have further subdivisions, although these differ in details depending on analyses of COI or ddRAD. Following Maia, Alves-Gomes (2012), we find high levels of geographic structure for the majority of these lineages. Lineage C is the most widespread, however each of the sub-groups identified by ddRAD analyses is more restricted, similar to the other lineages. Lineages B and E also occur only in two localities. The one E individual separate from the rest may represent a unique sub-lineage not identified here due to small sample size, results corroborated with an expanded COI database (JAG, unpublished data). Lineage B is also widespread, found in both the upper rio Negro and rio Branco. This lineage was also identified as wide-spread without a significant amount of inter-lineage genetic variation by Maia, Alves-Gomes.

The single D1 individual of COI that was grouped with Lineage C by ddRAD analyses is from the middle rio Negro (Fig. 4). While previously members of Lineage D have been captured in this region only, Lineage C was absent (Maia, Alves-Gomes, 2012). Given the geographic distance from the rest of Lineage C, admixture percentages (Fig. 4; Tabs. 3, **S8**), phylogenetic placement (Fig. 2), genetic distance (Tabs. 4–5), and the tendency for these methods to cluster individual taxa within larger groups (Puechmaille, 2016; Lawson *et al*., 2018; Prates *et al*., 2022) we consider this individual as part of D2 to make the ddRAD Lineage D.

While sample size may have had an effect on the single D1 individual of COI, we do not believe that sampling is what causes D/C3 to jump between lineages based upon data type. That Lineage C and D form a monophyletic group is well supported, although the genetic distances between them are significantly lower than between any other lineage comparison (Maia, Alves-Gomes, 2012) (Tabs. 4–5). It is possible that subsequent analyses with more data and samples will collapse these into a single lineage.

The previously unrecognized lineages reported here, Lineages F and G, are supported by both COI and ddRAD data. They are strongly monophyletic (ultrafast bootstrap of 100) in all phylogenetic reconstructions, represent unique groups based on COI delimitation and ddRAD population structure analyses, and are geographically separated, with Lineage G being found solely in the rio Branco drainage in Roraima State, and Lineage F in the Ducke Forest Reserve, near Manaus, in an affluent of the rio Amazonas drainage (Fig. 4).

While mito-nuclear discordance is a consistent result of many molecular studies (Toews, Brelsford, 2012), we find relatively strong congruence between mitochondrial (COI) and nuclear (ddRAD) markers for *Microsternarchus*. This may reflect the limited dispersal capabilities of *Microsternarchus,* which limits opportunities for introgression, one of the primary causes of mito-nuclear discordance (Bonnet *et al*., 2017). These fishes live closely associated with underwater vegetation and, especially due to their small adult size (8–15 cm) and having only their anal fin as a main source of thrust, have limited swimming capabilities (JAG, 2024, pers. obs.) Additionally, the vast distances between margins and high flow rate of rivers found in several regions of the Amazonas basin likely serve as significant barriers to gene flow among populations or putative species.

Current evidence suggests that the majority of the divergence in *Microsternarchus* has occurred in the last 6 million years (Tagliacollo *et al*., 2024). This corresponds with the decline of megawetlands in Western Amazonia and the rio Amazonas in its current distribution becoming fully established (Hoorn *et al*., 2010). This transition from

interconnected wetlands to high-energy river system may have created the isolation that resulted in the diversification observed today.

We refrain from suggesting revised species boundaries at this stage, as our sampling is uneven and will thus lack the statistical power needed for methods such as BPP (Yang, Rannala, 2010) and BFD* as implemented in SNAPP (Leache *et al*., 2014). However, the mito-nuclear concordance demonstrated between the COI and ddRAD datasets in both phylogeny and delimitation/clustering analyses provides further evidence that there remains undescribed diversity within *Microsternarchus*, with a minimum of six unique lineages, or putative species.

ACKNOWLEDGMENTS

The authors would like to thank José C. da Silva for his assistance in the collection of many of the samples analyzed here. We thank the Laboratório Temático de Biologia Molecular which generated the original COI sequences at the Instituto Nacional de Pesquisas do Amazonas and the Laboratory of Analytical Biology at the National Museum of Natural History, Smithsonian Institution for their help in preparing ddRAD libraries. We also thank R. A. Pyron (George Washington University) for help with methodology and theory, A. Santaquiteria (George Washington University) for helpful comments on drafts and figure design, and two anonymous reviewers whose comments helped improve the clarity of the manuscript. Analyses were performed on the Pegaus High Performance Computing cluster at the George Washington University unless otherwise stated.

REFERENCES

- **Alves-Gomes JA, Ortí G, Haygood M, Heiligenberg W, Meyer A.** Phylogenetic analysis of the South American electric fishes (Order Gymnotiformes) and the evolution of their electrogenic system: a synthesis based on morphology, electrophysiology, and mitochondrial sequence data. Mol Biol Evol. 1995; 12(2):298–318. https://doi.org/10.1093/oxfordjournals.molbev.a040204
- **Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K *et al.*** Cryptic species as a window on diversity and conservation. Trends Ecol Evol. 2007; 22(3):148–55. https://doi.org/10.1016/j.tree.2006.11.004
- **Bonnet T, Leblois R, Rousset F, Crochet PA.** A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution. 2017; 71(9):2140–58. https://doi.org/10.1111/evo.13296
- **Chifman J, Kubatko L.** Quartet inference from SNP data under the coalescent model. Bioinformatics. 2014; 30(23):3317–24. https://doi.org/10.1093/bioinformatics/btu530
- **Cox Fernandes C, Keefe RM, Pinilla CE.** *Microsternarchus schonmanni*, a new species of weakly electric fish (Gymnotiformes: Hypopomidae: Microsternarchini) from the Mamoré-Guaporé River Basin, Brazil. PNAS. 2024; 168(1):237–50. https://doi.org/10.1635/053.168.0110
- **Cox Fernandes C, Nogueira A, Alves-Gomes JA.** *Procerusternarchus pixuna*, a new genus and species of electric knifefish (Gymnotiformes: Hypopomidae, Microsternarchini) from the Negro River, South America. PNAS. 2014; 163(1):95–118. https://doi.org/10.1635/053.163.0107

- **Cox Fernandes C, Nogueira A, Williston A, Alves-Gomes JA.** A new species of electric knifefish from the rio Negro, Amazon basin (Gymnotiformes: Hypopomidae, Microsternarchini). PNAS. 2015; 164(1):213–27. https://doi.org/10.1635/053.164.0113
- **Cox Fernandes C, Williston A.** Redescription of *Microsternarchus bilineatus* (Fernández-Yépez 1968) (Gymnotiformes: Hypopomidae, Microsternarchini), with the designation of a neotype. PNAS. 2017; 165(1):105–15. https://doi.org/10.1635/053.165.0107
- **Crampton WGR, de Santana CD, Waddell JC, Lovejoy NR.** Phylogenetic systematics, biogeography, and ecology of the electric fish genus *Brachyhypopomus* (Ostariophysi: Gymnotiformes). PLoS ONE. 2016; 11(10):e0161680. https://doi.org/10.1371/journal.pone.0161680
- **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA *et al.*** The variant call format and VCFtools. Bioinformatics. 2011; 27(15):2156–58. https://doi.org/10.1093/bioinformatics/btr330
- **DeRaad DA, McCullough JM, DeCicco LH, Hime PM, Joseph L, Anderson MJ *et al.*** Mitonuclear discordance results from incomplete lineages sorting, with no detectable evidence for gene flow, in a rapid radiation of *Todiramphus* kingfishers. Mol Ecol. 2023; 32(17):4844–62. https://doi.org/10.1111/mec.17080
- **Eaton DAR, Overcast I.** ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformatics. 2020; 36(8):2592–94. https://doi.org/10.1093/bioinformatics/btz966
- **Englmaier GK, Rodríguez NV, Bravničar J, Zangl L, Persat H, Marić S *et al.*** SNP-based analysis of European *Thymallus* spp. (Salmonidae) reveals extensive mito-nuclear discordance relevant for biogeographic inferences, taxonomy and conservation. Divers Distrib. 2024; 30(7):e13845. https://doi.org/10.1111/ddi.13845
- **Fernández-Yépez A.** Contribución al conocimiento de los peces Gymnotiformes. Evencias. 1968; 20(6):1–29.
- **Firneno TJ Jr, O'Neill JR, Portik DM, Emery AH, Townsend JH, Fujita MK.** Finding complexity in complexes: assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. Mol Ecol. 2020; 29(18):3543–59. https://doi.org/10.1111/mec.15496
- **Frichot E, François O.** LEA: an R package for landscape and ecological association studies. Methods Ecol Evol. 2015; 6(8):925–29. https://doi.org/10.1111/2041-210X.12382
- **Frichot E, Mathieu F, Trouillon T, Bouchard G, François O.** Fast and efficient estimation of individual ancestry coefficients. Genetics. 2014; 4(1):973–83. https://doi.org/10.1534/genetics.113.160572
- **Funk DJ, Omland KE.** Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Ann Rev Ecol Evol Syst. 2003; 34(1):397–423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421
- **Good JM, Vanderpool D, Keeble S, Bi K.** Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. Evolution. 2015; 69(8):1961–72. https://doi.org/10.1111/evo.12712
- **Goudet J, Jombart T.** hierfstat: Estimation and tests of hierarchical f-statistics. R package version 0.5-11. 2022. Available from: https://CRAN.R-project.org/package=hierfstat
- **Harrison RG, Larson EL.** Hybridization, introgression, and the nature of species boundaries. J Hered. 2014; 105(S1):795–809. https://doi.org/10.1093/jhered/esu033
- **Herbert PDN, Cywinska A, Ball SL, deWaard JR.** Biological identifications through DNA barcodes. Proc R Soc B. 2003; 270(1512):313–21. https://doi.org/10.1098/rspb.2002.2218
- **Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS.** UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018; 35(2):518–22. https://doi.org/10.1093/molbev/msx281
- **Hoorn C, Wesselingh F, Ter Steege H, Bermudez M, Mora A, Sevink J *et al.*** Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science. 2010; 330(6006):927–31. https://doi.org/10.1126/science.1194585

- **Hughes LC, Cardoso YP, Sommer JA, Cifuentes R, Cuello M, Somoza GM *et al*.** Biogeography, habitat transitions and hybridization in a radiation of South American silverside fishes revealed by mitochondrial and genomic RAD data. Molec Ecol. 2020; 29(4):738–51. https://doi.org/10.1111/mec.15350
- **Jombart T.** adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008; 24(11):1403–05. https://doi.org/10.1093/bioinformatics/btn129
- **Jombart T, Ahmed I.** adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011; 27(21):3070–71. https://doi.org/10.1093/bioinformatics/btr521
- **Jombart T, Devillard S, Balloux F.** Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010; 11(1):94. https://doi.org/10.1186/1471-2156-11-94
- **Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS.** ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14(6):587–89. https://doi.org/10.1038/nmeth.4285
- **Lawson DJ, van Dorp L, Falush D.** A tutorial on how not over-interpret STRUCTURE and ADMIXTURE bar plot. Nat Comm. 2018; 9(1):3258. https://doi.org/10.1038/s41467-018-05257-7
- **Leaché AD, Fujieta MK, Minin VN, Bouckaert RR.** Species delimitation using genome-wide SNP data. Syst Biol. 2014; 63(4):534–42. https://doi.org/10.1093/sysbio/syu018
- **Mago-Leccia F.** Electric fishes of the continental waters of America. Caracas: Fundacion para el Desarrollo de las Ciencias Fisicas, Matematicas y Naturales. 1994.
- **Maia CR, Alves-Gomes JA.** Utilização do código de barras de DNA na estimativa da diversidade de peixes elétricos do gênero *Microsternarchus* (Ostariophysi: Gymnotiformes) na bacia do rio Negro, Amazonas. In: Souza LAG, Castellón EL, editors. Desvendando as fronteiras do conhecimento na região Amazônica do Alto Rio Negro. Manaus: INPA; 2012. p.185–201.
- **Minh BG, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A *et al*.** IQ-TREE2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020; 37(5):1530–34. https://doi.org/10.1093/molbev/msaa015
- **Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P *et al*.** vegan: Community Ecology Package. R package version 2.7-1. 2025. Available from: https://CRAN.R-project.org/package=vegan
- **Paradis E, Schliep K.** ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019; 35(3):526–28. https://doi.org/10.1093/bioinformatics/bty633
- **Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE.** Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE. 2012; 7(5):e37135. https://doi.org/10.1371/journal.pone.0037135
- **Pinilla CE, Cox Fernandes C, Alves-Gomes JA.** *Microsternarchus javieri*, a new species of weakly electric fish (Gymnotiformes: Hypopomidae: Microsternarchini) from the Negro River basin, Amazonas, Brazil. Acta Amaz. 2025; 55:e55bc24175. http://dx.doi.org/10.1590/1809-4392202401751
- **Prates I, Hutchinson MN, Singhal S, Moritz C, Rabosky DL.** Notes from the taxonomic disaster zone: evolutionary drivers of intractable species boundaries in an Australian lizard clade (Scincidae: *Ctenotus*). Mol Ecol. 2024; 33(20):e17074. https://doi.org/10.1111/mec.17074
- **Puechmaille SJ.** The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol Ecol Resour. 2016; 16(3):608–27. https://doi.org/10.1111/1755-0998.12512
- **Puillandre N, Brouillet S, Achaz G.** ASAP: assemble species by automatic partitioning. Mol Ecol Resour. 2021; 21(2):609–20. https://doi.org/10.1111/1755-0998.13281
- **Puillandre N, Lambert A, Brouillet S, Achaz G.** ABGD, automatic barcode gap discovery for primary species delimitation. Mol Ecol Resour. 2012; 21(8):1864–77. https://doi.org/10.1111/j.1365-294X.2011.05239.x

- **Pyron RA, O'Connell KA, Duncan SC, Burbrink FT, Beamer DA.** Speciation hypotheses from phylogeographic delimitation yield an integrative taxonomy for seal salamanders (*Desmognathus monticola*). Syst Biol. 2023; 72(1):179–97. https://doi.org/10.1093/sysbio/syac065
- **R Development Core Team.** R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2025. Available from: https://www.R-project.org/
- **Salles MMA, Carvalho ALG, Leaché AD, Martinez N, Bauer F, Motte M *et al*.** Ancient introgression explains mitochondrial genome capture and mitonuclear discordance among South American collared *Tropidurus* lizards. Mol Ecol. 2025; 34(1):e70130. https://doi.org/10.1111/mec.70130
- **Sambrook J, Russel DW.** Purification of nucleic acids by extraction with phenol: chloroform. Cold Spring Harb Protoc. 2006; 2006(1):pdb.prot4455. https://doi.org/10.1101/pdb.prot4455
- **Schlick-Steiner BC, Steiner FM, Seifert B, Stauffer C, Christian E, Crozier RH.** Integrative taxonomy: a multisource approach to exploring biodiversity. Ann Rev Entomol. 2010; 55(1):421–38. https://doi.org/10.1146/annurev-ento-112408-085432
- **Sousa T, Wosiacki WB, Muriel-Cunha J, Prudente BS, Sousa RS, Peixoto LW.** A new species of the electric fish *Microsternarchus* Fernández-Yépez 1968 (Gymnotiformes: Hypopomidae) from the lower Amazon basin, Brazil. Zootaxa. 2024; 5448(4):508–18. https://doi.org/10.11646/zootaxa.5448.4.4
- **Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI *et al*.** Finding evolutionary processes hidden in cryptic species. Trends Ecol Evol. 2018; 33(3):153–63. https://doi.org/10.1016/j.tree.2017.11.007
- **Swofford DL.** PAUP* phylogenetic analysis using parsimony (* and other methods). Version 4. 2003. Available from: http://paup.csit.fsu.edu/
- **Tagliacollo VA, Tan M, Reis RE, Gaia R, Carrijo V, Ranuzi M *et al*.** Time-calibrated phylogeny of Neotropical fishes. Front Bioinform. 2024; 4(1433995). https://doi.org/10.3389/fbinf.2024.1433995
- **Toews DPL, Brelsford A.** The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 2012; 21(16):3907–30. https://doi.org/10.1111/j.1365-294X.2012.05664.x
- **Vences M, Miralles A, Brouillet S, Ducasse J, Fedosov A, Kharchev V *et al*.** iTaxoTools 0.1: Kickstarting a specimen-based software toolkit for taxonomists. Megataxa. 2021; 6:77–92. https://doi.org/10.11646/megataxa.6.2.1
- **Ward RD, Zemlak TS, Innes BH, Last PR, Herbert PDN.** DNA barcoding Australia's fish species. Philos Trans R Soc B. 2005; 360(1462):1847–57. https://doi.org/10.1098/rstb.2005.1716
- **Wickham H.** ggplot2: elegant graphics for data analysis. Cham: Springer international publishing; 2016.
- **Willis SC, Farias IP, Ortí G.** Testing mitochondrial capture and deep coalescence in Amazonian cichlid fishes (Cichlidae: *Cichla*). Evolution. 2014; 68(1):256–68. https://doi.org/10.1111/evo.12230
- **Yang Z, Rannala B.** Bayesian species delimitation using multilocus sequence data. PNAS. 2010; 107(20):9264–69. https://doi.org/10.1073/pnas.0913022107
- **Zhang J, Kapli P, Pavlidis P, Stamatakis A.** A general species delimitation method with applications to phylogenetic placements. Bioniformatics. 2013; 29(22):2869–76. https://doi.org/10.1093/bioinformatics/btt499

AUTHORS' CONTRIBUTION

Matthew A. Burnett: Conceptualization, Formal Analysis, Investigation, Visualization, Writing-original draft, Writing-review and editing.
Guillermo Ortí: Conceptualization, Data curation, Writing-review and editing.
José Alves-Gomes: Conceptualization, Data curation, Writing-review and editing.

FUNDING INFORMATION

JAG was funded by process #01.02.016301.000992/2024–94 from Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), and by Internal Project #12307 of Instituto Nacional de Pesquisas do Amazonas (INPA). GO was funded by the George Washington University. During work on this project, MAB was supported by the Harlan Endowment to the Department of Biological Sciences at the George Washington University.

ETHICAL STATEMENT

All the animals used in this study were handled and processed under the rules and principles established by INPA's Comissão de Ética de Uso de Animais (CEUA). Capture, transport, maintenance in captivity, and experiments were done under IBAMA permits 14833–1 and 14834–1, and INPA/CEUA 036/2012 to JAG.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available on NCBI under BioProject PRJNA1460325. Accession numbers for ddRAD and COI data are provided in Tab. 2.

AI STATEMENT

ChatGPT was used in the editing and troubleshooting of R scripts.

COMPETING INTERESTS

The authors declare no competing interests.

SUPPLEMENTARY MATERIAL

Supplementary Material File

PEER REVIEW REPORT

Reviewers:
Anonymous reviewer #1
Anonymous reviewer #2
The peer review report is available at: Neotropical Ichthyology

HOW TO CITE THIS ARTICLE

- **Burnett MA, Ortí G, Alves-Gomes J.** New insights into the genetic diversity of *Microsternarchus* (Gymnotiformes: Hypopomidae) in the rio Negro using genomic markers. Neotrop Ichthyol. 2026; 24(4):e260014. https://doi.org/10.1590/1982-0224-2026-0014

Neotropical Ichthyology

OPEN ACCESS

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Distributed under
Creative Commons CC-BY 4.0

© 2026 The Authors.
Published by SBI

Official Journal of the
Sociedade Brasileira de Ictiologia
SBI