1-1-2019

Diversification In The Neotropics – Evolution And Population Genetics Of The Armored Catfish Hypancistrus Sp. From The Xingu River

Marcella Goncalves Santos

DIVERSIFICATION IN THE NEOTROPICS – EVOLUTION AND POPULATION GENETICS OF THE ARMORED CATFISH *HYPANCISTRUS* SP. FROM THE XINGU RIVER

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Biology
The University of Mississippi

By

MARCELLA GONÇALVES SANTOS

December 2019

ABSTRACT

The Xingu River, one of the largest tributaries of the Amazon River, is currently in peril due to the recent construction of hydroelectric dams, but little is known about the numerous fish species it supports. This dissertation focuses on three pleco catfish species belonging to the genus *Hypancistrus* from the Xingu River with partially overlapping distributions: *H. zebra*, *H.* sp. (L174), and *H.* sp. (L66/333). Chapter 1 is a bibliographic review of Amazonian freshwater fish diversity, with the goal of discussing the hypotheses of speciation mechanisms that can be tested in this system, including the relative importance of ecological adaptation and vicariance caused by topographical divides and waterfalls and rapids, and arguing this is an important overlooked model for the study of speciation processes. The goal of Chapter 2 was to use genomic data to unravel the basic relationships among eight described and eleven undescribed species belonging to the genus *Hypancistrus* distributed across the Orinoco and Amazon Basins. The phylogenetic analyses support the existence of two clades corresponding to each basin, but relationships among some of the species are poorly supported. Further exploratory analyses in combination of hypotheses testing indicate there are at least four admixed lineages in the Amazon clade. Chapter 3 investigated the evolution of *Hypancistrus* from the Xingu River based on genomic data. With dense sampling of *H.* sp. (L66/333), phylogenetic and population genetic analyses reveal a gradient of genetic structure along the river, with introgression from lineages of *Hypancistrus* from other Amazon River tributaries close to the mouth of the Xingu. On the upstream limit of the distribution of *H.* sp. (L66/333), a population hybridized with *H.* sp. (L174) is found just upstream of waterfalls, that act as a partial barrier to gene flow. Tests for past gene

flow suggest there is signal for multiple introgression events between these lineages, but the direction, timing, and intensity of these events is still unclear. Overall, these results indicate the evolution of *Hypancistrus* was exceptionally complex. Fascinating patterns of diversification are emerging from this system that is unfortunately in risk of extinction due to the impacts of damming.

DEDICATION

For Nico.

ACKNOWLEDGMENTS

This dissertation is the result of the combined efforts of numerous people and institutions. It all started with my desire to pursue a scientific carrier, aligned with the initiative of the Brazilian government to fund thousands of young students and researchers to obtain their degrees abroad, with the Science Without Borders program. Through this program, the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq) funded me for four years and four months of tuition, stipends, and lab supplies, for which I am grateful. I also had financial support from the University of Mississippi's Department of Biology, Graduate School, and Biology Graduate Students Society. Fieldwork in the Xingu in 2014 was supported by the iXingu Project (US National Science Foundation DEB 1257813) which also provided many of the samples used in this study. Furthermore, travel to the Academy of Natural Sciences of Philadelphia in 2017 was supported by the Böhlke Memorial Endowment Fund. I strongly believe in the power of science and education in changing society, and I strive to give back all the investment that was trusted in me.

I thank my advisor Dr. Brice Noonan for his support and friendship. Through his example I learned how to be a better scientist, to trust myself, to be ambitious, and also kind, compassionate, and understanding. He tought me important lessons about the disciplines of speciation and evolution, but mostly importantly, about leadership and how to communicate with students, other researchers, and the general public. Additionally, I would like to thank the members of my Committee: Dr. Leandro Sousa, Dr. Susan Balenger, Dr. Sarah Liljegren, and

Dr. Louis Zachos. I also thank Dr. Rebecca Symula who was temporarily part of my Committee. Their support and contribution was fundamental to this dissertation.

I was incredibly lucky to partner with Dr. Leandro Sousa, who co-advised me and kindly shared his expertise and enthusiasm for the Xingu River and its fishes. His involvement, along with the support of the leader of the iXingu Project, Dr. Mark Sabaj, was crucial to make my project a reality. I sincerely thank them for welcoming me in their field expedition in the Xingu and in my visit to Philadelphia, and for the lengthy discussions that enlightened the understanding of my findings throughout this entire project. In addition, I would like to especially acknowledge Dr. Nathan Lujan for introducing me to the iXingu Project and to the community of Neotropical ichthyologists. There were many people involved in the field expeditions in the Xingu River, including support staff, fishermen, and researchers of various areas, and I appreciate their involvement and dedication. Finally, this dissertation would not have been the same without the tissue samples donated by the aquarium hobby. I am grateful for Haakon Haagensen's enthusiasm in sharing these samples and photographing the fishes.

One of the most challenging parts of this work was to learn how to implement and interpret analyses. I would like to thank the people who were willing to share their experience with me, including Dr. Ísis Arantes, Dr. Mariana Vasconcellos, Dr. Fabricius Domingos, Chaz Hyseni, Dr. Ryan Garrick, João Paulo Marum, and Ben Phar. I am also thankful for the friends I made in the Noonan Lab, the Biology Department, and the Brazilians in Oxford, as well as the friends that stayed back home, but continue to cheer on me and are such an important part of my life.

Lastly, I thank my family for their unconditional love and support. My parents always encouraged me to pursue my dreams and have been my bedrock even from far. I thank Dr. Renan Bosque, my partner in science and in life, whose dedication and generosity continue to inspire me and push me forward. And I thank my son Nico, for bringing joy to our lifes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: AMAZONIAN FISHES AS MODELS FOR DIVERSIFICATION AND SPECIATION STUDIES

INTRODUCTION

Identifying and describing the mechanisms underlying the origin of new species remains one of the central questions in evolutionary biology (Butlin *et al.*, 2012). Given the complexity of biological phenomena, experimental approaches to speciation research have been limited to a few model organisms (Fry, 2009). With the revolution in molecular technologies, the discipline has seen great advancements towards identifying proximal causes for the origin of isolation between species in natural populations. These molecular techniques have allowed biologists to trace the signature of evolutionary change in DNA, RNA, and proteins of organisms in nature, as well as uncovering the underlying biochemical pathways of gene expression and the association between genotype, phenotype and fitness (Byers *et al.*, 2017). As we move towards gathering increasing amounts of empirical evidence from nature, we are revealing an unexpected variety of mechanisms and patterns of diversification. However, much remains to be done to shed light on the details of species origination, such as the roles of gene flow, natural and sexual selection, hybridization, and reinforcement (Butlin *et al.*, 2012). Identifying the natural laboratories where hypotheses can be tested harbors great potential for enhancing our understanding of evolution.

The goal for this chapter is to argue that Amazonian freshwater fishes are an understudied, important model system for speciation research. Herein we describe the highly diverse biota of Amazonian freshwater bodies and review published hypotheses about

evolutionary processes of diversification as well as biotic patterns in Amazonian fishes. Given this, we highlight areas of study that are ripe for exploration and propose tests and predictions that may contribute to our understanding of speciation mechanisms. Each of these aspects will be considered in the context of the unique qualities of Amazonian aquatic systems. While we will address and highlight the unique diversity and patterns present in this region that are attractive for the study of evolutionary patterns, we also highlight the urgent need for this work to commence immediately. The combination of threat and biological complexity offer a rapidly closing window for research.

Definitions

For this review we generally follow the nomenclature adopted by Albert & Reis (2011a). A basin is broadly defined as a lowland area surrounded by higher ground. We define a hydrogeographic basin as the drainage area of major rivers and their tributaries (e.g. Amazon Basin, Orinoco Basin, Paraná-Paraguay Basin). We reserve the term sub-basin for the drainage area of tributaries of the major basins (e.g. Negro sub-basin, Madeira sub-basin, Xingu sub-basin). We interchangeably use the terms drainage and catchment, defined as the area that drains into a single stream or river, at any hierarchical level (tributaries or major rivers). A divide is defined as the ridge or topographical division between drainages, at any hierarchical level.

Justification

The Amazon Basin harbors the largest biodiversity of freshwater fishes in the world, with more than 2,700 species (Dagosta & Pinna, 2019). This is approximately 15% of the world's total freshwater fish biodiversity. Within this area, Amazonian fish assemblages are structured

by, among other things, water chemistry, flood regimes, and distance to large rivers (Henderson & Crampton, 2009; Benone *et al.*, 2018; Stegmann *et al.*, 2019). Multiple patterns of biogeographic distribution are found among Amazonian fish lineages, evidencing the complexity of processes acting throughout their evolutionary history (Dagosta & Pinna, 2019). Despite the fact that the Amazon contains the largest freshwater network in the world and supports an extraordinary diversity of fishes, surprisingly little is known about diversification processes within this complex system (Albert & Reis, 2011a).

Neotropical freshwater systems currently face innumerable threats due to human activity. The main conservation problem in the region is related to habitat loss, caused by deforestation, pollution, damming, mining, and agricultural activities (Barletta *et al.*, 2010; Anderson *et al.*, 2018; Arantes *et al.*, 2018; Andrade *et al.*, 2019). Introduction of invasive species (Britton & Orsi, 2012) and fishing exploitation of native species for meat consumption or fish keeping purposes (Moreau & Coomes, 2007) have also caused enormous impacts and declines in population sizes. Moreover, the effects climate change in these ecosystems are yet to be elucidated, but a notable decrease in dissolved oxygen is expected, with profound effects in water chemistry (Britton & Orsi, 2012; Frederico *et al.*, 2016).

Among the numerous river courses in this region, those rivers that flow through steep slopes with rapids and waterfalls are particularly susceptible to anthropogenic modification in the form of impoundments given that these geological characteristics are targeted by constructors of hydroelectric power stations (e.g. Sabaj-Pérez, 2015; Alter *et al.*, 2017; Anderson *et al.*, 2018). The impacts include not only discontinuity of natural flow (Pelicice *et al.*, 2015), but also flooding above the dam accompanied by drastic changes in water quality and substrate, plus dewatering below the dam (Pringle *et al.*, 2000; Sawakuchi *et al.*, 2015; Fitzgerald *et al.*, 2018).

This so-called clean energy source has also been shown to release greenhouse-effect gases due to methane emissions from the decay of vegetation biomass in flooded reservoirs (Fearnside, 2009).

These imminent threats to the Amazonian freshwater ecosystems contrast with poor conservation efforts to minimize the impacts of human activities. The Amazon region has been at the center of conservation debate due to recent increases in deforestation rates after more than a decade of historically low deforestation (terrabrasilis.dpi.inpe.br, Freitas *et al.*, 2018). This is leading to concerns that we might be reaching a tipping point after which the ecosystem might become unsustainable, as the hydrological cycle in the Amazon will be severely affected (Lovejoy & Nobre, 2018). It is urgent to establish management strategies based on scientific evidence to make sure the Amazonian fish diversity is protected across all levels, along with the processes that maintain it (Castello *et al.*, 2013; Vitule *et al.*, 2017).

PAST FORMATION OF THE AMAZON BASIN

The Amazon Basin is a drainage system that encompasses approximately 7 million km^2 (Fig. 1, Bloom & Lovejoy, 2011). Extending from latitudes 10ºN to 15ºS, it is located in the tropics and largely covered by rainforest. The predominantly warm and wet climate of the Amazon region is marked by the alternation of wet and dry seasons. The distribution of moisture and seasonality patterns is highly influenced by the presence of the Andes (Vonhof & Kaandorp, 2009) from which much of the Basin's water flows. In this enormous region the South American freshwater biota has evolved in near isolation for the past 100 million years (Ma) after the breakup of western Gondwana and complete separation of South America from Africa in the Upper Cretaceous (Lundberg *et al.*, 1998; Maisey, 2000). Since that time, the uplift of the Andes, marine transgression and regression cycles, and headwater capture events have established the

modern South American drainage axes: the Amazon, Orinoco and Paraná-Paraguay Basins (Albert & Reis, 2011b). These river systems have been constrained by the persistent Precambrian (>540 Ma) Brazilian and Guiana Shields. Between the Brazilian and Guiana Shields lies the Amazon-Orinoco lowlands, a large sedimentary basin that was the stage for major changes in flow directions of the Amazon and Orinoco Basins over the past 100 Ma. During the Cretaceous and Paleogene (100-23 Ma) the proto-Amazon river comprised a west-flowing river and an east-flowing river (Hoorn *et al.*, 2010). During the same period, the divide between the proto-Amazon and the Paraná Basins was established due to the bending of the Bolivian orocline 30 Ma (Lundberg *et al.*, 1998).

The Nazca and Pacific Plate subduction along the Pacific coast of the South American Plate led to the uplift of the Andean Mountains in pulses, starting during the Cretaceous (125-112 Ma) but reaching its climax much later during the Miocene and Pliocene (10-4 Ma, Hoorn & Wesselingh, 2010). These orogenic events formed the Sub-Andean Forelands, a series of depressions to the east of the Andes divided by arches that have held sediment basins, lakes, and marine transgressions (Lundberg *et al.*, 1998). As a consequence, the uplift of the Vaupes Arch (10 Ma) was responsible for dividing the Amazon and Orinoco Basins (Albert & Reis, 2011b). Another important consequence of the Andean uplift was the onset of the modern transcontinental Amazon River's eastward flow, that started ~9 Ma but continued to expand through capture events up to 4.5 Ma (Hoorn *et al.*, 2017; Albert *et al.*, 2018). However, some authors challenge this view, arguing that the Amazon system only formed after the disappearance of the Pebas system of mega-lakes, which they suggest persisted until 2.5 Ma (Campbell *et al.*, 2006).

These geological past events along with the marine transgressions, biotic exchange due to closure of the Panamá Isthmus, and aquatic habitat diversity are the main processes believed responsible for structuring the major biogeographic patterns of freshwater organisms in South America (Albert & Reis, 2011b; Oberdorff *et al.*, 2019). However, we propose that at finer scales, the varying levels of complexity of Amazonian riverine systems and the diversity of lineages inhabiting this region present an opportunity to explore inter and intraspecific processes that are likely to provide insight into the origins of the diverse biota of Neotropical freshwater ecosystems.

HYPOTHESES

Freshwater fishes have provided many fascinating examples of speciation mechanisms acting on a local scale, such as the adaptive radiation of African cichlids (Salzburger & Meyer, 2004; Seehausen, 2006) and the use of the three-spined stickleback as a model to study genomics, behavior, and parasitism (Gibson, 2005; Barber, 2010). The uniqueness of directional water flow and hierarchical network structure in freshwater systems, unlike terrestrial and marine systems, calls for ecological and population genetic theories that account for the spatial constraints on dispersal of river inhabitants (Fausch *et al.*, 2002). Freshwater systems are organized in three discrete scales: continental, interbasin, and within basin (Rahel, 2007). The broader continental scale comprises Wallace's zoogeographic regions that are divided by seas, oceans, mountain ranges, and deserts (see also Balian *et al.*, 2008). Within continents, basins defined by topographical characteristics constitute the second scale, and sub-basins are found within basins. To reach a different basin or sub-basin, an organism must transpose topographical divides (i.e. stream capture) or disperse through salt water along coastlines (i.e. river mouth-to-

mouth). In addition to these geographical restrictions in reaching a different drainage, differences in water quality among river courses can be significant barriers to the movement of fishes between drainages. Finally, within the course of a river, the most important barriers to dispersal and gene flow are waterfalls and cascades, as well as anthropogenic modifications that disrupt continuity (i.e. reservoirs and associated dams). Long stretches of rapids may also act as ecological filters, as they may pose low availability of food and shelter and require high specialization to overcome the strong turbulence of water (Torrente-Vilara *et al.*, 2011). Inter and intra-specific interactions, as well as environmental variables are also important factors in delimiting the distribution of organisms.

Herein, we detail hypotheses for patterns of diversity and speciation mechanisms of Amazonian fishes considering the hierarchical level in which they are acting (Table 1). These are factors that we propose are important in causing reproductive isolation and diversification of fishes in the Amazon Basin. Such mechanisms certainly act simultaneously, and the use of integrative approaches with explicit hypothesis testing is essential to tease apart the relative roles of each mechanism.

We particularly highlight the role of ecological adaptation as the common thread across all hierarchical levels and hypotheses we describe. With rare exceptions of speciation by sexual selection, drift, and polyploid speciation, ecological adaptation is a fundamental part of the speciation process (Sobel *et al.*, 2010). To address this matter, biologists may use various methods like systematic, phylogeographic, population genetic, ecological, morphological, physiological, and behavioral studies. Furthermore, the integrative approach of landscape genetics has been shown to be a powerful tool at the local scale, detecting environmental filter effects on gene flow among populations (Manel *et al.*, 2003; Storfer *et al.*, 2006). The

combination of landscape genetic analytical methods with an extensive genomic dataset represents a solid foundation for a powerful test of questions about the role of a freshwater systems' environmental variables on fish population structure, helping elucidate the relative importance of ecological adaptation as a speciation mechanism (Grummer *et al.*, 2019). We emphasize the necessity to focus on the adaptive potential to environmental conditions expected with climate change, like tolerance to higher water temperature and lower dissolved oxygen levels (Frederico *et al.*, 2016).

Table 1. Hypotheses of patterns of diversity and speciation mechanisms for fishes in the Amazon Basin.

Hierarchical level	**Hypothesis**	**Process**
All levels	Ecological adaptation	Natural selection
Between drainages	Topographical divides	Vicariance and dispersal Capture events
	Marine transgressions	Vicariance of freshwater species in highlands Adaptation from saltwater to freshwater
Between/within river	Water color and chemistry	Physiological and morphological adaptation to different water chemistries
Within river	Waterfalls as barriers	Vicariance
	Isolation in rapids	Vicariance of lineages adapted to extreme rapids environment
	Downstream increase in genetic diversity	Directional gene flow Habitat availability Upstream colonization
	Seasonal structure	Isolation by time
	Biotic factors	Sexual selection Competition Predation

Topographical divides

On the largest level of the hierarchical structure of freshwater systems, topography is the major constraint to the distribution of obligatory freshwater organisms. During uplift or erosion

events, patterns of drainage change, and part of a river flow can be diverted joining a stream of an adjacent drainage system, in a river capture event (Bishop, 1995). Therefore populations might go through both dispersal, when reaching a previously inaccessible river, and vicariance, when its original distribution is divided. Genetic signatures of river capture events might show up when populations are more closely related between than within rivers (Burridge *et al.*, 2006). There is abundant evidence that river capture events are important contributors to promoting diversification of freshwater fauna (e.g. Waters *et al.*, 2001; Cardoso & Montoya-Burgos, 2009). In the Amazon Basin, the relatively flat landscapes with low elevational variance of the South American platform favor the occurrence of these phenomena, since small geological movements might drastically alter river courses (Albert & Reis, 2011b). Besides evidence of genetic structuring coincident with these geographic barriers, timing of diversification of lineages coupled with independent geological evidence for river course shifts is a crucial component in testing such hypotheses. This is currently an important limitation in Amazonian freshwater systems, since the literature detailing the formation of Amazon tributaries is scarce, and fossil records are rare for most fish lineages (Lovejoy *et al.*, 2010).

In terms of ichthyofaunal composition, the closest drainage to the Amazon Basin is the Orinoco Basin (Albert & Carvalho, 2011). Therefore, the many fish lineages with distributions spanning both basins are valuable models to investigate the role of topographical divides in promoting diversification. For the majority of such lineages, the division between the basins represents a barrier to gene flow, as in the case of the speckled peacock bass *Cichla temensis* (Willis *et al.*, 2015). However, there is evidence for past connections between these basins over the past 10 Ma via Guiana Shield rivers in the needlefish *Potamorrhaphis guianensis* (Lovejoy & de Araújo, 2000), and the seasonally flooded area that connects the Amazon and Esequibo Rivers

(Rupununi portal) in the pacu *Piaractus brachypomus* (Escobar *et al.*, 2015) that clearly support a deep history of biotic exchange. More recently, the diversion of part of the Orinoco headwaters to the Negro River created a new route of connection between the two basins in the past 10 thousand years (kya), although the dramatic differences in water quality still restrict the movement of many species of fishes (Willis *et al.*, 2010; Stokes *et al.*, 2018).

Marine transgressions

Past fluctuations in ocean levels coupled with the low elevation of South America led to multiple events of marine transgression throughout the Cenozoic, including the Late Paleogene (30-23 Ma), the Neogene (23-2.6 Ma), up to the climatic fluctuations in the Peistocene in the past 2.6 Ma (Lundberg *et al.*, 1998; Albert *et al.*, 2018). Saltwater wetlands and seas that invaded the lowlands presumably impeded the access of freshwater species to these areas, isolating such lineages in disjunct tributary headwaters (Lovejoy *et al.*, 2010). This leads to the prediction that freshwater lineages are older in highlands than in lowlands, the expectation of a genetic signature of population expansion from highlands to lowlands, and finally reciprocal monophyly among highland species from different areas, particularly the Andes, Brazilian Shield, and Guiana Shield, in a scenario analogous to the refugia hypothesis (Bloom & Lovejoy, 2011). Such cycles of intermittent periods of isolation and past connection favor divergence through adaptation and drift, followed by secondary contact with hybridization. Recent genomic evidence is highlighting the important role of hybridization in creating new combinations of old genetic variation to generate adaptive radiations and rapid speciation (Marques *et al.*, 2019). The complex dynamic of these highland refugia constitute an exciting and unexplored area for the study of hybridization and adaptive radiation.

On the other hand, marine transgressions have also contributed to freshwater biodiversity through the adaptation of marine lineages to freshwater systems (Bloom & Lovejoy, 2017). Molecular dating of diversification of croakers (*Plagioscion*) support the hypothesis of invasion of the Amazon Basin via marine transgressions in the early Miocene, with subsequent adaptation and speciation in drainages with different water colors/chemistry (Cooke *et al.*, 2012). However, the molecular mechanisms including which genes are responsible for the physiological adaptation to life in freshwaters remain poorly understood. Other taxa of presumed marine origin that are present in the Amazon Basin and merit further investigation to test include needlefish, stingrays, puffers, and soles, among others (Bloom & Lovejoy, 2017).

Water color and chemistry

In the Amazon Basin, rivers are classified as possessing one of three types of water color based on physical-chemical properties that are determined by the location of the river's source, type of soil along its course, and dominant vegetation cover (Albert & Reis, 2011b). White water is the result of high sediment and nutrient load draining from the Andes, as in the Amazon and Madeira Rivers. Such white water rivers have a neutral pH, and high electric conductivity. Black water, found for example in the Negro and Tefé Rivers, contains low sediments and nutrients, dark color from the high tannin content, low pH, low electric conductivity, intermediate transparency, and originates in the forested lowlands. Clear water rivers drain from ancient and well weathered crystalline rocks of the Brazilian and Guiana Shields, have high transparency because they carry low sediments, and have intermediate electric conductivity. The largest clear water tributaries of the Amazon include the Xingu and Tocantins Rivers. Water color is a particularly relevant ecological barrier for fish species, since it requires certain physiological

adaptations. Genetic structuring associated with water color has been documented in several species, for example, in the marbled hatchetfish *Carnegiella strigata* (Schneider *et al.*, 2012), the curimata *Prochilodus nigricans* and the tambaqui *Colossoma macropomum* (Ardura *et al.*, 2013), and the croaker *Plagioscion squamosissimus* (Cooke *et al.*, 2012). We hypothesize this to be one of the most important mechanisms driving speciation and isolation in the Amazon freshwater fishes.

In addition to the limitations imposed by the chemical differences of these water types, the color/clarity of the water also imposes differing selection pressures on resident species. In a recent study, Pires *et al.* (2019) found evidence of the combined effect of natural and sexual selection causing divergent adaptation in sexually dimorphic tetras in forested stream habitats. The authors suggest the primary driver of this divergence is the lighting condition in clearwater and blackwater streams. Unlike the transparent clearwater, blackwater filters out shorter wavelengths, tending to be infiltrated by diffuse red light. As a result of natural selection to promote better vision, fish from blackwater lineages have larger eyes than fish from clearwater lineages. On the other hand, sexual selection dependent on visual perception resulted in more conspicuous color patterns in fins in the clearwater lineages. This study highlights the potential of Amazonian fishes as models to study the origins of reproductive barriers, and the multiple mechanisms promoting diversification in the Amazon Basin.

Waterfalls as barriers

Few studies have focused on the role of waterfalls and cascades, as well as rapids, and whether they represent biogeographic barriers capable of promoting reproductive isolation at the finest of geographic scales. However, species occurrence data that identify waterfalls and rapids

as defining the limits of species distributions suggest these features likely play a role in intraspecific genetic continuity. For example, on the Madeira River of the Amazon Basin, fish assemblage surveys have identified waterfalls as coincident with major breaks in species composition (Torrente-Vilara *et al.*, 2011). However, other studies make it clear that waterfalls/rapids do not pose absolute barriers to dispersal, as significant levels of migration and gene flow have been observed downstream of waterfalls in the floodplain fish *Colossoma macropomum* (Farias *et al.*, 2010) and in the catfish *Brachyplatystoma platynemum* (Ochoa *et al.*, 2015), both in the Madeira River.

There are a few examples demonstrating the action of rapids and waterfalls as barriers to gene flow in the Amazon Basin. High genetic structure was observed in the brown pencilfish (*Nannostomus eques*) from the black water Negro River, in the Amazon Basin (Terencio *et al.*, 2012). While the main cause for population isolation in this pencilfish is the water color barrier at the confluence of the middle Negro River with the white water Branco River, rapids and waterfalls in the upper Negro River also correlated with observed patterns of genetic structure, implying barriers to gene flow (Terencio *et al.*, 2012). Additionally, the Madeira River Rapids were found to explain the high genetic structure in the black flannelmouth characin (*Prochilodus nigricans*) within this river (Machado *et al.*, 2017). Similarly, remarkable rapid stretches are found in the Xingu and Tapajós Rivers, tributaries of the Amazon River draining from the Brazilian Shield, yet no studies have investigated intraspecific genetic structure in these rivers.

In other river basins, population genetic, phylogeographic and phylogenetic studies are highlighting the role of waterfalls in restricting gene flow and promoting diversification. In a remarkable case, waterfalls were shown to drive parallel evolution in a fish complex in Iriomote Island, Japan. *Rhinogobius* sp. YB evolved independently in 11 rivers isolated upstream from the

parent species *R. brunneus* by waterfall formation (Kano *et al.*, 2012). High-energy rapids in the Lower Congo River are also driving strong differentiation among populations of African cichlids (Markert *et al.*, 2010; Schwarzer *et al.*, 2011). Within this system, Schwarzer *et al.* (2012) found evidence for complex reticulated evolutionary history with multiple hybridization events contributing to speciation in the *Steatocranus* group. Genomic sequence data revealed cryptic microallopatric divergence in *Teleogramma* spp., a small rapids specialized group of cichlids with extraordinary genetic structure driven by the Congo River rapids (Alter *et al.*, 2017). Waterfalls and rapids were also found to restrict gene flow among populations of cichlids in Cuanza and Okavango-Zambezi River systems from central Angola (Musilova *et al.*, 2013), and of the steelhead from Klickitat River, US (Narum *et al.*, 2006). Other forces may act in combination with waterfalls as barriers to dispersal in rivers, such as predation, sexual selection, and natural selection as evidenced by the Trinidadian guppies *Poecilia reticulata* (Labonne & Hendry, 2010).

Isolation in rapids

Rapids may not only act as barriers among populations in neighboring areas, but also as unique islands themselves where speciation via selection may occur. Organisms that live in rapid water flow and rocky substrates require specializations in order to survive, specializations that may be maladaptive outside these unique environments. Species that are limited to such fast water flow environments are called rheophilic. Fast-flow environments select for characteristics in locomotor morphology in fish that increase hydrodynamics and reduce drag (Blob *et al.*, 2008). Such adaptations may restrict these organisms to the patchy stretches of rivers with rapids, leading to isolation among these areas even within a single river. Many Amazonian

freshwater fishes are adapted to rapids habitats, but evidence for structuring within a river remains scarce (Lujan & Conway, 2015). Although draining to the Atlantic, the Araguaia River shares many characteristics with the Tapajós, Xingu, and Tocantins Rivers, flowing from the Brazilian Shield to the North, with stretches of rapids as those rivers leave the Brazilian Shield and enter in the Amazon basin. In one of the few studies to examine population structure associated with patchily distributed rapids in rheophilic species Hrbek *et al.*(2018) found strong genetic structuring in rheophilic fishes in the Araguaia River.

Downstream increase in genetic diversity

One prominent within-species spatial pattern typical of river systems is the downstream increase in genetic diversity (DIGD, Paz-Vinas *et al.*, 2015). S originally described this pattern may be explained by a series of non-exclusive processes: downstream-biased gene flow, due to unidirectional water flow biasing the direction of migration; variation in habitat availability, due to increased river width leading to increased effective population size near the river's mouth; and upstream-directed colonization, due to a series of bottleneck effects upstream, assuming the founding populations are closer to the river's mouth (Paz-Vinas *et al.*, 2015). These four processes can be differentiated with model-choice methods that are increasingly common in analyses of molecular datasets (Beaumont *et al.*, 2002). No studies have directly tested this hypothesis for Amazonian fishes. However, it is worth noting that two published studies have reported that taxonomic and functional diversity is higher in fish assemblages from the headwaters of Amazonian tributaries, which is explained by the past stability and lack of connectivity among rivers of the Guiana and Brazilian Shields (Oberdorff *et al.*, 2019; Stegmann *et al.*, 2019).

Seasonal structure

An interesting pattern observed in some migratory fish species is the lack of spatial structure but presence of temporal structure also known as isolation by time. This pattern is generated by genetically divergent populations using the same stretch of river in different reproductive seasons. Seasonal structure has not been demonstrated for Amazon fishes yet, but it was found in other South-American basins as in the case of *Prochilodus costatus* from the São Francisco River (Braga-Silva & Galetti, 2016) and in *Salminus brasiliensis* from the Uruguay River (Ribolli *et al.*, 2017). Intense seasonality is one of the most remarkable characteristics of Amazonian freshwater systems (Albert & Reis, 2011b), and though this diversifying mechanism has yet to be tested it would not be surprising to find species that follow this pattern.

Biotic factors

Inter and intra-specific interactions like predation, sexual selection, and competition are possible mechanisms driving ecological or sympatric speciation. Some freshwater fish groups are among the most important model systems to study this process, including African cichlids, and three-spined sticklebacks (Bernardi, 2013). Beheregaray *et al.* (2015) reviewed the existing body of research addressing ecological speciation in tropical systems. The authors concluded that this topic has been only superficially explored and proposed an integrative methodological framework to identify ecological speciation in this region. Broadly, this framework includes the collection of environmental, phenotypic, and molecular data, including genotypes and transcriptomes, in phylogenetic, phylogeographic, polupation genetic, and landscape genetic approaches. The last fundamental step of this proposed framework consists of creating models to

simulate selection and test the evidence for ecological speciation and elucidating reproductive isolation, selection gradients, and hybrid zones (Beheregaray *et al.*, 2015).

Beheregaray *et al.* (2015) demonstrate this framework with Amazonian fish examples. The electric fish *Steatogenys elegans* occurs in the Amazon and Orinoco lowlands, and in the Guiana Shield. Cooke (2014) sampled this species across the ecotone where the blackwater Negro River and the whitewater Amazon River meet. The authors collected mitochondrial sequence and genomic data to produce phylogenetic analyses, test for signatures of selection, and describe genetic structuring across the water type environments. Finally, they simulated riverscape genetics under scenarios of varying degrees of water color selection strength. Under this framework, this study found evidence for the presence of two sympatric cryptic species of the electric fish that diverged by ecological speciation. The differences in water chemistry affect the transmission of electric signals for intraspecific communication, which is important in mating. Therefore, this study was able to link assortative mating across the gradient of water types leading to speciation (Cooke *et al.*, 2014; Beheregaray *et al.*, 2015).

CONCLUSION

Amazonian freshwater habitats hold an extraordinary and unmatchable number of fish species that are threatened by the severe impacts of various human activities. Future studies should focus on how local processes contribute to the generation of one of the most diverse biotas on Earth. Understanding how populations are structured, how genes are exchanged between populations, and the role of river landscapes in shaping these patterns is extremely important and will represent a breakthrough in the knowledge of evolution in Amazonian freshwater systems. Furthermore, the proposed studies poise these Amazonian systems to serve

as excellent models to enlighten many poorly understood aspects of the speciation theory, and tropical freshwater diversification in particular, since past studies in the region have been primarily focused on past events.

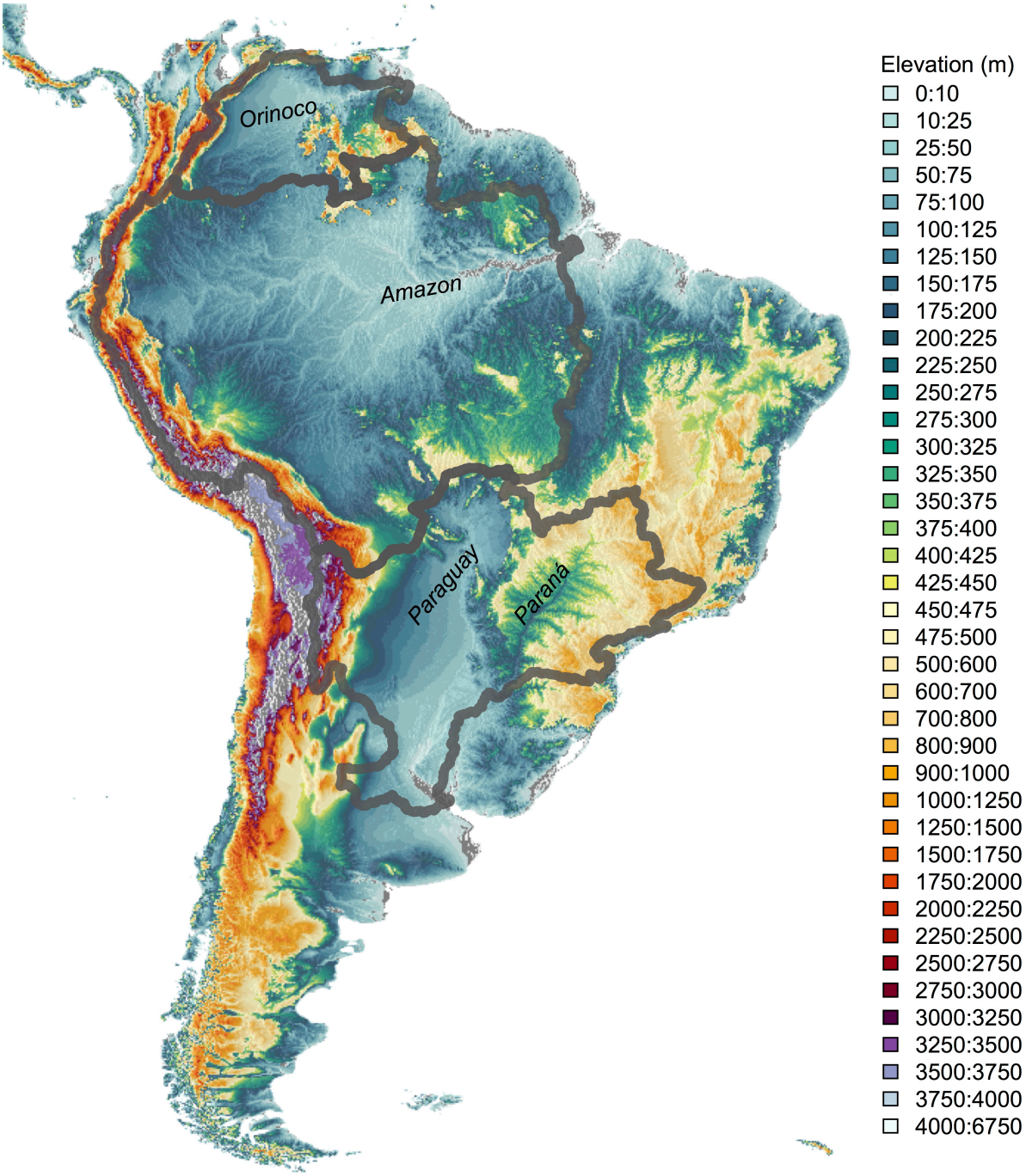


Figure 1. Elevation map of South America. Contoured areas correspond to Orinoco, Amazon, and Paraná-Paraguay Basins.

CHAPTER 2: DIVERSIFICATION OF *HYPANCISTRUS* CATFISHES IN THE AMAZON AND ORINOCO BASINS

INTRODUCTION

The Amazonian forest is one of the most remarkable examples of a highly diverse ecosystem that is suffering intense impacts caused by human activities (Hoorn & Wesselingh, 2010). Science still lacks basic knowledge of species inhabiting the Amazonian forest, their evolutionary history, ecology, physiology, and demography, which reduce the effectiveness of conservation actions. Notably, the region harbors the highest diversity of freshwater fishes. But even this incredibly high number of described species is low in relation to estimates of the total number of Amazonian fish species, as evidenced by the number of new species descriptions that are published every year. This lack of knowledge is due, in part, to the existence of cryptic diversity, but is primarily due to the lack of scientific, monetary, and human resources required to rectify the issue.

The high diversity of Amazonian fishes has its origins in the context of the geological history of the South American continent. For the past 100 million years (Ma), after isolation from Africa, the main modern South-American drainage axes were established by the orogenesis of arches in response to the formation of the Andes, the history of marine incursions, and headwater capture events (Albert & Reis, 2011b). The Amazon and Orinoco Basins, six and one million km^2 respectively, are the center of diversity of the extraordinarily rich Neotropical freshwater ichthyofauna (Lovejoy *et al.*, 2010). The modern configuration of these river systems

was established around 10-9 Ma, with the formation of the geomorphological division between the Amazon (east flow) and Orinoco (north flow) Rivers due to the uplift of the Vaupes Arch (Albert & Reis, 2011b; Hoorn *et al.*, 2017). However, a connection remains between these two distinct basins in the headwaters of the Orinoco River and the Negro River (a tributary of the Amazon) through the Casiquiare River (Winemiller & Willis, 2011). There are several examples of fish groups that span the two basins, but it is common to see a pattern of restricted gene flow within species, or the differentiation of evolutionary lineages correspond to the division between the rivers (Willis *et al.*, 2015).

Hypancystrus is a pleco catfish genus belonging to the Loricariidae, a South American family that is distributed across the Amazon and Orinoco. The foundation for the original description of the genus was a set of external morphological characters that were, surprisingly, not diagnostic in relation to other loricariids (Isbrücker & Nijssen, 1991). Armbruster (2002) addressed the matter and redefined the genus based on two unique synapomorphies of skull morphology: the presence of a wide separation between the metapterygoid and the lateral ethmoid and the presence of an angled adductor palatini crest of the hyomandibula. *Hypancistrus* are specialized for living in rocky outcrops and are thought to have omnivorous diets (Armbruster *et al.*, 2007). While known to be widespread across South America, and believed to be quite diverse, there are only nine described species. The genus was named with the description of *Hypancistrus zebra*, from the Xingu River, a southeastern tributary of the Amazon River (Isbrücker & Nijssen, 1991). Subsequently, five other species from the Orinoco (Armbruster, 2002; Armbruster *et al.*, 2007), and two new species from the Negro River drainage (Tan & Armbruster, 2016) were described. Additionally, the monotypic genus *Micracanthicus* was synonymized with *Hypancistrus* (Lujan *et al.*, 2017). Among the described species there is a

general lack of diagnostic morphological characters, apart from the differences in color pattern (Armbruster *et al.*, 2007).

As with other Loricariidae, *Hypancistrus* plecos are popular ornamental fishes, with an important role in the aquarium market (de Sousa *et al.*, 2018). They are captured by local fisherman who sell them to companies that export these animals to markets in Europe and North America, or resell them in the domestic market (de Araújo *et al.*, 2017; de Sousa *et al.*, 2018). The process is, in many cases, illegal, and can have negative impacts on the native populations (Carvalho Júnior *et al.*, 2009). The aquarium community classifies different morphological types of loricariids on a number-based system created by the German magazine DATZ (www.datz.de). In this system the code is composed by the letter L indicative of the Loricariidae family, followed by a number, e.g. *H. zebra* is known as L046. Types assigned to each number may correspond to true species, populations within species, or polymorphic variants that do not reflect genetic isolation. There are approximately 36 L-numbers assigned to *Hypancistrus* forms that have yet to be studied and may represent distinct species (www.planetcatfish.com). Though not all of those are likely to represent unique species, as is true of described species, color patterns are the most important character in defining *Hypancistrus*' L groups, suggesting much higher diversity of *Hypancistrus* than is recognized by the current taxonomy. Regardless of the correlation with taxonomic diversity, it has been noted that the great diversity of coloration in these fishes has been suggested to be an argument for its importance in their life history (Armbruster *et al.*, 2007). Due to the widespread use of the L-number system in the fish keeping community and in the scientific literature, as well as a lack of a better form of classifying them, we will be using L-numbers to refer to our samples that lack a formal scientific name with the

exception of one population from the Tapajós River that has not been assigned an L number (Jacareacanga).

While the majority of *Hypancistrus* described species are concentrated in the upper Orinoco Basin and upper Negro River drainage (Armbruster, 2002; Armbruster *et al.*, 2007; Tan & Armbruster, 2016), records from the aquarium community (www.planetcatfish.com) along with publicly available collection data (www.gbif.org; Cardoso *et al.*, 2016) indicate that *Hypancistrus* is widespread throughout the Amazon Basin, having been collected from the Curuá-Una, Jari, Madeira, Negro, Paruari, Purus, Tapajós, Tocantins, and Xingu River drainages. As the diversity and distribution of the genus is so poorly understood, it is unsurprising that few studies have included *Hypancistrus* in phylogenetic analyses. There are good phylogenetic reconstructions of Loricariidae that have included representatives of the genus (Lujan *et al.*, 2015; Lujan *et al.*, 2017; Roxo *et al.*, 2019), consistently supporting the position of *Hypancistrus* within the Peckoltia clade. Lujan *et al.* (2015; 2017) included six species of *Hypancistrus* in their phylogenetic reconstructions, finding species from the Orinoco Basin to be a monophyletic group. Support and resolution for other relationships within *Hypancistrus* were low, suggesting recent divergence and/or the inability of the chosen molecular markers to recover the relationships between these species. As these studies included only one species occurring outside of the Orinoco Basin, the ability to make inferences about evolutionary patterns in the genus is quite limited. To date, no attempt to undertake a comprehensive assessment of the evolutionary history and relationships within *Hypancistrus* has been made.

Here we present a phylogenetic analysis of the genus *Hypancistrus* and explore the evolutionary history of this group in the dynamic hydrological landscape of cis-Andean South America. We used genomic double digest restriction associated DNA sequencing (ddRAD) data

to unravel the basic relationships among *Hypancistrus* species. We tested the hypothesis that *Hypancistrus* from the Orinoco and from the Amazon Basins form reciprocally monophyletic groups in accordance to the finding of Lujan et.al. (2015) and Lujan et.al. (2017), and the common pattern of division between these basins in other fish groups. We give special emphasis to exploring the relationships among *Hypancistrus* from the Amazon Basin since these represent the larger gap in the literature, and in exploring evidence of admixture events among lineages.

MATERIALS AND METHODS

Sampling and Library prep

A total of 127 tissue samples representing six described and three undescribed species belonging to the genus *Hypancistrus* were obtained from the Ichthyology Collection of the Academy of Natural Sciences of Drexel University – ANSP (Fig. 2). To increase the relevance of our sampling, we partnered with fish keepers from Norway, who donated 38 additional samples (fin clips) representing four described and eight undescribed *Hypancistrus* species, accompanied by photographs of each sampled individual to confirm identifications. We also obtained ten samples for nine species belonging to the Peckoltia clade from the ANSP collection that were used as outgroups, matching the sampling of Roxo *et al.*(2019). An overview of sampled species is given in Table 2.

DNA was extracted following a salt-extraction protocol modified from Aljanabi & Martinez (1997) and concentration of DNA was quantified with a Qubit® Fluorometer (Thermo Fisher). Reduced-representation libraries were prepared following a modified version of the double digest restriction associated DNA sequencing (ddRAD, Peterson *et al.*, 2012). In this method total genomic DNA was digested using a pair of restriction enzymes (EcoRI and XbaI),

adding an additional enzyme (NheI-HF) with the goal of cutting adapter dimers (Glenn et al. unpublished). Digestion was followed by the steps of adaptor ligation, PCR, and clean-up. Unique pairs of barcodes were added during the PCR step to allow for sample pooling and post-sequencing assignation of reads to individual samples. DNA fragments of 324 to 416 base pairs were selected by a Pippin Prep (Sage Science). Library quality and quantification was obtained with qPCR in a Rotor-Gene Q (QIAGEN) using the KAPA Library Quantification kit for Illumina platforms (Kapa Biosystems). Sequencing was performed with an Illumina NextSeq 500 with single-end readings at the National Center for Natural Products Research at the University of Mississippi.

Table 2: List of *Hypancistrus* and outgroup samples obtained from the Ichthyology Collection of the Academy of Natural Sciences of Drexel University – ANSP and from fish keepers from Norway.

Species	**Locality**	**Number of samples**		
		ANSP	**Norway**	**Total**
Hypancistrus contradens	Ventuari River	8	2	10
Hypancistrus debilittera	Orinoco River	2		2
Hypancistrus furunculus	Orinoco River	10	3	13
Hypancistrus inspector	Negro River		3	3
Hypancistrus lunaorum	Orinoco River	3		3
Hypancistrus margaritatus	Takutu River	1		1
Hypancistrus vandragti	Orinoco River		2	2
Hypancistrus zebra	Xingu River	8		8
Hypancistrus sp. (Jacareacanga)	Tapajós River	5		5
Hypancistrus sp. (L66/333)	Xingu River	70		70
Hypancistrus sp. (L136)	Negro River		3	3
Hypancistrus sp. (L174)	Xingu River	20		20
Hypancistrus sp. (L201)	Orinoco River		3	3
Hypancistrus sp. (L260)	Tapajós River		3	3
Hypancistrus sp. (L270)	Curuá-Una River		1	1
Hypancistrus sp. (L316)	Jari River		6	6
Hypancistrus sp. (L411)	Jari River		3	3
Hypancistrus sp. (L499)	Paduari River		4	4
Hypancistrus sp. (L500)	Uatumã River		5	5
Outgroups				
Ancistomus feldbergae	Xingu River	2		2
Panaqolus sp.	Xingu River	1		1
Panaqolus sp. (L002)	Tocantins River	1		1
Peckoltia sabaji	Xingu River	1		1
Peckoltia vittata	Xingu River	1		1
Scobinancistrus aureatus	Xingu River	1		1
Scobinancistrus pariolispus	Iriri River	1		1
Spectracanthicus immaculatus	Xingu River	1		1
Squaliforma emaginata	Iriri River	1		1
Total		137	38	175

Sequence processing

All sequence processing was performed using the Mississippi Center for Supercomputing Research (MCSR) supercomputing clusters. Demultiplexing, the assignment of reads to individuals based on unique barcodes added to each library during PCR, was done using the

bcl2fastq software (available at support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html). We allowed for a maximum mismatch of two base pairs (bp) in barcodes, and trimmed all reads of restriction site overhangs, resulting in a final fragment length of 56 bp for all reads. Fragments were aligned and single nucleotide polymorphisms (SNPs) were extracted using the ipyrad software v. 0.7.28 (Eaton & Overcast, 2019). There is a lack of consensus regarding the most appropriate set of parameters for RAD data genotype calling and filtering (O'Leary *et al.*, 2018), therefore we generated a variety of datasets produced by changing one parameter at a time (Table 3). We decided on a combination of filters that selected high-quality reads to maximize the selection of homologous loci. In the first filtering step we defined base calls with phred Q score of less then 33 as ambiguous and reads with ambiguous sites were excluded. On the within-sample clustering step we allowed for a clustering threshold for *de novo* assembly of 0.85 and a maximum of two unique alleles per locus, since these are diploid species (da Silva *et al.*, 2014). Consensus sequences with more then two uncalled bases (N) or more then eight heterozygous sites were excluded and we required a minimum depth for base calling of ten reads. For among sample assembly we required a clustering threshold of 0.85, minimum of 50% of samples represented per locus (maximum 50% missing data per locus), and maximum of ten SNPs, four indels, and 50% of shared polymorphic sites per locus. To generate final output files we used one of two approaches: 1) complete sequences were concatenated in a supermatrix; or 2) one SNP per locus was selected to build a matrix of putatively unlinked SNPs. Shared locus selection with *de novo* asembly methods is intrinsicaly dependent on which samples are used, therefore we ran step six of ipyrad (clustering reads among samples) separetly for each one of the different sampling combinations we used (Table 4; datasets 1 to 4 cover the Orinoco and

Amazon Basins; 5 to 11 focus on the Amazon Basin). Additionally, to test the effect of low quality samples on the final number of loci and on the outcome of phylogenetic analyses, we generated alternative datasets excluding samples with more then 25% of missing data after step six of ipyrad (datasets 3 and 4), according to O'Leary *et al.* (2018).

Table 3: Combinations of ipyrad parameters tested to generate SNP datasets of *Hypancistrus* and outgroups including 92 individuals and 28 species. Parameters that are not relevant for genotyping are omitted.

Affected step*	Parameter	Original	Tested parameter									Final
			9	11/12	14	19	20	21	22	23	24	
3	[5] Assembly method	denovo										denovo
1 to 7	[7] Datatype	ddrad										ddrad
2	[9] Max low quality base calls (Q<20) in a read	5	0									0
2	[10] phred Q score offset	33										33
4 & 5	[11] Min depth for statistical base calling	6		10								10
4 & 5	[12] Min depth for majority-rule base calling	6		10								10
4 & 5	[13] Max cluster depth within samples	10000										10000
3 & 6	[14] Clustering threshold for de novo assembly	0.85			0.9							0.85
2	[16] Filter for adapters/primers	0										0
4 & 7	[18] Max alleles per site in consensus sequences	2										2
5	[19] Max uncalled bases in consensus	5				2						2
5	[20] Max heterozygotes in consensus	8					5					5
7	[21] Min number of samples per locus for output	50%						80%				50%
7	[22] Max number of SNPs allowed in final locus	20							10			10
7	[23] Max number of indels per locus	8								4		4
7	[24] Max number of heterozygous sites per locus	0.5									0.25	0.5
2	[25] Trim raw read edges	0, 0										0, 0
7	[26] Trim locus edges	0, 0										0, 0
Total number of loci in assembly		12585	12569	10973	12897	12616	12731	8168	9848	11644	12552	9125

*ipyrad assembly steps (Eaton & Overcast, 2019):

1 Loading fastq files
2 Filtering / editing reads
3 Clustering / mapping reads within samples and alignment
4 Joint estimation of heterozygosity and error rate
5 Consensus base calling and filtering
6 Clustering / mapping reads among samples and alignment
7 Filtering and formatting output files

Table 4: Summary of analyzed datasets in relation to samples included, number of loci, and percentage of missing data. Datasets 1 to 4 include *Hypancistrus* species from the Orinoco and Amazon Basins, and datasets 5 to 11 concentrate sampling on the Amazon Basin. Combination of ipyrad parameters is the same as described in the final scheme in Table 3, to the exception of datasets 5 and 6 for TreeMix analyzes, which required exclusion of all indels and missing data (minimum number of samples per locus = 100%; maximum number of indels per locus = 0)

	N species (ingroup / outgroup)	**N samples (ingroup / outgroup)**	**Data matrix**	**Total loci**	**% missing data**	**Analysis**
1	19 *Hypancistrus* / 9 outgroups	82 / 10	concatenated	9125 (518358 bp)	22.6%	IQ-TREE
2			unlinked SNPs	9018	23.3%	SVDquartets SplitsTree PCA
3	18 *Hypancistrus* / 4 outgroups	56 / 5	concatenated	11449 (650726 bp)	14.7%	IQ-TREE
4			unlinked SNPs	11153	13.9%	SVDquartets
5	11 Amazon clade *Hypancistrus* / 1 Orinoco clade *Hypancistrus*	26 / 2	unlinked SNPs	2021	0%	TreeMix Fourpop test
6	7 Amazon clade *Hypancistrus* / 1 Orinoco clade *Hypancistrus*	39 / 4	unlinked SNPs	2612	0%	TreeMix
7	14 Amazon clade *Hypancistrus* / 1 Orinoco clade *Hypancistrus*	132 / 4	unlinked SNPs	11670	19.0%	Structure
8	8 Amazon clade *Hypancistrus* / 2 Orinoco clade *Hypancistrus*	22 / 6	concatenated	13328 (757465bp)	14.5%	IQ-TREE
9			unlinked SNPs	11052	13.8%	SVDquartets
10	7 Amazon clade *Hypancistrus* / 2 Orinoco clade *Hypancistrus*	19 / 6	concatenated	12992 (737615 bp)	13.7%	IQ-TREE
11			unlinked SNPs	10438	13.1%	SVDquartets

Phylogenetic and network analyses

Maximum likelihood (ML) phylogenetic trees were estimated with IQ-TREE v. 1.6.10 (Nguyen *et al.*, 2014) in the CIPRES Science Gateway portal v. 3.3 (Miller *et al.*, 2010). We ran

analyses using datasets 1, 3, 8, and 10, with complete sequences concatenated in supermatrices (Table 4). We set *Spectracanthicus imaculatus* as the outgroup for dataset 1, *Scobinancistrus aureatus* as the outgroup for dataset 3, and no outgroup was defined for datasets 8 and 10. The automatic model selection method ModelFinder implemented by IQ-TREE was used to determine the best-fit substitution models based on the Bayesian information criterion. The transversion model and unequal base frequencies estimated empirically from the alignment with the FreeRate model of heterogeneity across sites with four categories (TVM+F+R4) was selected as the best-fit model for dataset 1. The same model but with three categories on the FreeRate model of heterogeneity (TVM+F+R3) was selected for datasets 3 and 8. For dataset 9, the best-fit model was the general time reversible (GTR) model and unequal base frequencies estimated empirically from the alignment, with the FreeRate model of heterogeneity across sites with three categories (GTR+F+R3). We ran 1,000 replicates for Ultrafast bootstrap branch support estimation, optimizing tree search by a nearest neighbor interchange search, which reduces overestimation of branch support values (Hoang *et al.*, 2017).

Relationships among individual samples were also estimated with the coalescent model used by single value decomposition scores for species quartets SVDquartets (Chifman & Kubatko, 2014, 2015) implemented in PAUP* v. 4.0 (Swofford, 2003). Matrices composed of unlinked SNPs (datasets 2, 4, 9, and 11; Table 4) were used to run exhaustive quartet sampling. Similar to the ML analyses, we set *S. imaculatus* as the outgroup for dataset 2 and *S. aureatus* as the outgroup for dataset 4. Both *H. contradens* and *H. fururnculus* were used as the outgroup for datasets 9 and 11. Branch support was obtained with 100 bootstrap replicates and results were written into the 50% majority rule consensus tree.

Finally, to allow for identification of reticulation events like gene flow and the detection of conflicting phylogenetic signals we built a phylogenetic network using SplitsTree (Huson & Bryant, 2006). We used the UncorrectedP distance method to calculate genetic distances between samples and the NeighborNet distance method to compute the network. The matrix of unlinked SNPs in dataset 2 was used in this analysis (Table 4). To simplify visualization, we omitted the outgroups from the network graphics.

Principal components analysis

To explore how samples are distributed within the genetic variance space we ran a principal components analysis (PCA) using *Hypancistrus* samples from dataset 2 with unlinked SNPs (Table 4), as implemented by the adegenet package v. 2.1.1 (Jombart & Ahmed, 2011) and the ade4 package v. 1.7.13 (Dray & Dufour, 2007) in R v. 3.6.1 (R Development Core Team, 2019). This is a good exploratory method, as non-parametric multivariate methods are robust to varying choices of SNP calling and filtering in reduced representation library datasets (Linck & Battey, 2018). We replaced missing data by the mean allele frequencies as suggested by adegenet's developer (Jombart, 2008). We examined the three first principal components and plotted the results from the first two axes, grouping samples based on species identifications. We repeated the PCA analysis with the same dataset, selecting only the samples from the Amazon Basin.

Population history inference with migration

We used TreeMix v. 1.13 (Pickrell & Pritchard, 2012) to jointly estimate the historical relationships among populations along with migration edges to model non-bifurcating processes.

We created two distinct sampling schemes focusing on the Amazon Basin populations (Table 4): for dataset 5 we selected two samples per population, excluding populations with only one sample, and *H. contradens* was used as the outgroup; for dataset 6 we selected from two to seven high coverage samples per population (more then 15k loci per sample prior to step 6 of ipyrad) and *H. furunculus* was used as the outgroup. We divided *H.* sp. (L66/333) into three populations according to their distribution along the Xingu River: collection site 2 from the upstream limit, sites 3-16 from the middle portion, and sites 17-22 (referred to as x2, x3-16, x17-22 on the results) on the furthest downstream reaches of their distribution, near the confluence of the Xingu with the Amazon River. We generated datasets with ipyrad that excluded indels and missing data since TreeMix doesn't perform well with incomplete datasets. Additionally, non-binary loci were excluded, and input files with allele frequencies were generated with the dartR package v. 1.3.4 for R (Gruber *et al.*, 2019).

The TreeMix algorithm works by estimating a ML tree of relationships among populations, calculating a residual covariance matrix for the model, adding migration edges (*m*) to increase the likelihood of the model, and re-adjusting the tree (Pickrell & Pritchard, 2012). We assumed loci are independent, testing zero to ten migration edges ($m = 0$ to 10) allowing for free tree rearrangement, and generating jackknife estimates of standard errors for the weight of migration edges. We accessed the residuals of the population trees, the likelihood of the models, and the proportion of variance explained by the models to compare runs. We defined the best *m* based on when the model's likelihood reached an asymptote. To check for consistency, we conducted ten independent runs of TreeMix for each *m* and reported results on the replicate with the highest likelihood.

Population structure

We inferred genetic structure among populations from the Amazon Basin with ParallelStructure v. 2.3.4 (Besnier & Glover, 2013) which allows the implementation of the Structure algorithm (Pritchard *et al.*, 2000) across multiple processors using the R language. We built a matrix of unlinked SNPs including 132 samples of *Hypancistrus* form the Amazon Basin and added four samples of *H. contradens* from the Orinoco Basin to check for admixture between basins (Table 4). We heavily sampled populations from the Xingu River including three species: *H. zebra* and *H.* sp. (L174) that, by and large, co-occur in the middle reaches of the Xingu River, and *H.* sp. (L66/333) that is restricted to the lower Xingu.

We ran ParallelStructure in the CIPRES Science Gateway portal, using the admixture model and the F model of allele frequencies correlation across populations, which allows for better estimation of clustering for closely related populations. We tested clustering individuals in one to ten genetic groups (k = 1 to 10), and ran five replicas of each k for 200,000 iterations, using a burn-in of 50,000 iterations. We accessed the best-fit k based on the Evanno method (Evanno *et al.*, 2005) implemented by Structure Harvester web v. 0.6.94 (Earl & vonHoldt, 2012), and summarized results across runs with the main pipeline of Clumpak (Kopelman *et al.*, 2015). To test for the effect of unequal sample sizes across populations, we re-ran ParallelStructure selecting three samples belonging each of the main lineages from the Xingu River.

Four-population tests

To test the robustness of our population history inferences we ran four-population tests developed by Reich *et al.* (2009) and implemented with TreeMix (Pickrell & Pritchard, 2012).

Again we focused on the Amazon Basin populations and used dataset 5 (Table 4). Four populations are selected at a time, for which there are three possible unrooted trees to explain their relationships in the absence of admixture: (A,B)(C,D), (A,C)(B,D), (A,D)(B,C). The four-population test gives a D-statistic for each topology, and an associated z-score to assess significance. A D-statistic significantly differing from zero rejects the hypothesis for a given topology. We calculated p-values based on two-tailed tests for the z-scores. For all our tests we used *H. contradens* as one of the populations, which we used to infer the root of the trees.

RESULTS

Sequence processing

After sequencing, our samples averaged 2,532,456 reads. Our tests for parameter selection on ipyrad assembly for our ddRAD sequences of *Hypancistrus* and outgroups (dataset 1) resulted in between 8,168 and 12,897 total loci (Table 3). The parameters that most influenced the final number of loci in the assemblies were 1) minimum number of samples per locus and 2) maximum number of SNPs allowed per locus, with a 35% and 22% decrease respectively in relation to our initial combination of parameters. None of our tests produced a relevant increase in the number of loci relative to our original parameter settings. Our final parameter combination choice generated an average of 14,599 loci per sample after within-sample assembly and an average of 7,060 loci per sample after between-sample assembly and filtering in a total of 9,125 loci for dataset 1, with 22.6% of missing data (Table 4). Our other sampling schemes generated additional datasets varying from 2,012 loci (0% missing) in dataset 5 to 13,328 loci (14.5% missing) in dataset 8, with a maximum proportion of missing data of 23.3% in dataset 2.

Concatenated datasets varied between 518,358 bp and 757,465 bp, with individual loci varying between 56 bp and 61 bp in length.

Phylogenetic and network analyses

Maximum likelihood and SVDquartets coalescent trees show high support for the monophyly of the genus *Hypancistrus* (Fig. 3-6). Relationships among outgroup species are consistent with the findings of Roxo *et al.*(2019), except for the SVDquartets analysis with 92 samples (dataset 2, Fig. 4), where *Ancistomus* was recovered as sister to *Panaqolus* instead of *Peckoltia*. Notably, support values for these relationships are low and thus not contradictory to previous findings. Analyses with datasets that excluded low quality samples (ML: dataset 3, Fig. 5; SVDquartets: dataset 4, Fig. 6) did not produce strongly supported differences when results are compared to the analyses with complete sampling.

Both ML an SVDquartets methods support a basal division within *Hypancistrus* between two geographically exclusive monophyletic clades inhabiting the Orinoco and the Amazon Basins, with the exception of *H. inspector* from the Casiquiare River, which connects the two basins. One of the *H. inspector* specimens is sister to all Orinoco lineages, while the other two samples group within the Amazon clade. *Hypancistrus vandragti* is sister to the remaining species of the Orinoco clade, but the two samples are not recovered as monophyletic in the SVDquartets tree (Fig. 4). Both samples of *H. debilittera* form a monophyletic lineage in all analyses, although their relation to other species remain unclear due to the lack of strong support for their placement. All specimens identified as *H. furunculus* consistently form a monophyletic group, as does the grouping of *H. contradens*, *H. lunaorum,* and *H.* sp. (L201) though support for the monophyly of each lineage is lacking (Fig. 3-6).

In the Amazon clade, the species from the Tapajós River *H.* sp. (Jacareacanga) and *H.* sp. (L260), along with *H. zebra* from the Xingu River consistently form a strongly supported monophyletic group (Fig. 3-6). Monophyly is also supported for the group of species native to Amazonian tributaries draining from the Guiana Shield, including *H. margaritatus*, *H.* sp. (L136), *H.* sp. (L316), *H.* sp. (L499), and *H.* sp. (L500). The relationships among these species remain uncertain, except for the high support for the sister relationship of *H. margaritatus* and *H.* sp. (L136) and *H.* sp. (L270) within *H.* sp. (L316) in all analyses, and one of the samples identified as *H.* sp. (L316) is recovered as closely related to *H.* sp. (L411). The most relevant discrepancies between the ML and SVDquartets results involve *H.* sp. (L174) and *H.* sp. (L66/333) from the Xingu River. Specifically, *Hypancistrus* sp. (L174) appears as a derived lineage nested within *H.* sp. (L66/333) in the ML trees (Fig. 3 & 5), yet it clusters with *H. zebra* in the SVDquartets trees (Fig. 4 & 6). Contrary to the ML method, SVDquartets splits *H.* sp. (L66/333), placing the two samples closest to the mouth of the Xingu (sample IDs 10329 and 12634) in the Guiana Shield drainages clade though support for this relationship is low (Fig. 4 & 6). Aside from the uncertainties regarding these Xingu River lineages, *H.* sp. (L411) is well supported as a sister group to the Guiana Shield drainages clade.

The phylogenetic network built with SplitsTree is largely concordant with the phylogenetic trees (Fig. 7). The largest divergence is represented by the split between the Amazon and the Orinoco clades. Interestingly, *H. inspector* is positioned between both major clades. Within the Orinoco clade, *H. furunculus* and *H. vandragti* are recovered as monophyletic, while *H. contradens*, *H. lunaorum* and *H.* sp. (L201) are again recovered as members of a clade that fails to differentiate individual taxa. Overall, the Amazon clade appears to possess greater genetic diversity (i.e. longer branch lengths). The Xingu species *H.* sp. (L174) and *H.* sp.

(L66/333) share connections to both *H. zebra* and the Guiana Shield drainage lineages, suggesting introgression among them.

Principal components analysis

The first axis of the PCA including all *Hypancistrus* species (Fig. 8A) explains 26% of the variance in the data and divides the Orinoco Basin and the Amazon Basin lineages. The second axis explains 13% of the variance, and shows a gradient with the Tapajós and *H. zebra* samples in one extreme and Guiana Shield drainages lineages in the other. For the PCA with Amazon Basin *Hypancistrus* the first principal component explains 24% of the variance and the second explains 13% of the variance (Fig. 8B). *Hypancistrus* sp. (Jacareacanga) and *H. zebra* have similar scores on the first principal component but are clearly separated by the second principal component. Samples from *H.* sp. (L174), *H.* sp. (L66/333), and *H.* sp. (L260) have intermediate distributions, suggesting potential admixture.

Population history inference with migration

TreeMix analysis including 14 populations with a lower restriction on sequencing coverage is depicted in Figure 9A-C. The maximum likelihood tree without migration edges explains 88.7% of the variance in relatedness among populations. The topology shows the same consistent patterns observed in the IQ-Tree and SVDquartets results, with a clade including *H. zebra*, *H.* sp. (Jacareacanga) and *H.* sp. (L260), and *H.* sp. (L411) as sister to the Guiana Shield drainages clade. The residual matrix indicates *H.* sp. (L174) and *H. zebra* are more closely related to each other than the tree suggests. Other pairs with high residuals are *H.* sp. (L66/333) x2 and *H. zebra, H.* sp. (L66/333) x3-16 and *H.* sp. (L66/333) x17-22, and *H.* sp. (L499) and *H.*

inspector. Based on when the model's likelihood reached an asymptote, we determined four was the optimal number of migration edges for this dataset, in a model that explains 98.6% of the variance in the data. The strongest admixture event was from *H. inspector* to an ancestral population of *H.* sp. (L411) and the Guiana Shield drainages clade (weight = 0.50, $p < 10^{-30}$). This ancestral population is also involved in admixture edges towards *H.* sp. (L66/333) x17-22 (weight = 0.44, $p < 10^{-30}$) and *H.* sp. (L260) (weight = 0.23, $p = 3.8 \times 10^{-11}$). The last admixture event is from *H. zebra* to an ancestral of *H.* sp. (L174) and *H.* sp. (L66/333) x2 (weight = 0.22, $p = 1.8 \times 10^{-12}$).

The second TreeMix analysis with a higher restriction on sequence coverage included 10 populations (Fig. 9D-F). The maximum likelihood tree without migration edges explains 90.7% of the variance in relatedness among populations and has the same topology as the tree for the previous dataset. The residual matrix is also similar to that of the less restrictive dataset. For this dataset there was strong evidence for the model with three migration edges that explain 99.1% of the data variance. An ancestral of *H.* sp. (L174) and *H.* sp. (L66/333) x2 is the origin of an admixture edge towards *H.* sp. (L66/333) x17-22 (weight = 0.42, $p < 10^{-30}$). That same ancestral population received admixture from *H. zebra* (weight = 0.22, $p = 2.3 \times 10^{-5}$). The third edge is from an ancestral population of *H. zebra* and *H.* sp. (Jacareacanga) to the ancestor of *H.* sp. (L174), *H.* sp. (L66/333) x2, and *H.* sp. (L66/333) x3-16 (weight = 0.38, $p < 10^{-30}$).

Population structure

The Structure analysis for our larger dataset recovered eight clusters, while the dataset excluding most samples from populations from the Xingu River recovered seven clusters (Fig. 10). The additional cluster in the larger dataset is correspondent to *H.* sp. (L66/333). Regardless

of the sampling scheme, four species showed strong signal of admixed origins: *H. inspector*, *H.* sp. (L260), *H.* sp. (L66/333), and *H.* sp. (L499). Notably, in the Guiana Shield Clade samples identified as *H. margaritatus*, *H.* sp. (L136), *H.* sp. (L316), and *H.* sp. (L500) are grouped in the same cluster.

Based on these results we decided to re-run the phylogenetic analyses for the Amazon clade in IQ-Tree and SVDquartets excluding the admixed species with the goal of finding a well-supported tree for the remaining lineages (Fig. 11). Both methods recovered trees with high support. The only difference among them is in the placement of *H.* sp. (L174), which appears as the sister to *H. zebra* in the ML trees (Fig. 11A & C), as the sister to the clade including *H. zebra* and *H.* sp. (Jacareacanga) in the SVDquartets tree (Fig. 11B), and as the sister to all Amazon clade species in the SVDquartets analyses that excluded *H. zebra* (Fig. 11D).

Four-population tests

Our four-population tests to evaluate the robustness of our phylogenetic inferences for non-admixed clades support those results (Table 5). The tests are compatible with the grouping of *H. zebra* and *H.* sp. (Jacareacanga) (supported topology 6; Table 5) and the positioning of *H.* sp. (L411), as the sister clade to the Guiana Shield drainages clade (GSD in Table 5; supported topology 2). Interestingly, the four-population tests suggest *H.* sp. (L174) is the sister clade to all other populations in the Amazon clade (supported topology 7), agreeing with the SVDquartets analyses that excluded *H. zebra* (Fig. 11D). The tests that included both *H. zebra* and *H.* sp. (L174) gave unexpected results. There is weak support for topology 8 that places *H.* sp. (L174) in the base of the tree, with p-values at 0.01, which might suggest there is introgression between *H. zebra* and *H.* sp. (L174). The most paradoxical result was the passing score supporting

topology 9, that places *H.* sp. (Jacareacanga) closer to *H.* sp. (L174) then to *H. zebra*, going against all other evidence. Four-population tests can give false passing scores for populations that have admixture levels that cancel each other in just the right way, and we believe this is the best explanation for this finding (Reich *et al.*, 2009).

Table 5. Four populations tests for putatively non-admixed populations of *Hypancistrus* belonging to the Amazon clade. Topologies that agree with the true population history without admixture are expected to give non-significant z-scores. The best-fit topologies are indicated in bold. *Hypancistruus contradens* was used as the fourth population in all tests and assumed to be the basal taxon (omitted for clarity). Results are presented in triplets of possible combinations of trees for a given set of three populations. GSD – Guiana Shield drainages clade.

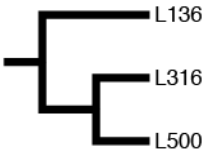
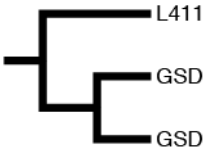
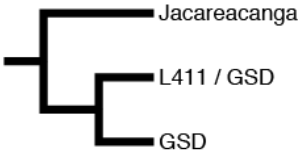
Supported topology	Tested topology		z-score / p-value
1 	**L136**	**L316, L500**	**-0.37 / 0.71**
	L500	L136, L316	-1.46 / 0.15
	L316	L136, L500	-2.11 / 0.03
2 	**L411**	**L136, L316**	**-0.11 / 0.92**
	L136	L316, L411	-3.61 / 3.1 x 10^{-4}
	L316	L136, L411	-3.87 / 1.0 x 10^{-4}
	L411	**L136, L500**	**-2.25 / 0.02**
	L136	L411, L500	3.68 / 2.3 x 10^{-4}
	L500	L136, L411	-5.32 / 1.0 x 10^{-7}
	L411	**L316, L500**	**-1.59 / 0.11**
	L316	L411, L500	2.56 / 0.01
	L500	L316, L411	-3.82 / 1.3 x 10^{-4}
3 	**Jacareacanga**	**L136, L316**	**0.92 / 0.36**
	L136	Jacareacanga, L316	9.6 / < 10^{-16}
	L316	Jacareacanga, L136	9.09 / < 10^{-16}
	Jacareacanga	**L136, L411**	**0.67 / 0.5**
	L411	Jacareacanga, L136	7.15 / 8.7 x 10^{-13}
	L136	Jacareacanga, L411	7.64 / 2.2 x 10^{-14}
	Jacareacanga	**L136, L500**	**-2.19 / 0.03**
	L136	Jacareacanga, L500	9.77 / < 10^{-16}
	L500	Jacareacanga, L136	10.64 / < 10^{-16}
	Jacareacanga	**L316, L411**	**-0.1 / 0.92**
	L316	Jacareacanga, L411	6.83 / 8.6 x 10^{-12}
	L411	Jacareacanga, L316	6.86 / 6.9 x 10^{-12}
	Jacareacanga	**L316, L500**	**-2.43 / 0.01**
	L316	Jacareacanga, L500	8.27 / 2.2 x 10^{-16}
	L500	Jacareacanga, L316	9.72 / < 10^{-16}
	Jacareacanga	**L411, L500**	**-1.96 / 0.05**
	L411	Jacareacanga, L500	6.26 / 3.9 x 10^{-10}
	L500	Jacareacanga, L411	7.68 / 1.5 x 10^{-14}

Table 5. Continued.

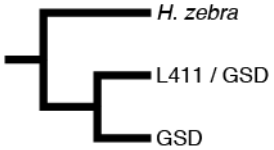
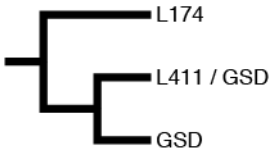
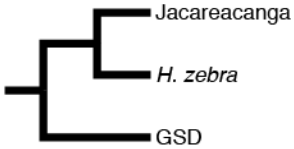
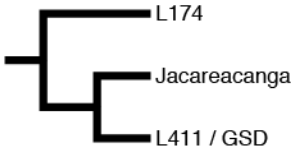
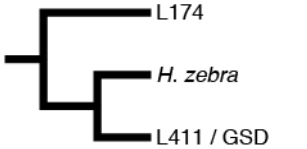
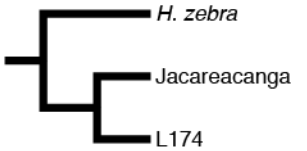
4 	***H. zebra***	**L136, L316**	**0.63 / 0.53**
	L136	L316, *H. zebra*	-9.75 / < 10^{-16}
	L316	L136, *H. zebra*	-9.46 / < 10^{-16}
	H. zebra	**L136, L411**	**0.38 / 0.7**
	L411	L136, *H. zebra*	-7.45 / 9.3 x 10^{-14}
	L136	L411, *H. zebra*	-7.65 / 1.9 x 10^{-14}
	H. zebra	**L136, L500**	**-2.21 / 0.03**
	L136	L500, H. zebra	-9.82 / < 10^{-16}
	L500	L136, H. zebra	-10.69 / < 10^{-16}
	H. zebra	**L316, L411**	**-0.09 / 0.93**
	L316	L411, *H. zebra*	-7.12 / 1.1 x 10^{-12}
	L411	L316, *H. zebra*	-7.18 / 6.9 x 10^{-13}
	H. zebra	**L316, L500**	**-2.27 / 0.02**
	L316	L500, *H. zebra*	-8.62 / < 10^{-16}
	L500	L316, *H. zebra*	-9.87 / < 10^{-16}
	H. zebra	**L411, L500**	**-1.72 / 0.09**
	L411	L500, *H. zebra*	-6.63 / 3.3 x 10^{-11}
	L500	L411, *H. zebra*	-7.78 / 7.1 x 10^{-15}
5 	**L174**	**L136, L316**	**0.2 / 0.84**
	L316	L136, L174	-6.13 / 8.7 x 10^{-10}
	L136	L174, L316	6.18 / 6.2 x 10^{-10}
	L174	**L136, L411**	**0.19 / 0.85**
	L411	L136, L174	-3.43 / 6.0 x 10^{-4}
	L136	L174, L411	3.48 / 4.9 x 10^{-4}
	L174	**L136, L500**	**-2.02 / 0.04**
	L136	L174, L500	6.33 / 2.4 x 10^{-10}
	L500	L136, L174	-7.54 / 4.7 x 10^{-14}
	L174	**L316, L411**	**0.01 / 0.99**
	L411	L174, L316	3.03 / 2.4 x 10^{-3}
	L316	L174, L411	3.04 / 2.3 x 10^{-3}
	L174	**L316, L500**	**-1.75 / 0.08**
	L316	L174, L500	5.05 / 4.3 x 10^{-7}
	L500	L174, L316	6.4 / 1.6 x 10^{-10}
	L174	**L411, L500**	**-1.51 / 0.13**
	L411	L174, L500	2.21 / 0.03
	L500	L174, L411	3.52 / 4.3 x 10^{-4}
6 	**L136**	**Jacareacanga, *H. zebra***	**-0.75 / 0.45**
	Jacareacanga	L136, *H. zebra*	7.2 / 6.1 x 10^{-13}
	H. zebra	Jacareacanga, L136	-7.66 / 1.8 x 10^{-14}
	L316	**Jacareacanga, *H. zebra***	**-0.97 / 0.33**
	Jacareacanga	L316, H. zebra	6.59 / 4.3 x 10^{-11}
	H. zebra	Jacareacanga, L316	-7.27 / 3.4 x 10^{-13}
	L411	**Jacareacanga, *H. zebra***	**-0.99 / 0.32**
	Jacareacanga	L411, *H. zebra*	6.75 / 1.4 x 10^{-11}
	H. zebra	Jacareacanga, L411	-7.44 / 1.0 x 10^{-13}
	L500	**Jacareacanga, *H. zebra***	**-0.76 / 0.44**
	Jacareacanga	L500, *H. zebra*	7.7 / 1.3 x 10^{-14}
	H. zebra	Jacareacanga, L500	-8.2 / 2.2 x 10^{-16}

Table 5. Continued.

7 	**L174**	**Jacareacanga, L136**	**0.61 / 0.54**
	Jacareacanga	L136, L174	5.21 / 1.8 x 10^{-7}
	L136	Jacareacanga, L174	5.64 / 1.7 x 10^{-8}
	L174	**Jacareacanga, L316**	**0.69 / 0.49**
	Jacareacanga	L174, L316	-4.37 / 1.2 x 10^{-5}
	L316	Jacareacanga, L174	4.99 / 5.9 x 10^{-7}
	L174	**Jacareacanga, L411**	**0.7 / 0.49**
	Jacareacanga	L174, L411	-4.45 / 8.4 x 10^{-6}
	L411	Jacareacanga, L174	5.05 / 4.5 x 10^{-7}
	L174	**Jacareacanga, L500**	**-0.13 / 0.9**
	L500	Jacareacanga, L174	5.53 / 3.1 x 10^{-8}
	Jacareacanga	L174, L500	-5.77 / 7.8 x 10^{-9}
8 	**L174**	**L136, *H. zebra***	**2.54 / 0.01**
	L136	L174, *H. zebra*	-6.17 / 6.9 x 10^{-10}
	H. zebra	L136, L174	8.36 / < 10^{-16}
	L174	**L316, *H. zebra***	**2.45 / 0.01**
	L316	L174, *H. zebra*	-5.96 / 2.4 x 10^{-9}
	H. zebra	L174, L316	-8.03 / 8.8 x 10^{-16}
	L174	**L411, *H. zebra***	**2.45 / 0.01**
	L411	L174, *H. zebra*	-6 / 1.9 x 10^{-9}
	H. zebra	L174, L411	-8.09 / 6.6 x 10^{-16}
	L174	**L500, *H. zebra***	**3.17 / 1.5 x 10^{-3}**
	L500	L174, *H. zebra*	-6.18 / 6.2 x 10^{-10}
	H. zebra	L174, L500	-8.93 / < 10^{-16}
9 	***H. zebra***	**Jacareacanga, L174**	**0.6 / 0.55**
	Jacareacanga	L174, *H. zebra*	4.45 / 8.7 x 10^{-6}
	L174	Jacareacanga, *H. zebra*	5.01 / 5.5 x 10^{-7}

DISCUSSION

Data quality

We built an unprecedented dataset of the *Hypancistrus* genus, both in terms of taxonomical sampling and in number of loci. We are aware that reconstruction of phylogenetic trees with RAD data presents potential problems, particularly related to missing data expected due to mutations in the restriction sites among clades that diverged a long time ago and the difficulty in inferring orthology and linkage among loci (Rubin *et al.*, 2012). However, recent research has demonstrated success in the reconstruction of phylogenetic relationships from RAD data, especially among recently diverged species, as it is the case for this study (Rubin *et al.*,

2012; Herrera & Shank, 2015; Pante *et al.*, 2015). The best estimate for timing of origin for the most recent common ancestor for *Hypancistrus* is around 2.25 million years ago (Ma), between the Neogene and Quaternary (Roxo *et al.*, 2019). We were not able to estimate a time calibrated tree for our dataset, due to the failure of runs to converge on the MCMC analysis performed by the package SNAPP (Bryant *et al.*, 2012) in BEAST (Drummond & Rambaut, 2007). We believe this is due to severe violations by our data of the model that assumes no gene flow between populations (Bryant *et al.*, 2012).

Nevertheless we are confident we have a high quality dataset, that while unsuitable for estimating divergence times, is sufficiently robust for estimating genetic structure among *Hypancistrus* populations. By testing different parameters for *de novo* assembly of our RAD sequences, we were able to select a set of conservative filters that should reduce the amount of orthologous loci by eliminating reads with low quality bases and low depth of coverage, and filtering out loci that have an elevated number of SNPs, indels, and heterozygous sites in consensus (Table 3). The minimum number of samples per locus was the parameter that had the greatest influence on our final number of loci. Although many authors prefer to implement more restrictive filters in order to reduce the amount of missing data, recent papers highlight the fact that these loci contain information that is particularly useful in the inference of relationships among closely related taxa (Huang & Knowles, 2016; Crotti *et al.*, 2019). In situations such as this it is preferable to employ downstream analyses that deal well with missing data or to use alternative approaches to remove missing data like imputation.

We were limited by the availability of tissue samples in ichthyological collections, but our sampling was greatly improved by the addition of specimens from the aquarium hobby. However, such sources of samples for phylogeographic study has a downside due to uncertainty

in the accuracy of the origin information associated with these specimens and their genealogical history (in the case of animals bred in captivity). We raise this issue in order to provide an explanation for two samples that stand out in our analyses. First, the specimen identified as *H.* sp. (L270) from the Curuá-Una River, a small southern tributary of the Amazon River located between the Tapajós and the Xingu rivers. This sample clusters within *Hypancistrus* sp. (L316) from the Jari River, that meets the Amazon River approximately 350 km downstream of the Jari. Additionally, sample 033 was identified as *H.* sp. (L316) but consistently grouped with *H.* sp. (L411), also in the Jari River. In this later case a natural hybridization may as likely explain our findings, since the Strucuture analysis assigned 12% of ancestry to *H.* sp. (L316) and 87% to *H.* sp. (L411), and both species occur in sympatry. Nevertheless, in all other cases the genetic data coupled with inspection of color pattern in photos taken from each individual make biological sense, and the inclusion of these samples added invaluable information to our data. Therefore, we choose to proceed with caution in drawing conclusions about biogeographical history based on these samples, but we believe they provide crucial insights for interpretation of our results.

Phylogenetic relationships and the Casiquiare River

Our phylogenetic, network, and PCA analyses support the basal division between the Orinoco and the Amazon clades, in agreement with Lujan *et al.* (2015) and Lujan *et al*. (2017). This pattern is common among other fish clades (Escobar *et al.*, 2015; Willis *et al.*, 2015; Lujan *et al.*, 2017). The Amazon and Orinoco Basins separated around 10 Ma with the orogenesis of the Vaupes Arch dividing the upper Orinoco and upper Negro Rivers (Albert *et al.*, 2018). Considering the estimate for the origin of *Hypancistrus* around 2.5 Ma (Roxo *et al.*, 2019), an ancient dispersal event must have occurred from one basin to the other. Based on the evidence of

higher genetic diversity in the Amazon clade, as seen in longer branches in the ML and network analyses (Fig. 3, 5 & 7), and the distribution of points across the PCA axes (Fig. 8A), we hypothesize the genus to have originated in the Amazon Basin.

Connecting the upper Orinoco and the upper Negro Rivers is the Casiquiare River, which itself originated approximately 10 thousand years ago (kya), in a bifurcation of the Orinoco into the Negro River (Laraque *et al.*, 2019). Geological evidence suggests the Casiquiare is an active headwater capture event, that is diverting waters of the Orinoco into the Negro River drainage (Stokes *et al.*, 2018). The type locality of *H. inspector* is 10 km above the confluence of the Casiquiare with the Negro River (Armbruster, 2002). Of the three samples of *H. inspector* we sequenced, two are more closely related to the Amazon clade, and the third is more closely related to the Orinoco clade. The intermediary placing of these samples in the network and PCA plots (Fig. 7 & 8), in addition to the admixed profile in the Structure plot (Fig. 10) suggests this species is admixed with both the Amazon and the Orinoco clades. TreeMix residual plots indicate *H. inspector* is closely related to the Guiana Shield drainages clade, particularly *H.* sp. (L499), a pattern also supported by the Structure results (Fig. 10). Notably *H.* sp. (L499) is from the Paduari River, itself a tributary of the Negro River to which the Casiquiare is connected. There is a strong ecological gradient along the Casiquiare River that moves from the clear water of the Oricono to the black water of the Negro River, that serves as a semi-permeable filter for movement of species between the two basins (Winemiller *et al.*, 2008). Thorough sampling across the distribution of *H. inspector* would provide an excellent model for the study of adaptation, introgression, diversification, and evolution across the Amazon and Orinoco Basins.

Orinoco clade

Our results show, for the first time, that the sympatric species *H. contradens* and *H. lunaorum* are not reciprocally monophyletic, but together form a monophyletic clade. The only diagnostic character to differentiate the two species is based on color pattern and consists of the relative diameter of white spots in relation to the nasal aperture (absent or smaller spots in *H. lunaorum*) (Armbruster *et al.*, 2007). As a result, we recommend synonymization of *H. lunaorum* with *H. contradens* since both were described in the same publication and *H. contradens* has the page priority (Armbruster *et al.*, 2007). Samples identified as *H. sp.* (L201) also clustered with *H. contradens* and should be treated as the same species. The striped *Hypancistrus* from the Orinoco Basin include *H. debilittera* and *H. furunculus*. They are morphologically distinguished by incomplete bands on the dorsal fin, indistinct dark E mark on the snout, and indistinct anterior dark bars in *H. debittera* as opposed to complete and distinct in *H. furunculus* (Armbruster *et al.*, 2007). All species of the Orinoco clade have very similar distributions in the Ventuari and upper Orinoco Rivers, to the exception of *H. debilittera* that occurs further downstream of the Orinoco River. All species occupy the same types of habitat.

Hypancistrus vandragti is well-supported as the sister taxon to all other members of the Oricoco clade, which is reflected by its morphological distinctiveness that lead to its original assignment to a separate genus (Lujan & Armbruster, 2011). Since then, a few studies have corroborated the placement of *H. vandragti* within *Hypancistrus*. The relationships among *H. contradens*, *H. debilittera* and *H. furunculus* remain unclear despite the massive number of loci analyzed in this study. Population genetic and phylogeographic studies with more comprehensive sampling of each of these taxa would be necessary to clarify the genetic structure within and among these species.

Amazon clade

Our ML and SVDquartets phylogenetic trees (Fig. 3 & 4) showed agreement in a few relationships, but were also inconsistent about the positioning of some lineages. In particular, *H. inspector*, *H.* sp. (L66/333), and *H.* sp. (L175) were recovered in a number of different positions in these different trees. Additionally, the interrelationships among the taxa from the Guiana Shield drainages clade are poorly supported. Based on the network, PCA, TreeMix, and Structure analyses (Fig. 7-10) we believe these inconsistencies are caused by events of introgression among lineages across the evolutionary history of *Hypancistrus* (detailed below), violating important assumptions of traditional phylogenetic estimation methods (Leaché *et al.*, 2013). The field of systematics has seen many methodological advancements in multilocus and multi-species phylogenetic inferences, accounting for natural processes like incomplete lineage sorting, long branch attraction, gene tree coalescence, and attacking questions of species delimitation, all while incorporating demographic parameters. However, dealing with introgression considerably increases the complexity of the comparatively few models that accommodate introgression, in a limited number of analytical packages capable of dealing with this process (Leaché *et al.*, 2013). A number of these methods have recently become available, like PHRAPL (Jackson *et al.*, 2017), IMa3(Hey *et al.*, 2018), and PhyloNetworks (Solís-Lemus *et al.*, 2017), but they were developed for multilocus sequence data and are not suitable for SNPs or even for a matrix of short reads such as ours.

To our knowledge, the most suitable model to infer the history of relationships among populations and simultaneously account for admixture events using SNP data is that implemented by TreeMix. Still, in our application of this method the results were variable and difficult to interpret (Fig. 9). TreeMix inference decreases in accuracy when true migration

happens between closely related populations, incorrectly inferring introgression into a third population if that population is, in reality, simply exchanging migrants with another population that is, in fact, exchanging migrants with the original population (e.g. A⇔B⇔C may infer A⇔C) (Pickrell & Pritchard, 2012). To deal with these issues other authors have fixed the tree topology based on a previously-determined, known species tree, and excluded highly admixed samples (more then 20% ancestry from 2 or more lineages) that overwhelmed the model (Puckett *et al.*, 2016). In our case, we still do not have a definite, well-supported species tree and it is understanding the origins of those admixed populations themselves that is our primary focus at this point. Therefore, we view the results of our TreeMix analyses as an exploratory method that points to interesting hypotheses for potential pairs of admixed lineages that can be further tested. However, our this system is far too complex given our sparse sampling of these Guiana Shield lineages for TreeMix to provide clear insight into the origins of these taxa.

So our large dataset with presumed incomplete taxon sampling (clearly there are many undescribed species, many more than we were able to sample) placed us in a paradoxical loop: on one hand we are not able to estimate a species tree for all populations because the models assume no introgression; on the other hand, many models to infer introgression history require prior knowledge of the species tree for adequate interpretation of results. We dealt with this problem by using the population assignment model of Structure for the entire Amazon clade to generate hypotheses of potentially admixed populations (Fig. 10). We then excluded these populations to run new phylogenetic analyses assuming our reduced dataset has a better fit to the inference models (having removed samples providing direct evidence of introgression), and checked for consistencies in topology and branch support across methods. Finally, we ran four-population tests that use a less-parameterized model and are robust in rejecting hypotheses of

data that do not fit the expected tree in the absence of migration. With this combination of methods we believe we established a well-supported tree for populations with little or no admixture, which will allow formal hypotheses testing for intensity and direction of introgression among populations within particular drainages given adequate sampling of each population/species.

Hypancistrus are widespread across the Amazon Basin, with populations found in both large and small tributaries, but not in the Amazon River. To the North of the Amazon we sampled rivers draining from the Guiana Shield, including *H. margaritatus* from the Takutu River, *H*. sp. (L136) from the Negro River, *H*. sp. (L499) from the Paduari River, *H*. sp. (L500) from the Uatumã River, and *H*. sp. (L316) and *H*. sp. (411) from the Jari River, and these all share a recent common ancestor. To the south of the Amazon, draining from the Brazilian Shield, we sampled *H*. sp (Jacareacanga) and *H*. sp. (L411) from the Tapajós River; *H. zebra*, *H*. sp. (L66/333), and *H.* sp. (L174) from the Xingu River, and *H*. sp. (L270) from the Curuá-una River. Although we have sampled eight out of the nine described species, there are still many gaps in our sampling relative to the total diversity of the genus, and we would be particularly interested to see how undescribed lineages from the Tapajós, Purus, and Madeira Rivers would fit into our findings.

In these rivers south of the Amazon we recover strong support for the grouping of the spotted *H.* sp. (Jacareacanga) from the Tapajós River with the Xingu's *H. zebra*. As the rivers draining from the Brazilian Shield are ancient, an ancestral lineage has likely dispersed from one drainage to the other. As for the route of dispersal, this may have resulted from headwater capture in which a small, upper tributary of one of these rivers changed course and began draining into the adjacent river, taking with it the species inhabiting that tributary. Alternatively,

it may be that during periods of reduced flow water quality changes due to climatic instability in the Quaternary, it may have been possible for ancestral populations to disperse from one river to the next through the Amazon River itself (Höppner *et al.*, 2018). The second species from the Tapajós River, *H*. sp. (L260) is possibly a hybrid between *H.* sp. (Jacareacanga) and the Guiana Shield drainages clade, based on the intermediary position in the PCA (Fig. 8) and in the admixed assignement in the Structure analysis (Fig. 10). Interestingly, *H*. sp. (L260) has a worm-lined color pattern well defined by a strong contrast between black and white lines, as opposed to less defined light spots on *H.* sp. (Jacareacanga). Other lineages from the Guiana Shield drainages we sampled present worm-lined patterns, but none with as thin and well-defined lines. Therefore, it seems the color in *H*. sp. (L260) is not an intermediary state between putative parental lineages.

The lineages from the Guiana Shield drainages are so closely related they are not recovered as distinct clusters in the PCA and Structure analyses (Fig. 8 & 10). They are spread across a very large area of the Amazon Basin, from the headwaters of the Branco River drainage all the way to the lower Amazon in the Jari. *Hypancistrus* are territorial fish that typically occupy small home ranges (Leandro Sousa, personal observation), making this broad distribution particularly intriguing. We were able to confidently establish *H*. sp. (L411), which occupies the Jari River, as the lineage sister to all others of the Guiana Shield drainages clade. All analyses that excluded admixed lineages recovered *H*. sp. (L316) as the sister clade to that comprised of *H*. sp. (L136), *H. margaritatus*, and *H*. sp. (L500) (Fig. 11). Though the three possible topologies within this clade all pass the four-population test, that with *H*. sp. (L136) as the sister lineage to *H. margaritatus* + *H*. sp. (L500) presented the highest p-value (Table 5). We also found support for a hybrid origin of *H*. sp. (L499) from the Paduari River, a tributary of the Negro River, as

evidenced by the TreeMix and Structure analyses (Fig. 9 & 10). The parental lineages are *H. inspector* (14 – 23% ancestry) and a Guiana Shield drainages population (86 – 77% ancestry), with *H*. sp. (L136) of the Negro River a likely source due to its geographical proximity. In this case it is noteworthy that both parental taxa have a spotted color pattern, while *H*. sp. (L499) presents a striped pattern suggesting the possibility of incomplete sampling of one parental lineage or rapid selection for divergent coloration. The ML trees (Fig. 3 & 5) support the monophyly of *H*. sp. (L499) and *H*. sp. (L136), but sampling more populations of *Hypancistrus* from the Guiana Shield drainages would be necessary to elucidate relationships within this clade. We hypothesize the Guiana Shield drainages clade is young and diversified very recently. We believe the populations we sampled are independent evolutionary lineages, but species delimitation analyses along with morphological assessment of multiple samples per population will be imperative for dependable description of new species in this group.

The populations of *Hypancistrus* from Xingu River present a complex genetic structure. Our findings support the basal position of *H*. sp. (L174) in relation to the other Amazon Basin taxa (Fig. 11D; Table 5 – test 7) , but the inclusion of *Hypancistrus* lineages we did not sample in this study may affect this conclusion. The biogeographic scenario for the divergence between *H*. sp. (L174) and the other Amazon Basin species remains uncertain, though, if the position of *H*. sp. (L174) as a sister clade to other Amazon lineages is true, we speculate there was an early vicariance event that separated *H*. sp. (L174) in the Xingu from the Tapajós lineage. After this time, the Tapajós lineage would have dispersed to the Xingu River, founding the population that would become *H. zebra*. More recently, the Guiana Shield drainage lineage colonized the lower Xingu from the Amazon River giving origin to *H*. sp. (L66/333). Alternatively, *H*. sp. (L66/333) diverged from *H*. sp. (L174) in the Xingu, and received introgression from the Guiana Shield

drainage lineage in the lower portion of the Xingu River. We have evidence for admixture between *H*. sp. (L66/333) and *H*. sp. (L174), and possibly between *H*. sp. (L174) and *H. zebra*. Population genetics and hypotheses of hybridization events in the Xingu River will be further explored in Chapter 3.

CONCLUSIONS

In this study, we present the first phylogenetic reconstruction of the genus *Hypancistrus* based on genomic SNP data. We found two major clades corresponding to the Orinoco and Amazon Basins, and find *H. inspector* from the Casiquire River to be an admixed species with genetic influence from both basins. Our results suggest the synonymization of two species (*H. contradens* and *H. lunaorum*) in the Orinoco clade, but confirm the monophyly of the remaining species. Within the Amazon Basin, the major lineages correspond to Amazon River tributaries flowing from the Guiana Shield in the North, and from the Brazilian Shield in the South. We report evidence of admixture in lineages native to the Xingu River, the Tapajós River, and the Paduari River. The inclusion of samples donated by the aquarium trade had considerable impact on the relevance of our sampling and our ability to reconstruct the evolutionary history of the genus, allowing us to include several undescribed lineages of *Hypancistrus*. Nevertheless, important gaps in our sampling ensure this will not be the final word on *Hypancistrus* phylogeny and evolutionary history. The complexity of demographic histories in this system poses important challenges and highlights limitations in our ability to infer phylogenetic relationships and demographic history in species with a history of admixture, but we believe we have made important advances towards understanding the evolution of the genus *Hypancistrus*. With this

study we propose several new hypotheses regarding the relationships among *Hypancistrus* lineages and highlight their complex biogeographic history.

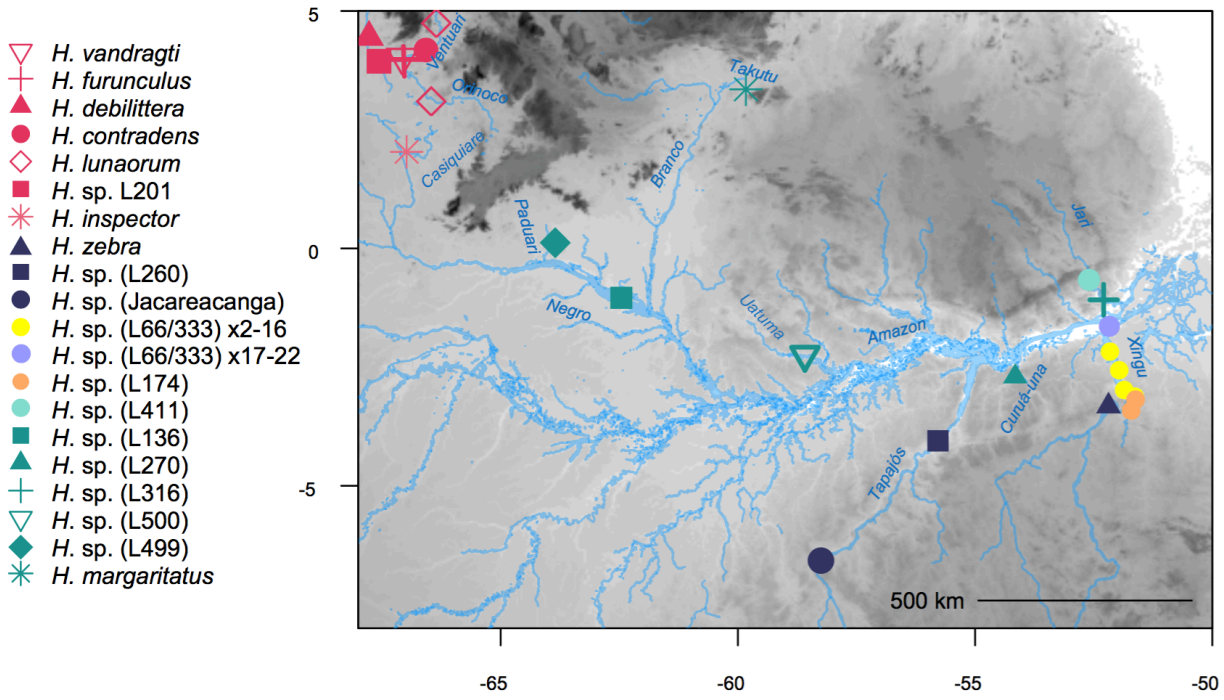


Figure 2. Map of *Hypancistrus* sampled in the Amazon and Orinoco Basins.

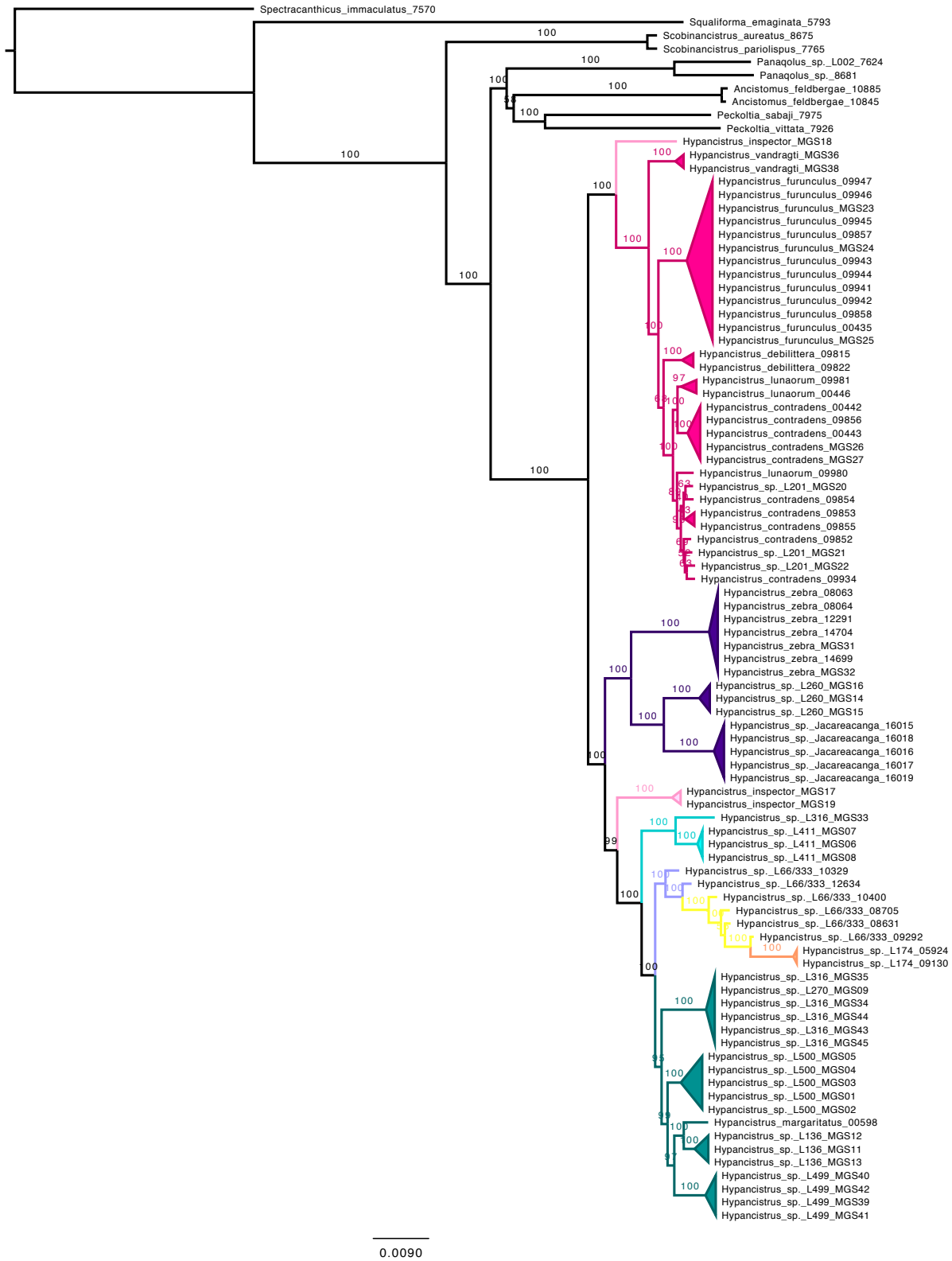


Figure 3. IQ-Tree phylogeny of *Hypancistrus* based on maximum likelihood of 9125 concatenated loci including 92 individuals and 28 species (dataset 1, see Table 4). Support is indicated by FastBootstrap values (≥ 95 indicate good support.

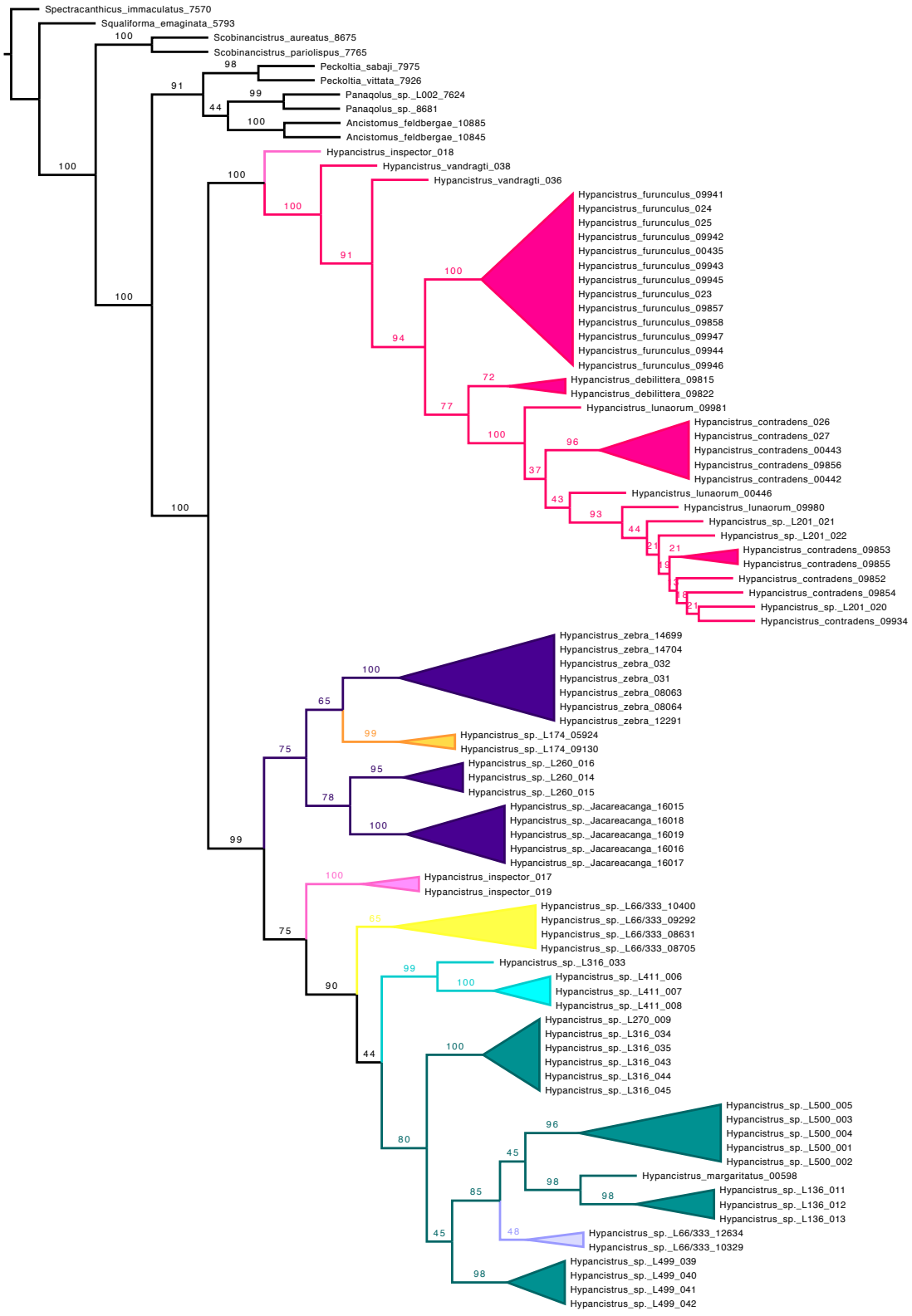


Figure 4. Phylogeny of *Hypancistrus* based on coalescent method of SVDquartets analysis of 9018 unlinked SNPs including 92 individuals and 28 species (dataset 2, see Table 4). Support is indicated by bootstrap values.

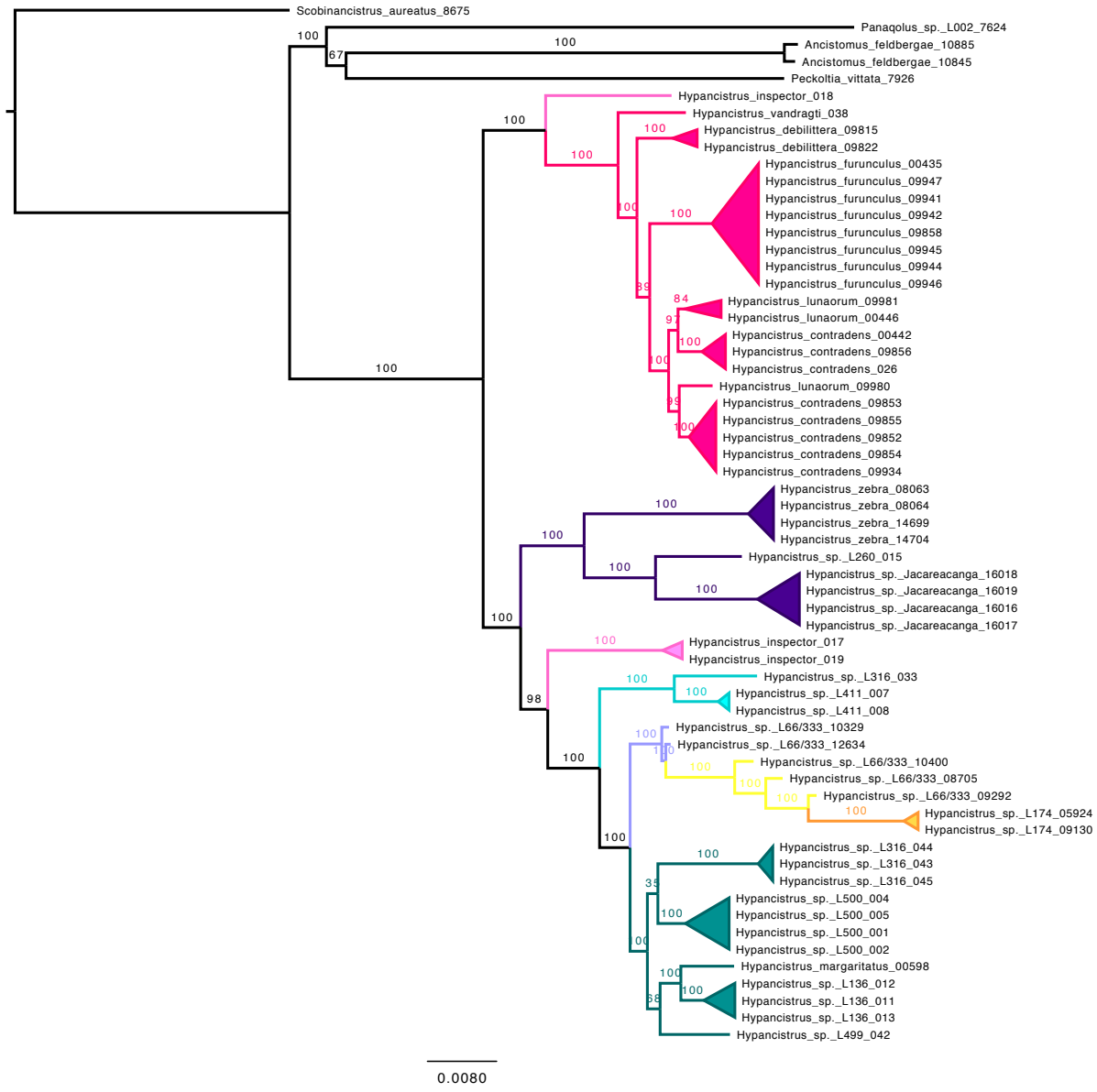


Figure 5. IQ-Tree phylogeny of *Hypancistrus* based on maximum likelihood of 11449 concatenated loci including 61 individuals with high coverage sequencing and 22 species (dataset 3, see Table 4). Support is indicated by FastBootstrap values (≥ 95 indicate good support).

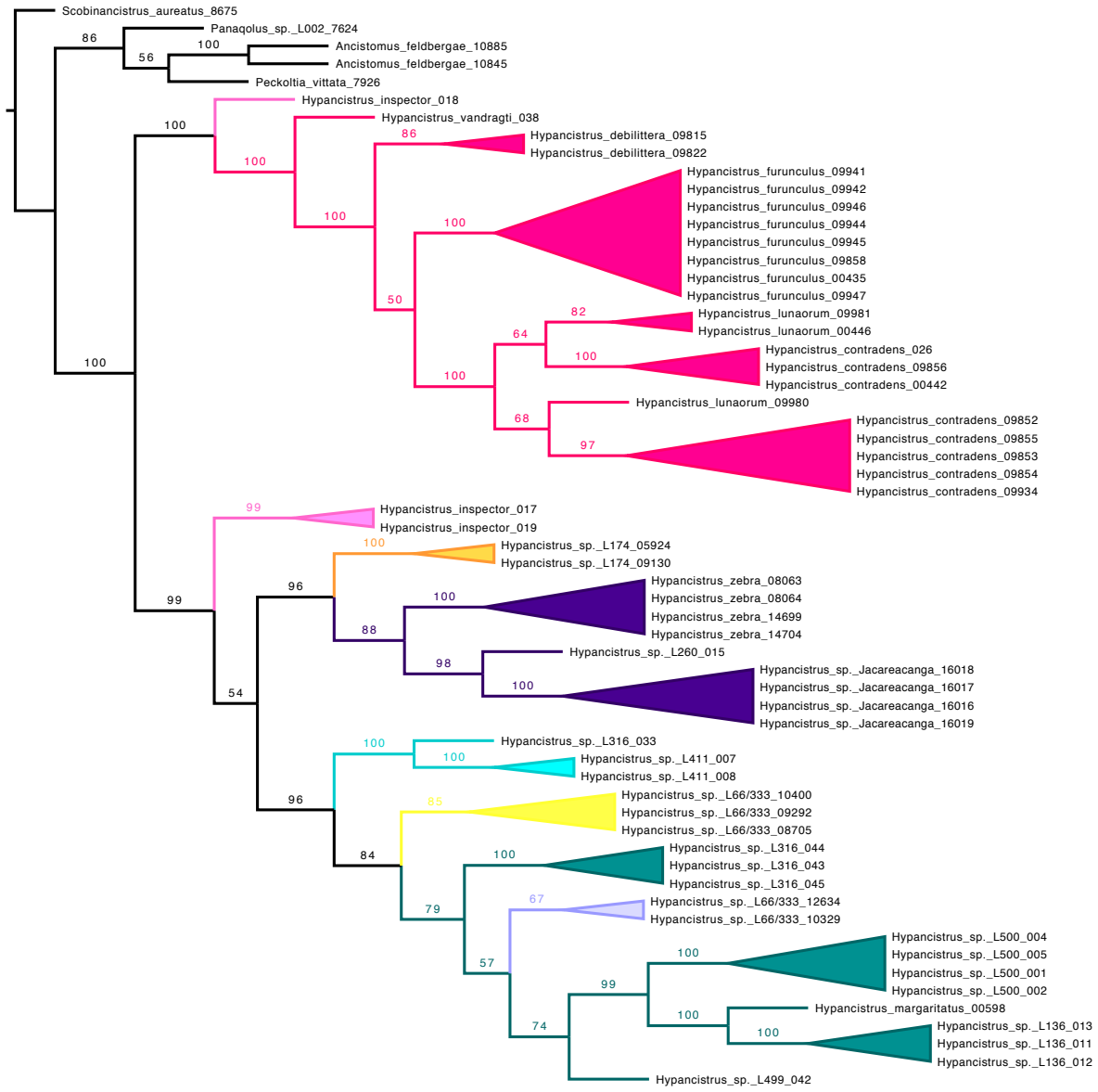


Figure 6. Phylogeny of *Hypancistrus* based on coalescent method of SVDquartets analysis of 11153 unlinked SNPs including 61 individuals with high coverage sequencing and 22 species (dataset 4, see Table 4). Support is indicated by bootstrap values.

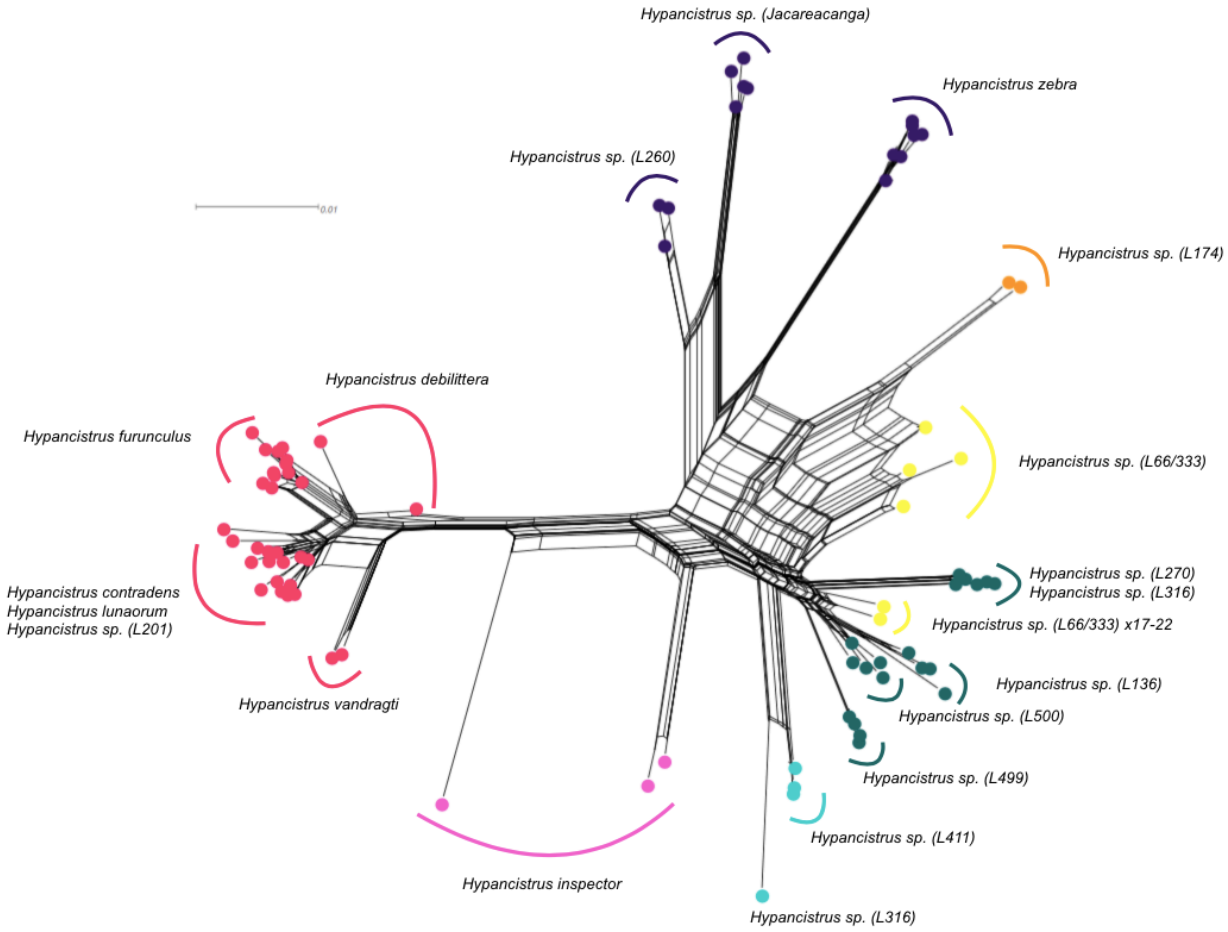


Figure 7. Network analysis of *Hypancistrus* based on SplitsTree of 9018 concatenated loci including 92 individuals and 28 species. Outgroups were excluded from graphic for clarity.

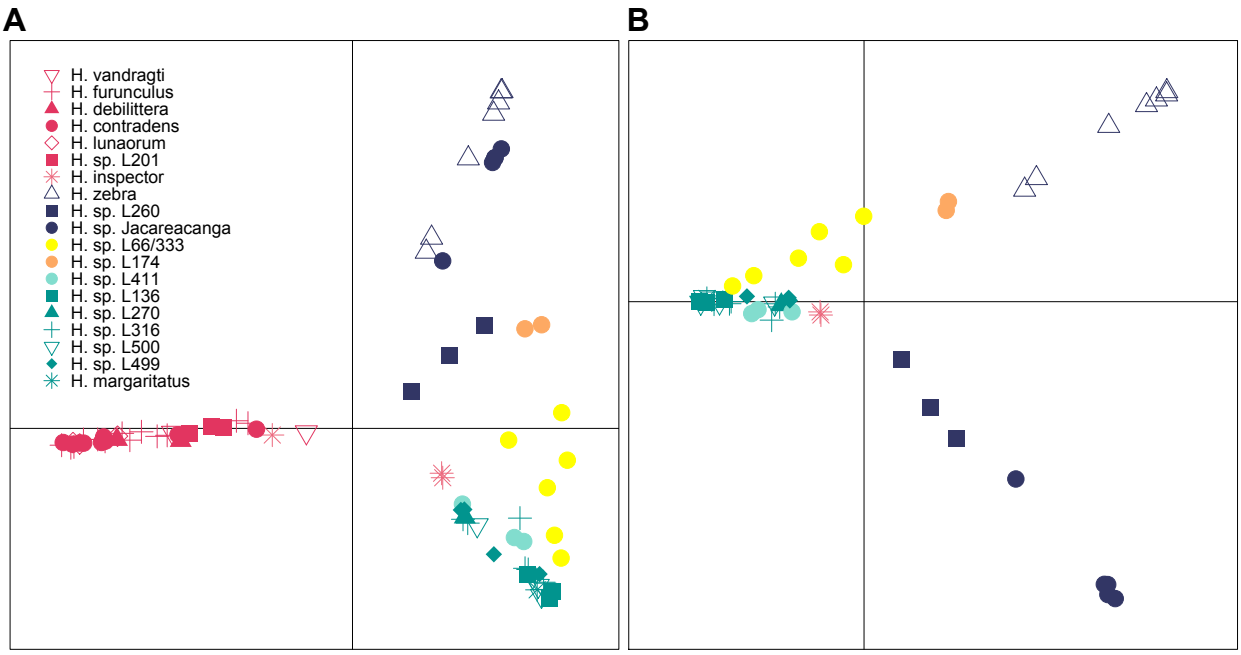


Figure 8. Principal components analysis of *Hypancistrus* based on 9018 SNP loci. A. *Hypancistrus* lineages from the Amazon and Orinoco Basins. B *Hypancistrus* lineages from the Amazon Basin.

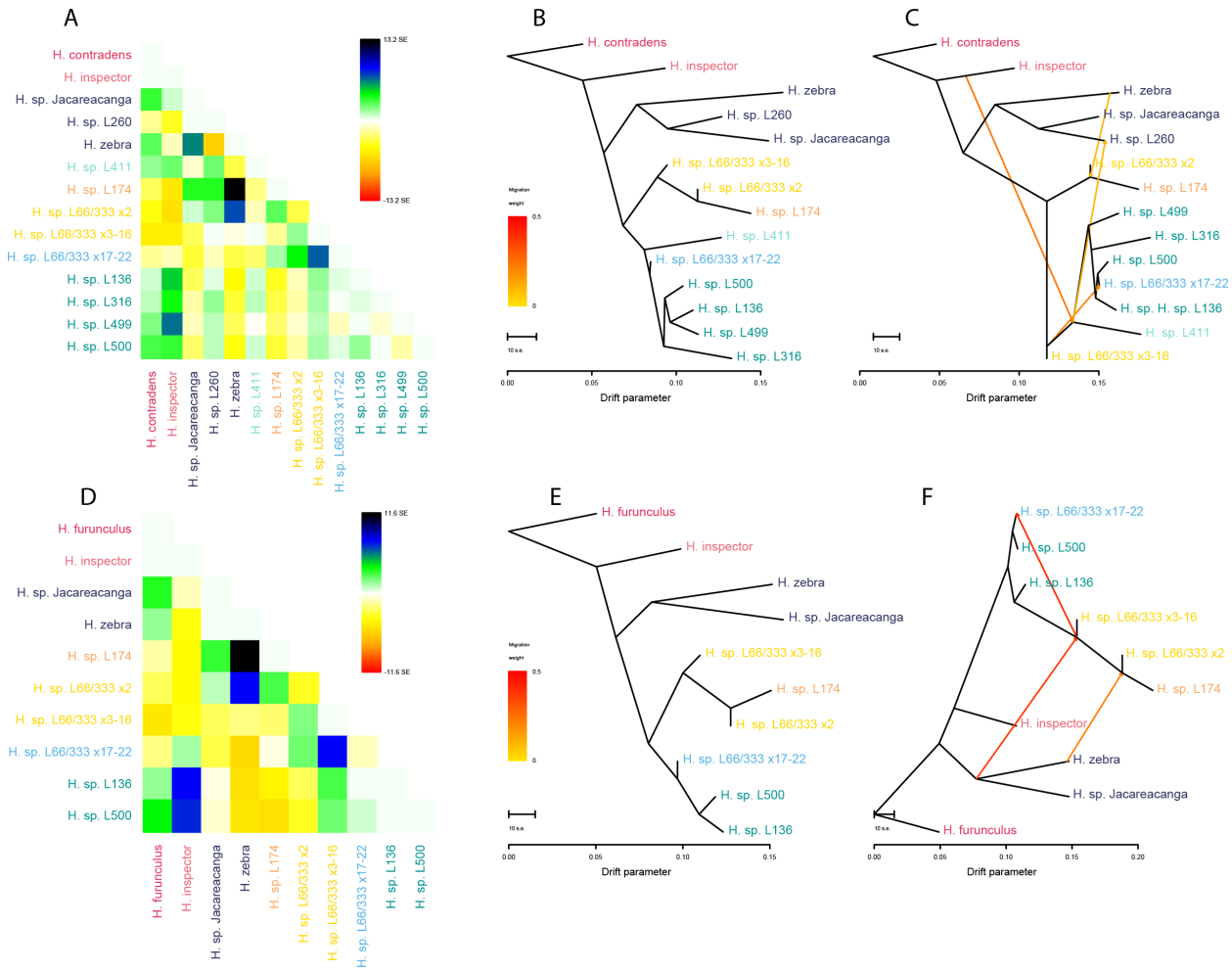


Figure 9. TreeMix analysis of *Hypancistrus* populations from the Amazon Basin. A-C refer to dataset 5 and D-F refer to dataset 6 (see Table 4). A and D represent the matrix of residual fit from the maximum likelihood trees without migration depicted in B and E. Large positive residuals (black and blue colors) indicate poor fit of the data to the model, pointing to likely candidates for pairs of populations that have been admixed. C and F show maximum likelihood trees allowing for respectively four and three migration edges. Color of edges indicate their weight.

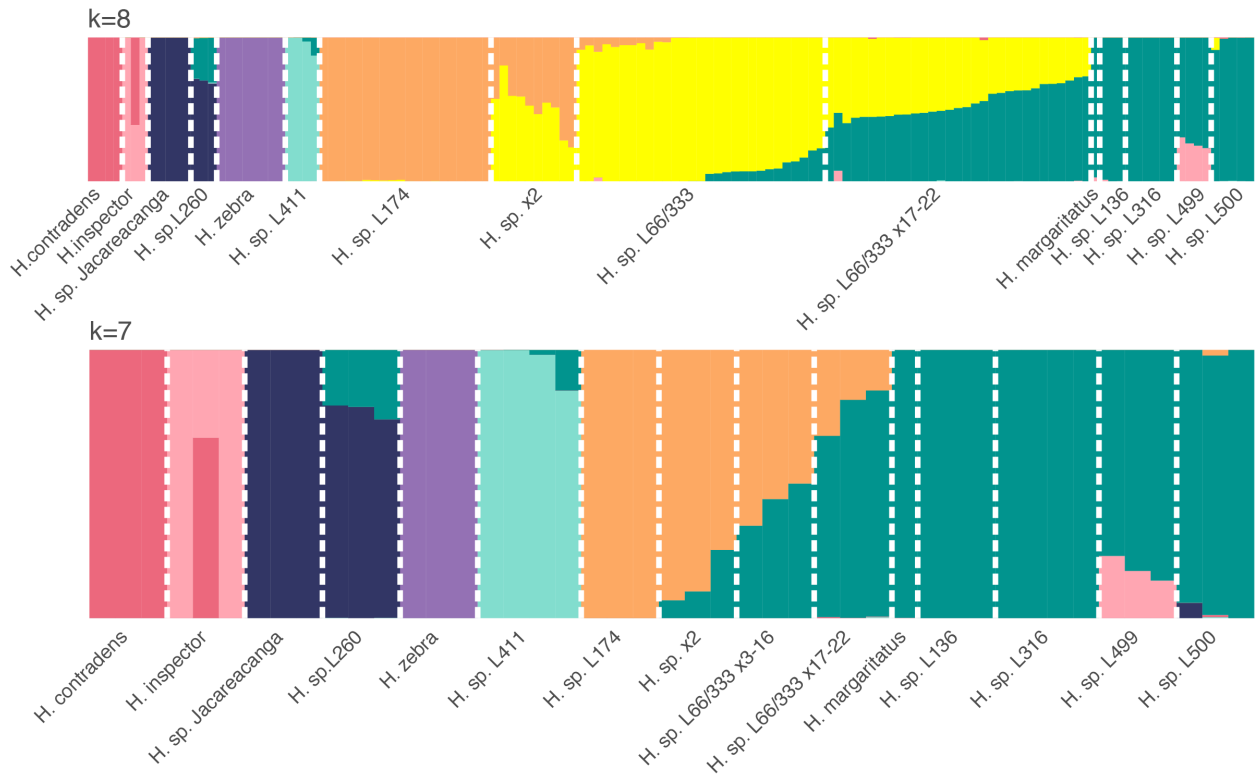


Figure 10. Population assignment plots for best k for Structure analyses of *Hypancistrus* from the Amazon Basin based on 11670 SNP loci. Top plot shows analysis including 136 individual samples and bottom plot represents the analysis with 45 samples to reduce variance in sampling among populations.

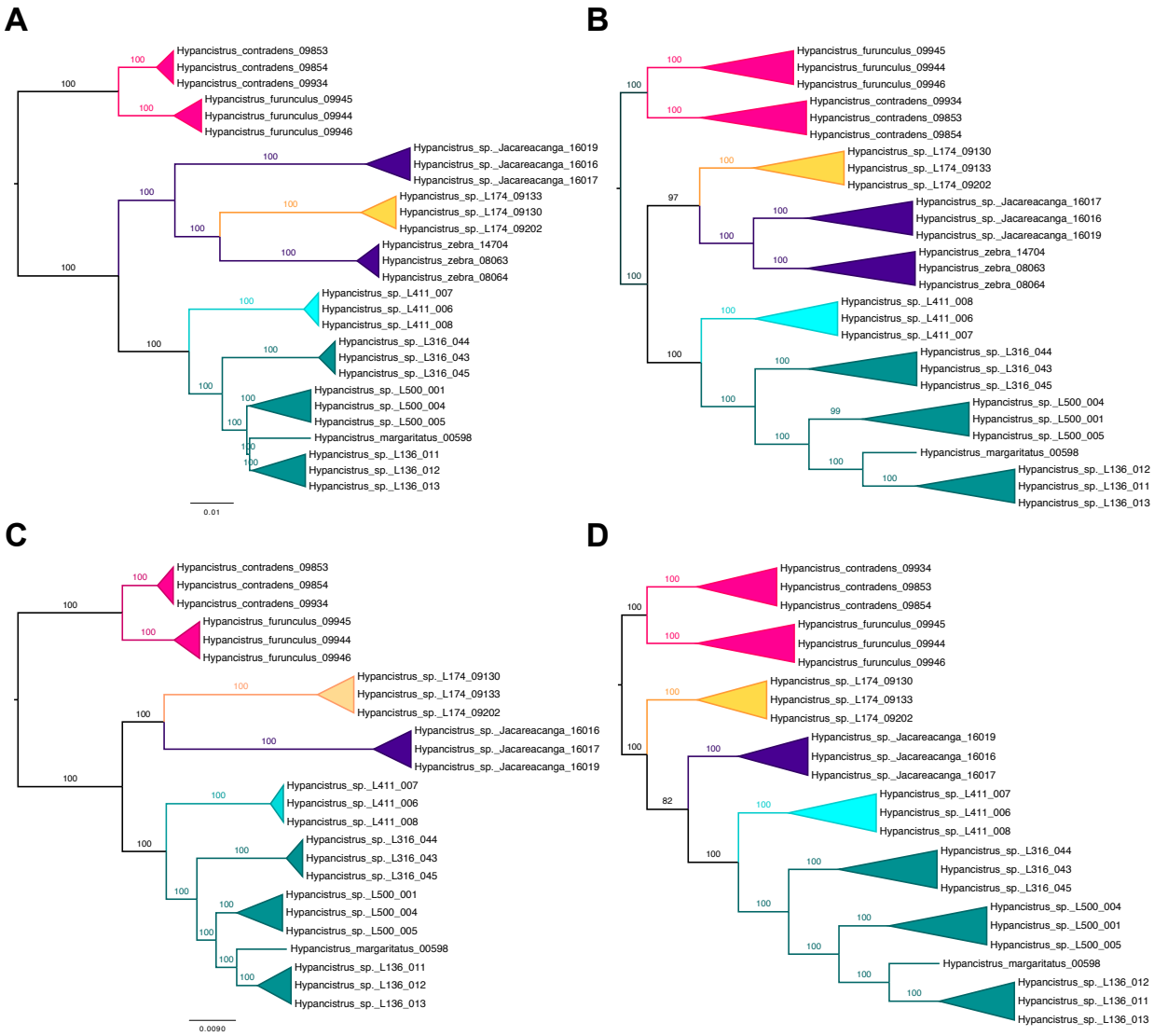


Figure 11. Phylogenies of *Hypancistrus* from the Amazon Basin excluding populations with evidence of admixture (A and B) and additionally excluding *Hypancistrus zebra* (C and D). A and C are IQ-Tree phylogenies based on maximum likelihood of respectively 13328 and 12992 concatenated loci (datasets 8 and 10, see Table 4). Support is indicated by FastBootstrap values (≥ 95 indicate good support). B and D are phylogenies based on coalescent method of SVDquartets analysis of respectively 11052 and 10438 unlinked SNPs (datasetd 9 and 11, see Table 4). Support is indicated by bootstrap values.

CHAPTER 3: EVOLUTION OF PLECO CATFISHES BELONGING TO THE GENUS *HYPANCISTRUS* IN THE HIGHLY IMPERILED XINGU RIVER, BRAZIL

INTRODUCTION

Freshwaters across the globe have been fertile systems to study various bigeographic patterns and processes at large geographic and evolutionary scales. However, effects of fine-scale features of the freshwater landscape in generating and maintaining genetic diversity within species are poorly understood (Paz-Vinas *et al.*, 2015). As freshwater bodies, and rivers in particular, face imminent threats of habitat loss and exploration/exploitation (e.g. fishing, mining, pollution) it is urgent to establish management strategies based on scientific evidence to ensure diversity is protected across all levels, along with the processes that generate and maintain it. The Amazon Basin is among the largest and most species-rich freshwater systems on Earth and traverses a course that, for the most part, lies upon relatively flat lowlands (Albert & Reis, 2011b). Its waters drain from the Andean foothills in the east, Guiana Shield to the north, and Brazilian Shield to the south (Irion & Kalliola, 2010), an area spanning some 6.3 million km^2. The course of the rivers draining the Amazon Basin contain characteristic rapids and waterfalls found in the areas in which their courses transition from these higher altitude areas to the Amazon lowlands, and it is the impact of these complexifying features of the riverscape on biotic diversity that we sought to explore in this study.

Among the largest of the Amazonian tributaries, accounting for roughly 5% of the water in the basin, is the Xingu River. With headwaters in the Brazilian shield, the Xingu River flows

northwards until it meets the Amazon River, about 350 km before enters the Atlantic Ocean (Fig. 12). It has clear waters characterized by low sediment and high transparency due to the crystalline rocks typical of its catchment area (Albert & Reis, 2011b). Stretching more than 1,600 km, the course of this river makes a very characteristic shift as it leaves the Brazilian Shield. Here the Xingu River makes a sharp inflection, turning back upon itself to the south, then back to the north, in an area known as Volta Grande (Big Bend). In this area where it leaves the Brazilian shield the river suddenly drops 90 m over the course of 130 km, in a series of rapids, waterfalls, and anastomosing channels (Sawakuchi *et al.*, 2015). Downstream of Volta Grande, the mouth of the Xingu River opens into a wide channel when it enters the Amazon Basin, flowing for a stretch of 180 km before joining the Amazon River. This lower portion, known as the Xingu Ria, lays over a sedimentary basin with predominantly sandy substrate but still containing patchy remnants of rocky substrate (Sabaj-Pérez, 2015; Sawakuchi *et al.*, 2015). The Xingu River harbors a rich and unique biota with over 450 fish species, many of which are endemic to the Volta Grande (Zuanon, 1999; Camargo *et al.*, 2004). Among this incredible fish diversity, many are species that have been described only recently, a testament to just how poorly known the region remains (e.g. Netto-Ferreira & Moreira, 2018; Sousa *et al.*, 2018; Silva-Oliveira *et al.*, 2019).

Among the most remarkable groups of animals of the lower Xingu are the armored catfishes, or plecos, of the South American family Loricariidae (Sabaj-Pérez, 2015). These fishes are notorious for their exuberant coloration, making them highly sought after in the aquarium trade. Among this diverse family, *Hypancistrus* is a genus of loricariid catfishes that is distributed across the Amazon and Orinoco Basins (Isbrücker & Nijssen, 1991; Armbruster *et al.*, 2007; Tan & Armbruster, 2016). These species specialize in living in fast-flowing

environments with rocky substrates. In the Xingu river, there is one described and two undescribed *Hypancistrus* species known (Fig. 12). The only described species in the Xingu, *Hypancistrus zebra* (Isbrücker & Nijssen, 1991), is a boldly-patterned, black and white striped favorite of fish collectors that is endemic to the Volta Grande and considered a critically endangered species by the Brazilian authorities. The undescribed taxa known from this river include a species exhibiting a much finer, worm-lined pattern known in the aquarium trade as *H.* sp. (L66/333) or by the popular name king tiger pleco, which is found in both the Volta Grande and the Xingu Ria. The third species of Xingu *Hypancistrus* is a spotted species known as *H.* sp. (L174), also found only in the Volta Grande.

While the Volta Grande endemics *H. zebra* and *H.* sp. (L174) exhibit little variation in their color pattern, the more abundant and widespread *Hypancistrus* sp. (L66/333) presents intra-specific polymorphisms in color pattern, with background color ranging from white to tan and delineation of their pattern varying dramatically. This variation within such a widespread taxon initially led to the conclusion that this represented more than one species, hence the two different L numbers (66 & 333). Recently though Cardoso *et al*. (2016) analyzed cytogenetic and mitochondrial DNA data of the two phenotypes *H.* sp. (L66) and *H.* sp. (L333), finding no chromosomal or DNA sequence differences between these phenotypes other than a polymorphism at chromosome 21 unrelated to color pattern or sex (Cardoso *et al.*, 2016). The color pattern variation in *H.* sp. (L66/333) contrasts with the lack of differentiation in other morphological traits, highlighting the importance of using adequate molecular markers in order to detect potential population structuring in variable species such as this.

The geological characteristics that make the Xingu River's Volta Grande such a unique environment have attracted dam builders and the Brazilian government to explore its

hydroelectric power. Two main dams, Pimental and Belo Monte, located respectively at the beginning and end of Volta Grande were constructed between 2011 and 2016 (Nadilo, 2012) and went online in 2016. The Pimental dam deviates the river's course, flooding approximately 382 km^2 of the Volta Grande area and dewatering most of the rapids (Sawakuchi *et al.*, 2015). Fishes of the Xingu River including *Hypancistrus* species are threatened by the direct impact of loss of habitat of the dewatered stretch (Fitzgerald *et al.*, 2018). The indirect effects of substrate, water chemistry, and flood regimes alterations in the area below Belo Monte dam are less clear (Sawakuchi *et al.*, 2015), and *Hypancistrus* is an excellent candidate taxon to monitor as these changes occur, due to its distribution in both the Volta Grande and the Xingu Ria.

In Chapter 2 we explored the relationships among *Hypancistrus* from the Orinoco and Amazon Basin. Our findings indicate that the Xingu River is an exceptionally interesting drainage, encompassing three distinct lineages of *Hypancistrus*, that show signs of a complex history of genetic structure and introgression. The goal of this study is to investigate the evolution of *Hypancistrus* in the Xingu River. We collected samples along the Volta Grande e Xingu Ria prior to the closure of the Belo Monte dams, making this a unique study in reconstructing the effects of such complex rapids on species that rely on these features of the riverscape. We used phylogenomic and population genomic tools to unravel the relationships among the species of *Hypancistrus* inhabiting the Xingu River in order to understand the evolutionary history of a species assemblage within a major Amazonian tributary while sampling in a way that allows us to explore fine-scale population structure in a way that has rarely been done in a tropical river system. We described the patterns of genetic structure of *H.* sp. (L66/333) also tested hypotheses of historical introgression among populations to address the hypotheses

generated by our findings in Chapter 2. Finally, we discuss the implications of our findings to species delimitation in this system.

MATERIALS AND METHODS

Sampling and library prep

A total of 153 tissue samples of *Hypancistrus* sp. (L66/333) spread across 22 sampling sites, 21 samples of *Hypancistrus* sp. (L174) from two sampling sites, and six samples of *H. zebra* from two sampling sites were obtained from the Ichthyology Collection of the Academy of Natural Sciences of Drexel University – ANSP (Fig. 12). The disproportionate sampling of *Hypancistrus* sp. (L66/333) is due in part to their much broader distribution and higher abundance, but also the focus of collectors on this species and the protected status of *H. zebra*. The majority of these samples (172) were collected during a series of expeditions between 2012 and 2014 by a group of international collaborators intent upon surveying fishes and bivalves of the Xingu River prior to the closure of the Belo Monte complex dams. Additionally, depending on the analysis, samples of *H. contradens* or *H. furunculus* from the Orinoco River were used as an outgroup. To investigate possible introgression between *Hypancistrus* from the Xingu River and other species of *Hypancistrus* from the Amazon Basin we sampled *H*. sp. (L136) from the Negro River, *H*. sp. (L316) from the Jari River, *H*. sp. (L500) from the Uatumã River, and *H*. sp. (Jacareacanga) from the Tapajós River.

DNA was extracted following a salt-extraction protocol modified from Aljanabi & Martinez (1997) and concentration of DNA was quantified with a Qubit® Fluorometer (Thermo Fisher). Reduced-representation libraries were prepared following a modified version of the double digest restriction associated DNA sequencing (ddRAD, Peterson *et al.*, 2012). In this

method total genomic DNA was digested using a pair of restriction enzymes (EcoRI and XbaI), adding an additional enzyme (NheI-HF) with the goal of cutting adapter dimers (Glenn *et al.*, 2017). Digestion was followed by the steps of adaptor ligation, PCR, and clean-up. Unique pairs of barcodes were added during the PCR step to allow for sample pooling and post-sequencing assignation of reads to individual samples. DNA fragments of 324 to 416 base pairs were selected by a Pippin Prep (Sage Science). Library quality and quantification was obtained with qPCR in a Rotor-Gene Q (QIAGEN) using the KAPA Library Quantification kit for Illumina platforms (Kapa Biosystems). Sequencing was performed with an Illumina NextSeq 500 with single-end reads at the National Center for Natural Products Research at the University of Mississippi.

Sequence processing

All sequence processing and downstream analysis was performed using the Mississippi Center for Supercomputing Research (MCSR) supercomputing clusters, unless stated otherwise. Demultiplexing, the assignment of reads to individuals based on unique barcodes added to each library during PCR, was done using the bcl2fastq software (available at support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html). We allowed for a maximum mismatch of two base pairs (bp) in barcodes, and trimmed all reads of restriction site overhangs, resulting in a final fragment length of 56 bp for all reads. Fragments were aligned and single nucleotide polymorphisms (SNPs) were extracted using the ipyrad software v. 0.7.28 (Eaton & Overcast, 2019). We decided on a combination of filters that selected high-quality reads to maximize the selection of homologous loci (Chapter 2). In the first filtering step we defined base calls with phred Q (quality) score of less then 33 as ambiguous and

reads with ambiguous sites were excluded. In the within-sample clustering step we allowed for a clustering threshold for *de novo* assembly of 0.85 and a maximum of two unique allleles per locus, since these are diploid species (da Silva *et al.*, 2014). Consensus sequences with more then two uncalled bases or more then eight heterozygous sites were excluded and we required a minimum depth for base calling of ten reads. For among-sample assembly we required a clustering threshold of 0.85, minimum of 50% of samples represented per locus (maximum 50% missing data per locus), and maximum of ten SNPs, four indels, and 50% of shared polymorphic sites per locus. To generate final output files we used one of two approaches: 1) complete sequences were concatenated in a supermatrix; or 2) one SNP per locus was selected to build a matrix of puttatively unlinked SNPs.

Shared locus selection with *de novo* asembly methods is intrinsicaly dependent on which samples are used, therefore we ran step six of ipyrad (clustering reads among samples) separetly for each one of the different sampling combinations we used (Table 6). In datasets 1, 2 and 3 we used all samples of *Hypancistrus* from the Xingu River we had available. Dataset 4 is a subsample (28 samples) of dataset 7 in Chapter 2 (136 samples), for which we ran ipyrad including *Hypancistrus* from the Xingu River and the additional samples of *Hypancistrus* from other Amazon Basin drainages (ten species), with *H. contradens* as the outgroup. For dataset 5 we ran ipyrad with three samples of *H*. sp. (L174), *H*. sp. (L66/333), and *H.* sp (L316), which generated 16,465 loci, and selected 10,000 loci for our analyses.

Table 6: Summary of analyzed datasets of Xingu *Hypancistrus* in relation to samples included, number of loci, and percentage of missing data.

	Species; N samples	Data matrix	Total loci	% missing data	Analysis
1	*H*. sp. (L174); 21 *H*. sp. (L66/333); 153 *H. zebra*; 6 *H. furunculus*; 1	concatenated	13230 (759380 bp)	20.0%	IQ-TREE
2		unlinked SNPs	10984	17.7%	SVDquartets
3	*H*. sp. (L174); 21 *H*. sp. (L66/333); 153 *H. zebra*; 6	unlinked SNPs	11013	18.1%	Network PCA Structure
4	*H*. sp. (L174); 3 *H*. sp. (L66/333); 9 *H. zebra*; 3 *H*. sp. (L136); 3 *H*. sp. (L316); 3 *H*. sp. (L500); 3 *H*. sp. (Jacareacanga); 3 *H. contradens*; 1	unlinked SNPs	11670	10.9%	compD
5	*H*. sp. (L174); 3 *H*. sp. (L66/333); 3 *H*. sp (L316); 3	full sequences without concatenation	10000 (567908 bp)	14.4%	BPP

Phylogenetic and network analyses

A maximum likelihood (ML) phylogenetic tree was estimated with IQ-TREE v. 1.6.10 (Nguyen *et al.*, 2014) in the CIPRES Science Gateway portal v. 3.3 (Miller *et al.*, 2010). We ran IQ-TREE using dataset 1, with complete sequences concatenated in a supermatrix, assigning *H. furunculus* as the outgroup (Table 6). The automatic model selection method ModelFinder implemented by IQ-TREE was used to determine the best-fit substitution model based on the Bayesian information criterion. The transversion model and unequal base frequencies estimated empirically from the alignment with the FreeRate model of heterogeneity across sites with three categories (TMV + F + R3) was selected as best-fit model for our dataset. We ran 1,000 replicates for Ultrafast bootstrap branch support estimation, optimizing tree search by a nearest

neighbor interchange search, which reduces overestimation of branch supports (Hoang *et al.*, 2017).

Relationships among individual samples were also estimated with the coalescent model used by single value decomposition scores for species quartets SVDquartets (Chifman & Kubatko, 2014, 2015) implemented in PAUP* v. 4.0 (Swofford, 2003). We used a matrix composed of unlinked SNPs (dataset 2; Table 6) to sample 10 million quartets, representing 23.1% of the total number of possible distinct quartets for our dataset. Similar to the ML analyses, we set *H. furunculus* as the outgroup. Branch support was obtained with 100 bootstrap replicates and results were written onto the 50% majority rule consensus tree.

To allow for identification of reticulation events like gene flow and the detection of conflicting phylogenetic signals we built a phylogenetic network using SplitsTree (Huson & Bryant, 2006). We used the UncorrectedP distance method to calculate genetic distances between samples and the NeighborNet distance method to compute the network. The matrix of unlinked SNPs in dataset 3 without an outgroup was used in this analysis (Table 6).

Principal components analysis

To explore how samples are distributed within the genetic variance space we ran a principal components analysis (PCA) using dataset 3 with unlinked SNPs (Table 6), as implemented by the adegenet package v. 2.1.1 (Jombart & Ahmed, 2011) and the ade4 package v. 1.7.13 (Dray & Dufour, 2007) in R v. 3.6.1 (R Development Core Team, 2019). This is a good exploratory method, as non-parametric multivariate methods are robust against varying choices of SNP calling and filtering in reduced representation library datasets (Linck & Battey, 2018). We replaced missing data by the mean allele frequencies as suggested by adegenet's developer

(Jombart, 2008). We examined the three first principal components and plotted the results from the first two axes, grouping samples based on species identification and sampling locality for *H.* sp. (L66/333).

Population structure

We inferred genetic structure among populations from the Xingu River with fastStructure v. 1.0 (Raj *et al.*, 2014) which builds a model approximate to Structure (Pritchard *et al.*, 2000) that allows for the use of thousands of genome-wide SNPs. This method uses allele frequency data to assign a probability of identity of each sample to membership in one of k groups. We used the matrix of 11,013 unlinked SNPs including 180 samples of *Hypancistrus* from the Xingu River (dataset 3, Table 6). We tested the clustering individuals in one to ten genetic groups (k = 1 to 10), running five replicates of each k to check for consistency and using the simple prior. We ran the choosek script included in fastStructure's package to select the best k based on two criteria, according to Raj *et al.* (2014): 1) the model complexity that maximizes marginal likelihood; 2) the minimal number of model components used to explain structure in data. We used the package pophelper v. 2.3 in R to plot our results (Francis, 2017). Additionally, we estimated the pairwise fixation distance (F_{st}) as a measure of genetic distance among populations, using the package hierfstat v. 0.04 in R (Goudet, 2005). For this, we separated *H.* sp. (L66/333) into four populations based on the different patterns of genetic composition resultant from the Structure analysis.

Tests for introgression

We used ABBA-BABA (Durand *et al.*, 2011) tests for introgression as implemented by comp-D (Mussmann *et al.*, 2019). Given a set of four populations related by the topology (((P1,P2)P3)O), where O is the outgroup, the gene trees for an informative SNP (alleles A and B) concordant with the population history would be (((B,B)A)A) or (((B,B)B)A). There are two possible patterns of discordant gene trees: (((A,B)B)A) or ABBA; and (((B,A)B)A) or BABA. These discordant gene trees may be explained by incomplete lineage sorting, when gene lineages do not coalesce within a population, a phenomenon that is more pronounced among recently diverged species. Under incomplete lineage sorting the frequencies of ABBA and BABA are expected to be the same. Introgression between P3 and P2 would result in a higher frequency of ABBA and introgression between P3 and P1 would result in a higher frequency of BABA. These relative frequencies are calculated by a *D* statistic.

We ran ABBA-BABA tests to check for signs of past introgression in populations of *H*. sp. (L174), *H*. sp. (L66/333), and *H. zebra*. We used the matrix of unlinked SNPs from dataset 4 (Table 6) that was originally generated including populations of *Hypancistrus* from the other Amazon Basin drainages, and *H. contradens* as the outgroup (dataset 7 of Chapter 2). We divided *H*. sp. (L66/333) into four populations according to sampling area and Structure results. We chose a single sample of *H. contradens* as the outgroup for all tests. For P1, P2, and P3, we selected three samples for each population keep our sampling sizes consistent across populations, prioritizing samples with the highest number of reads and choosing different sampling localities when those were available. In the comp-D input for each run we define which samples belong to which population (P1, P2, P3, or O) and the program automatically performs the tests for all possible combinations of samples. We included heterozygote information in calculations.

Although ABBA-BABA tests are performed at the individual level, we report the summary z-scores and p-values combined for populations provided for each run by comp-D (Mussmann *et al.*, 2019).

BPP analysis

In the previous chapter, our Structure results (Chapter 2, Fig. 10) detected a signature of admixed origin of *H*. sp. (L66/333), with ancestral populations correspondent to *H*. sp. (L174) and the lineages from the Guiana Shield drainages. To further explore this hypothesis of hybridization between *H*. sp. (L174) and the lineages from the Guiana Shield drainages to form *H*. sp. (L66/333) we used the multispecies-coalescent-with-introgression model (Flouri *et al.*, 2019) implemented by BPP v. 4.1.4 (Flouri *et al.*, 2018). We estimated parameters of speciation and introgression coalescent times (τ), population size ($\theta = 4N\mu$, where N is the effective population size and μ is the mutation rate), and introgression probability (φ). We chose *H*. sp. (L316) from the Jari River to represent the lineages from the Guiana Shield drainages due to its close proximity to the Xingu River. We selected samples from localities Xin5, Xin7 and Xin16 belonging to *H*. sp. (L66/333) (Fig. 12). We avoided sampling localities on the extremities of the distribution of *H*. sp. (L66/333) to guarantee there is no current overlap in distribution with other species. We built a dataset including three samples from each species (*H*. sp. (L174), *H*. sp. (L66/333),and *H*. sp. (L316)) to run ipyrad, which resulted in 16,465 loci, but we ran BPP with the first 10,000 loci due to computational time constraints (dataset 4, Table 6). We used an inverse-gamma prior for τ ($\alpha = 3$, $\beta = 0.002$), and θ ($\alpha = 3$ and $\beta = 0.02$), and a beta prior for φ ($\alpha = 1$, $\beta = 1$). We used a burn-in of 16,000 iterations, and took 5×10^5 samples, sampling every 2 iterations. We ran two different models: in model A ancestral lineages of *H*. sp.

(L174) and *H*. sp. (L316) first diverge, and then hybridize, forming *H*. sp. (L66/333) (Fig. 13A); in model B there is no initial divergence within ancestral lineages (Fig. 13B). We ran each BPP model twice to check for consistency in parameter estimation.

RESULTS

Sequence processing

After sequencing, each of our samples averaged 3,055,775 unfiltered reads. We obtained in average 15,738 loci per sample after within-sample assembly and an average of 10,710 loci per sample after among-sample assembly and filtering, for a total of 13,230 loci and 759,380 bp for the concatenated dataset 1, allowing up to 20% missing data and individual loci varying between 56 bp and 61 bp in length (Table 6). When selecting one SNP per locus, we obtained 10,984 putatively unlinked SNPs, with 17.7% missing data (dataset 2). Our other sampling schemes resulted additional datasets varying from 10,000 loci (14.4% missing) in dataset 5 to 11,670 loci (10.9% missing) in dataset 4.

Phylogenetic and network analyses

For both our ML (Fig. 14) and SVDquartets (Fig. 15) trees we recovered high support for the monophyly of *H. zebra* and *H*. sp. (L174), but *H*. sp. (L66/333) was found to be paraphyletic. However, this latter conclusion holds only if the assignment of samples from collection site 2 is accurate. This sampling site is the the uppermost locality for *H*. sp. (L66/333), but may also represent the lowermost occurrence of *H*. sp. (L174) and these specimens exhibit an intermediate phenotype. Given this, it is possible all three taxa are indeed monophyletic, if identity of those samples of *H*. sp. (L66/333) are reassigned to *H*. sp. (L174). As expected for within-species

phylogenies, support for most branches is low. The ML analysis placed *H. zebra* as sister to the clade formed by the other two species, a pattern consistent with phylogenetic findings reported in Chapter 2. As noted above, among *H*. sp. (L174) and *H*. sp. (L66/333), five samples from sampling site 2 (Xin2; Fig. 12) are recovered as being most closely related to *H*. sp. (L174), despite being identified phenotypically as *H*. sp. (L66/333). Within the otherwise monophyletic *H*. sp. (L66/333) the tree is hierarchically structured from upstream to downstream, with high support for the clade containing all samples from Xin13 and below, and also support for the monophyly of populations from Xin17 and below. The SVDquartets analysis places *H. zebra* close to *H*. sp. (L174), with both species nested within *H*. sp. (L66/333) (Fig. 15). Samples from locality Xin2 are again closely related to *H*. sp. (L174). Although weakly supported, the phylogenetic structuring of *H*. sp. (L66/333) along the Xingu River is also recovered in the SVDquartets tree.

The phylogenetic network shows a clear split between the three *Hypancistrus* species from the Xingu (Fig. 16). *Hypancistrus zebra* is separated by the longest branch, supporting previous findings that it is the most genetically distinct species of the Xingu. Samples collected from locality Xin2 are placed in an intermediate position between *H*. sp. (L174) and *H*. sp. (L66/333), which unlike the phylogenetic analysis strongly suggests this is a hybrid population and not an issue of species identification as was a possible interpretation of the phylogenetic results. The remaining samples of *H*. sp. (L66/333) are, in agreement with the phylogenetic trees, genetically structured along the Xingu River, with the lowermost populations (Xin17-22) the most distantly related.

Principal components analysis

The first and second axes of the PCA respectively explain 18.6% and 11.2% of the variance in the data (Fig. 17). The first axis separates *H*. sp. (L174) and *H. zebra* on one extreme and *H*. sp. (L66/333) on the other, with Xin2 samples showing intermediate scores. *Hypancistrus zebra* is differentiated from the other species on PC2, which also shows a gradient of differentiation among *H*. sp. (L66/333).

Population structure

The k selection in fastStructure returned a k = 4 as optimal for both the model complexity that maximizes marginal likelihood (marginal likelihood = -0.25) criterion and the minimal number of model components used to explain structure in data criterion (Fig. 18). Adding more clusters did not change the composition of ancestry of individuals. *Hypancistrus zebra* and *H*. sp. (L174) correspond to two distinct and well delimited clusters. Populations of *H*. sp. (L66/333) are assigned to two clusters on a gradient from upstream to downstream. Individuals sampled in location Xin2, above the waterfalls, are clearly recovered as admixed between *H*. sp. (L174) and *H*. sp. (L66/333) lineages. Downstream of the waterfalls, samples from locations Xin3 to Xin12 are primarily assigned to a single cluster, while from location Xin13 to Xin22 there is increasing probability of assignment to the lower Xingu cluster.

Values for pairwise F_{st} are presented in Table 7. These results suggest the pairs of populations that are the least isolated genetically are Xin3-12 and Xin13-16, and Xin13-16 to Xin17-22. *Hypancistrus* sp. (L174) is closer to Xin2 than to any other populations, but Xin2 is more closely related to the population clusters of *H.* sp. (L66/333) than to *H.* sp. (L174). The largest observed differentiation is between *H. zebra* and *H.* sp. (L174).

Table 7. Pairwise F_{st} between populations of *Hypancistrus* from the Xingu River. Populations of *H.* sp. (L66/333) are divided according to sampling localities and Structure results (Xin2, Xin3-12, Xin3-16, Xin17-22).

	L174	Xin2	Xin3-12	Xin13-16	Xin17-22
Xin2	0.339				
Xin3-12	0.582	0.243			
Xin13-16	0.593	0.293	0.103		
Xin17-22	0.571	0.349	0.243	0.163	
H. zebra	0.839	0.742	0.758	0.746	0.689

Tests for introgression

Results of the ABBA-BABA tests for introgression are presented in Table 8. Tests 1 to 4 assume the populations of *H.* sp. (L66/333) Xin3-12 and Xin17-22 form a monophyletic group. Tests 1 to 3 support introgression between the Xin17-22 population and lineages from the Guiana Shield drainages (*H.* sp. (L136), *H.* sp. (L316), and *H.* sp. (L500)). The result of test 4 indicates introgression between Xin3-12 and *H.* sp. (L174). We also considered the alternative hypothesis of Xin17-22 being more closely related to the Guiana Shield drainages than to any populations from the Xingu River (tests 5-10), and our results clearly support instead introgression between Xin3-12 and Xin17-22, and between *H.* sp. (L174) and Xin17-22. If we assume populations Xin3-12 and Xin2 are monophyletic (tests 11 and 12), there is support for admixture between Xin2 and *H.* sp. (L174) and between Xin2 and *H. zebra*. In the case of Xin2 being more closely related to *H.* sp. (L174), test 13 suggests introgression between Xin2 and Xin3-12. Finally, when we assign *H.* sp. (Jacareacanga) and *H. zebra* as P1 and P2 (tests 14 to 17), there is support for admixture of *H. zebra* with *H.* sp. (L174), Xin2, and Xin2-12, but not with Xin17-22.

Table 8. ABBA-BABA test results, assuming topology ((P1,P2)P3). Significant positive z-scores indicate introgression between P1 and P3, significant negative z-scores indicate introgression between P2 and P3. *Hypancistrus contradens* was used as the outgroup for all tests. Populations of *Hypancistrus* sp. (L66/333) are divided according to sampling localities and Structure results (Xin2, Xin3-12, Xin17-22).

Test	P1	P2	P3	z-score / p-value	Conclusion
1	Xin17-22	Xin3-12	L136	6.4 / 1.6×10^{-10}	Xin17-22 ⇔ L136
2	Xin17-22	Xin3-12	L316	3.7 / 2.3×10^{-4}	Xin17-22 ⇔ L316
3	Xin17-22	Xin3-12	L500	5.9 / 4.1×10^{-9}	Xin17-22 ⇔ L500
4	Xin17-22	Xin3-12	L174	-5.9 / 3.2×10^{-9}	Xin3-12 ⇔ L174
5	Xin17-22	L136	Xin3-12	6.3 / 4.0×10^{-10}	Xin17-22 ⇔ Xin3-12
6	Xin17-22	L316	Xin3-12	5.7 / 1.0×10^{-8}	Xin17-22 ⇔ Xin3-12
7	Xin17-22	L500	Xin3-12	7.0 / 2.7×10^{-12}	Xin17-22 ⇔ Xin3-12
8	Xin17-22	L136	L174	5.1 / 4.2×10^{-7}	Xin17-22 ⇔ L174
9	Xin17-22	L316	L174	4.1 / 4.2×10^{-5}	Xin17-22 ⇔ L174
10	Xin17-22	L500	L174	4.9 / 1.2×10^{-6}	Xin17-22 ⇔ L174
11	Xin3-12	Xin2	L174	-7.4 / 1.6×10^{-13}	Xin2 ⇔ L174
12	Xin3-12	Xin2	*H. zebra*	-5.8 / 8.6×10^{-9}	Xin2 ⇔ *H. zebra*
13	Xin2	L174	Xin3-12	5.8 / 6.4×10^{-9}	Xin2 ⇔ Xin3-12
14	Jacareacanga	*H. zebra*	L174	-14.8 / $< 10^{-16}$	*H. zebra* ⇔ L174
15	Jacareacanga	*H. zebra*	Xin2	-7.6 / 2.0×10^{-14}	*H. zebra* ⇔ Xin2
16	Jacareacanga	*H. zebra*	Xin3-12	-6.6 / 3.0×10^{-11}	*H. zebra* ⇔ Xin3-12
17	Jacareacanga	*H. zebra*	Xin17-22	-0.8 / 0.45	none

BPP analysis

Results of the BPP analyzes are reported in Table 9. The repeated runs for each model provided consistent results, suggesting appropriate convergence of paramaters in our runs. Model A (Fig. 13A) estimates a probability of introgression of *H.* sp. (L316) to *H*. sp. (L66/333) of 13% (95% confidence interval CI 8.4 – 15%). However, the estimate for the coalescent time of hybridization (τ_h) was 0 for this model. Model B (Fig. 13B) gives narrower confidence intervals, as expected due to the smaller number of parameters to be estimated (9 in model B versus 13 in model A). The probability of introgression of *H.* sp. (L316) to *H*. sp. (L66/333) is 36% in this model (95% CI 34 – 39%). Population size estimates indicate *H.* sp. (L316) has the smallest population size in both models. Model A indicates *H*. sp. (L66/333) has a larger population than *H*. sp. (L174), but model B supports *H*. sp. (L174) having the larger population.

Table 9. Parameter estimates of BPP for models of hybridization of *Hypancistrus* sp. (L66/333) (refer to Fig. 13). Values correspond to posterior means and 95% HPD confidence intervals. φ is the probability of introgression ,τ is the coalescent time of divergence or hybridization; θ is the population size (4Nμ).

Parameter	Model A	Model B
φ	0.13332 (0.08451, 0.14900)	0.36464 (0.34283, 0.38699)
τ_r	0.00071 (0.00024, 0.00482)	0.00018 (0.00017, 0.00020)
τ_s	0.00018 (0.00009, 0.00025)	= τ_h
τ_t	0.00008 (0.00005, 0.00021)	= τ_h
τ_h	0.00000 (0.00000, 0.00000)	0.00007 (0.00007, 0.00008)
θ_{L316}	0.00011 (0.00007, 0.00017)	0.00005 (0.00005, 0.00006)
θ_{L174}	0.00015 (0.00008, 0.00036)	0.00369 (0.00077, 0.00877)
$\theta_{L66/333}$	0.00115 (0.00019, 0.00296)	0.00023 (0.00020, 0.00025)
θ_r	0.01052 (0.00013, 0.01195)	0.01182 (0.01141, 0.01223)
θ_s	0.03638 (0.01443, 0.06093)	0.00200 (0.00122, 0.00296)
θ_t	0.00077 (0.00032, 0.00428)	0.00011 (0.00010, 0.00012)
θ_{hs}	0.06028 (0.01242, 0.12293)	NA
θ_{ht}	0.00008 (0.00003, 0.00040)	NA

DISCUSSION

Data quality

Our study is the first assessment of the evolutionary history of *Hypancistrus* in the Xingu River. Our ddRAD sequencing and data processing resulted in datasets that includes 10,000 to 13,230 genomic loci, a number that remained robust to variation in sequencing processing parameters, as showed in Chapter 2. We were able to describe major patterns of genetic structuring along the Xingu River, but our interpretations are limited our low sampling of *H. zebra* and *H*. sp. (L174). We did not find evidence for genetic structuring within these species, which may reflect the real pattern in these populations, or may be an effect of sampling. *Hypancistrus zebra* is a threatened species that has suffered intense pressure from harvesting for the aquarium trade, where individual fishes can fetch hundreds of dollars on the market, since its description in 1991 (Isbrücker & Nijssen, 1991; Evers *et al.*, 2019). The low abundance and threatened conservation status of this species hinders access to samples from the wild

(Gonçalves, 2011). *Hypancistrus* sp. (L174), is also difficult to sample due to its habitat preference for great depths in rapids (Leandro Sousa, personal observation). The second factor adversely affecting our ability to interpret the evolutionary history of these Xingu taxa is the complex demographic history in the genus detailed in Chapter 2, which may confound our interpretations. There is a multitude of processes that can generate the patterns observed in our data, and these have to be taken in consideration when interpreting our results. A combination of methods must be used in order differentiate past from ongoing events (Falush *et al.*, 2016), and we are taking the first steps towards unraveling the origins of *Hypancistrus* diversity.

Patterns of genetic structure

We were able to detect four major lineages within our data, corresponding to *H. zebra*, *H.* sp. (L174), and two lineages in *H.* sp. (L66/333). Among the three species, *H. zebra* is the most genetically distinct, as evidenced by the long branches in the ML tree and network (Fig. 14 & 16), PCA plot (Fig. 17), and F_{st} statistics (Table 7). This is consistent with previously reported phylogenetic results of the genus in which *H. zebra* was recovered as being more closely related to species of the Tapajos River than the Xingu River. Although there is evidence for the proximity of *H.* sp. (L174) to *H.* sp. (L66/333), its monophyly in both phylogenetic trees (Fig. 14 & 15) and long branches in the network (Fig. 16) support its identification as a distinct species. Both *H. zebra* and *H.* sp. (L174), occur in the Volta Grande, but occupy different habitats. *Hypancistrus zebra* occurs on large granitic rocks, up to 8 m deep (Gonçalves, 2011) whereas *H*ypancistrus sp. (L174) occurs at greater depths (14 to 40 m) on ironstone-pebble conglomerate rocks (Leandro Sousa, personal observation). While *H. zebra* has a contrasting pattern of well-defined black and white stripes, *H.* sp. (L174) has irregular brown spots on a lighter, tan

background. There is a short stretch of the Xingu River where all three species overlap in distribution, immediately upstream of the Itapaiúna, Itamaracá, and Ananinduba waterfalls, although we were unable to sample any *H. zebra* from this area.

Our comprehensive sampling of *H*. sp. (L66/333) covers the entire distribution of this species, from the terminus of the Volta Grande extending all the way down the Xingu Ria until it reaches the Amazon River. This species presents a pattern of dark worm-like lines on a light background that is highly polymorphic among individuals. Aquarium hobbyists recognize two distinct phenotypes co-distributed across the Xingu Ria based on the pattern of lines on the fins and the tone of the light lines (L66 and L333), but we did not find evidence of population genetic structuring corresponding with this differentiation, consistent with the findings of Cardoso *et al.* (2016). Instead, we detected a gradient of genetic structure from upstream to downstream, supported by the network, PCA, and Structure analyses (Fig. 16-18). This pattern can also be identified in the phylogenetic trees, with a hierarchical topology recovered in the ML tree and clustering of samples from the same area in the SVDquartets tree (Fig. 14 & 15). The results from Chapter 2 indicate samples from localities Xin17-22 are closely related to the *Hypancistrus* lineages from the Guiana Shield tributaries of the Amazon River. The Xingu Ria runs over a sedimentary basin with patches of granitic, sandstone, and conglomerate rocks, where *H*. sp. (L66/333) is found occupying the entire depth range (0 to 40 m) . For this reason we believe *H*. sp. (L66/333) to be a more generalist species than *H. zebra* and *H*. sp. (L174). Although, to the best of our knowledge no *Hypancistrus* has been found in the main course of the Amazon River, this plasticity in habitat use might have favored sporadic migrations of lineages to and from the Guiana Shield drainages through the Amazon. Additionally, changes in rainfall in the Andes and sea levels throughout the Quaternary may have altered the Amazon River's chemical

composition and level, creating conditions that might allow for migration between these tributaries (Höppner *et al.*, 2018).

One of the findings that was most supported throughout these analyses is the hybrid genetic composition of the samples collected from site Xin2 (Fig. 14-18). This population is located immediately above the Ananinduba waterfall. Less than 2 km downstream is locality Xin3, and approximately 8 km upstream is the source from which most of our *H*. sp. (L174) samples were collected (Fig. 12C). The waterfalls appear to constitute a semi-permeable barrier to gene flow, as no mixture appears above these, the lowest of the rapids of the Volta Grande. *Hypancistrus* sp. (L174) does not occur downstream of the waterfalls, but Xin2 samples present genetic signatures of both *H*. sp. (L174) and *H*. sp. (L66/333), implying that the later was at some point able to move upstream of the waterfalls. As we discussed previously, *H*. sp. (L174) appears to be more of a habitat specialist than *H*. sp. (L66/333), so the different geologic composition of the two areas could be the critical barrier preventing the downstream migration of *H*. sp. (L174) (Sawakuchi *et al.*, 2015; Fitzgerald *et al.*, 2018). Evaluating the effect of such waterfalls on the genetic structuring of *H*. sp. (L174) was more limited since we only had two samples from the locality above the Jericoá waterfall (ID 5924 and 5925, Fig. 12B). Although we did not specifically test these different populations, our phylogenetic trees and network did not suggest these two samples were genetically distinguishable from their conspecifics (Fig. 14 & 15).

Evolutionary processes

We found important differences between the results from the ML and SVDquartets phylogenetic trees, specifically, the position of *H*. sp. (L174) relative to *H*. sp. (L66/333) and *H.*

zebra (Fig. 14 & 15). The ML method places *H. zebra* in a separate clade from *H*. sp. (L174), however, SVDquartets places *H*. sp. (L174) close to *H. zebra* both in Chapter 2 and 3. The four population tests indicate *H*. sp. (L174) is the sister clade to all Amazon Basin *Hypancistrus* lineages (Chapter 2, Table 5), but more evidence is warranted to confirm this hypothesis. We additionally suggest *H*. sp. (L174) and *H. zebra* experienced past introgression events, which would explain the inconsistencies in our analyses. This hypothesis is supported by the highly significant ABBA-BABA test (test 14, Table 8).

The phylogenetic analysis of *Hypancistrus* including other Amazon Basin drainages in Chapter 2 shows a complex history with signs of admixture among many species. The Xingu was the Amazon tributary with the most intriguing phylogenetic composition, including three or four different lineages of *Hypancistrus*, depending on the analysis. When we subsampled *H*. sp. (L66/333) to match the sampling size of the populations of *H. zebra* and *H*. sp. (L174), Structure *H*. sp. (L66/333) appeared as a hybrid between *H*. sp. (L174) and the lineages from the Guiana Shield drainages (Chapter 2, Fig. 10). Structure is sensitive to sampling bias, and doesn't make a distinction between patterns of recent gene flow or past introgression followed by drift (Falush *et al.*, 2016). For that reason, we investigated the possibility of a hybrid origin of *H*. sp. (L66/333) with BPP. For model A (Fig. 13), the estimate of the parameter of coalescent time of hybridization ($\tau_h = 0$) and the low probability of introgression from *H*. sp. (L316) into *H*. sp. (L66/333) (mean $\varphi = 13\%$) failed to support the hybrid origin hypothesis. However, model B provided moderate support for introgression of Guiana Shield drainages (mean $\varphi = 36\%$). Considering that we had purposefully sampled only Xin2-12, excluding samples nearest the Amazon that would be presumed to be the most likely to be introgressed, this model supports our hypothesis. To determine the best model, further testing would be necessary, adding additional

populations of the Guiana Shield drainages, and running a test to compare models, again using BPP (Rannala & Yang, 2017; Flouri *et al.*, 2019).

Species boundaries

Our findings suggest multiple events of gene flow among *Hypancistrus* lineages in the Xingu, making the species boundaries hard to define in this system despite the apparent support for monophyly when only results of phylogenetic analyses are considered. The traditional species concept of a group interbreeding individuals that is reproductively isolated does not seem to fit our findings and perhaps we should consider a combination of conditions to delimit species of *Hypancistrus*. Such conditions might include aspects of color pattern, geographic and ecological isolation, as well as patterns of genetic isolation and exchange.

CONCLUSION

In this study we investigate, for the first time, the evolution of *Hypancistrus* in the Xingu River using a genome-wide dataset. Application of a large, genomic-scale dataset to this complex system has revealed a history of past and present introgression events. Our findings support the presence of four lineages of *Hypancistrus* in the Xingu, namely *H. zebra*, and *H*. sp. (L174) in the Volta Grande, and *H*. sp. (L66/333) from upper and middle Xingu Ria, and *H*. sp. (L66/333) from the lower Xingu Ria. Additionally, two hybrid zones were detected. First, Xin2, the locality immediately above the Itapaiúna, Itamaracá, and Ananinduba waterfalls, has an admixed genetic signal from *H*. sp. (L174) and *H*. sp. (L66/333). Second, the lowermost portion of the distribution of *H*. sp. (L66/333) is genetically structured on a gradient, with increasing genetic signature from the lineages from the Guiana Shield drainages towards the mouth of the

Xingu River. While we were able to detect signals of past introgression between these lineages, the direction, timing, and intensity of these events requires further, more intensive sampling of taxa outside the Xingu River. This study system serves as an excellent model to enlighten many poorly understood aspects of tropical freshwater diversification, making important advancements towards our understanding of evolution in tropical freshwater systems.

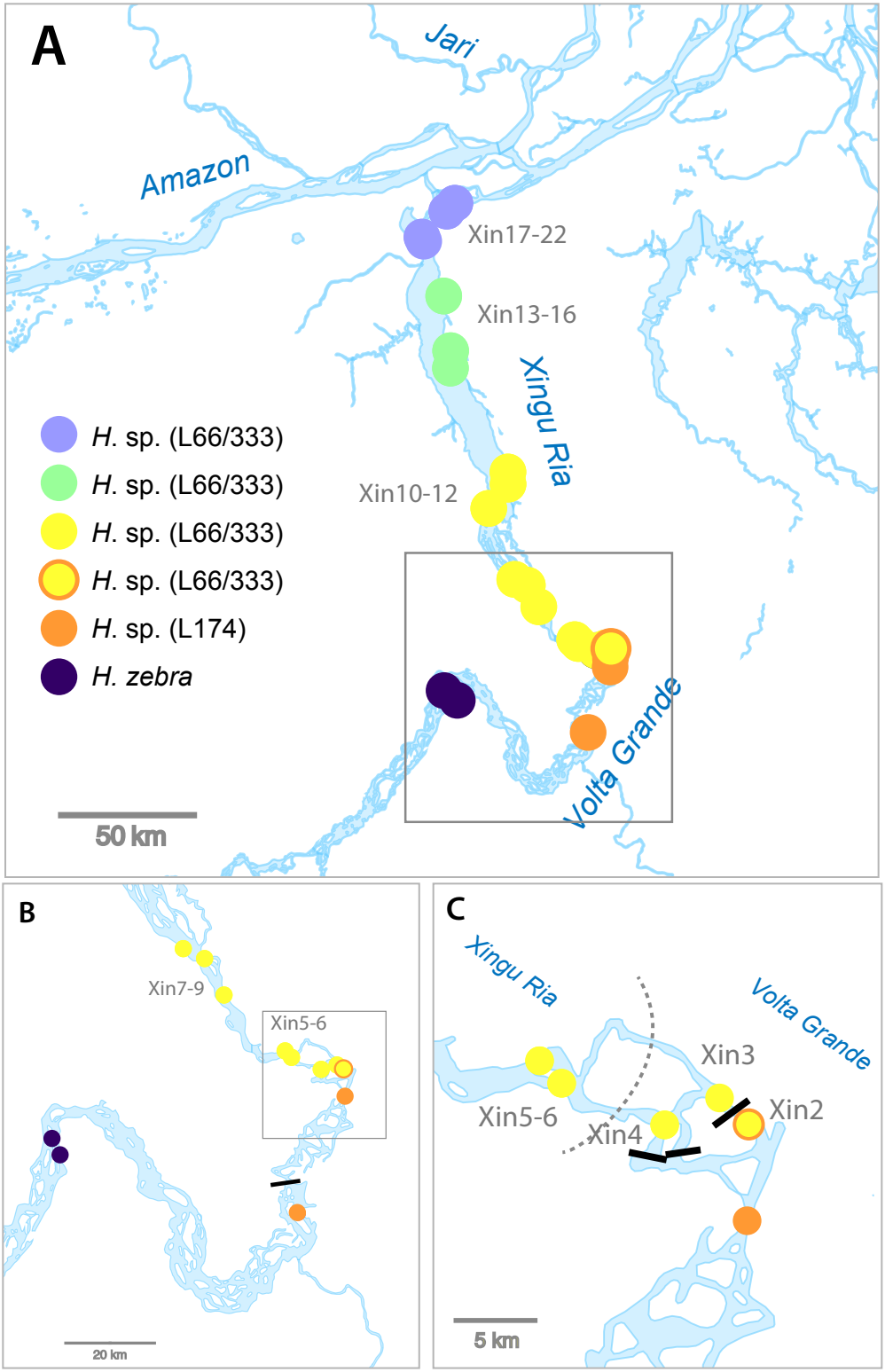


Figure 12. Distribution of *Hypancistrus* from the Xingu sampled for this study. In A and B the highlighted areas correspond respectively to B and C. Bar in B represents Jericoá waterfall. Bars in C represent, from left to right, Itapaiuna, Itamaracá, and Ananinduba waterfalls. Dashed line in C indicates the division between the Xingu Ria and Volta Grande. Collection site 2 (Xin2) is indicated by the yellow circle outlined in orange.

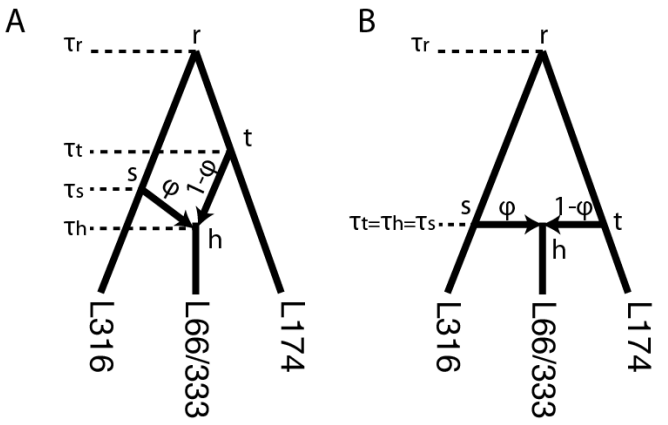


Figure 13. Models of hybrid origin of *Hypancistrus* sp. (L66/333) tested with BPP. A) Parental lineages first diverge, and then hybridize. B) Ancestral populations come into contact and hybridize, without prior divergence. r, s, and t represent the ancestral populations; h is the hybridization node; τ is the coalescent time of divergence or hybridization; φ is the probability of introgression. Modified from Flouri *et al.* (2019).

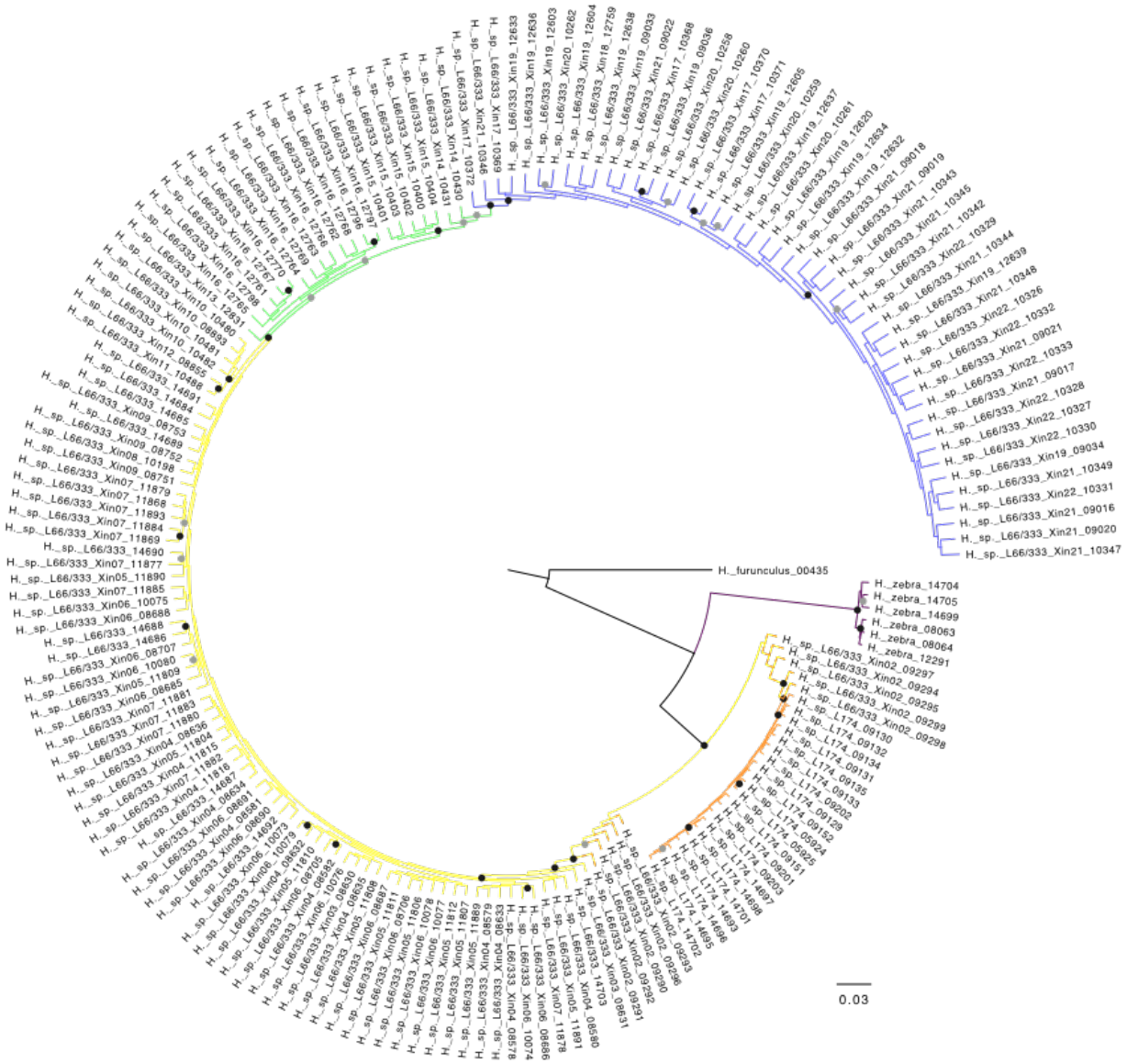


Figure 14. IQ-TREE phylogeny of *Hypancistrus* from the Xingu River based on maximum likelihood of 13230 concatenated loci including 181 individuals and 4 species (dataset 1, see Table 6). Circles indicate Ultrafast bootstrap support (black 95-100; grey 90-94).

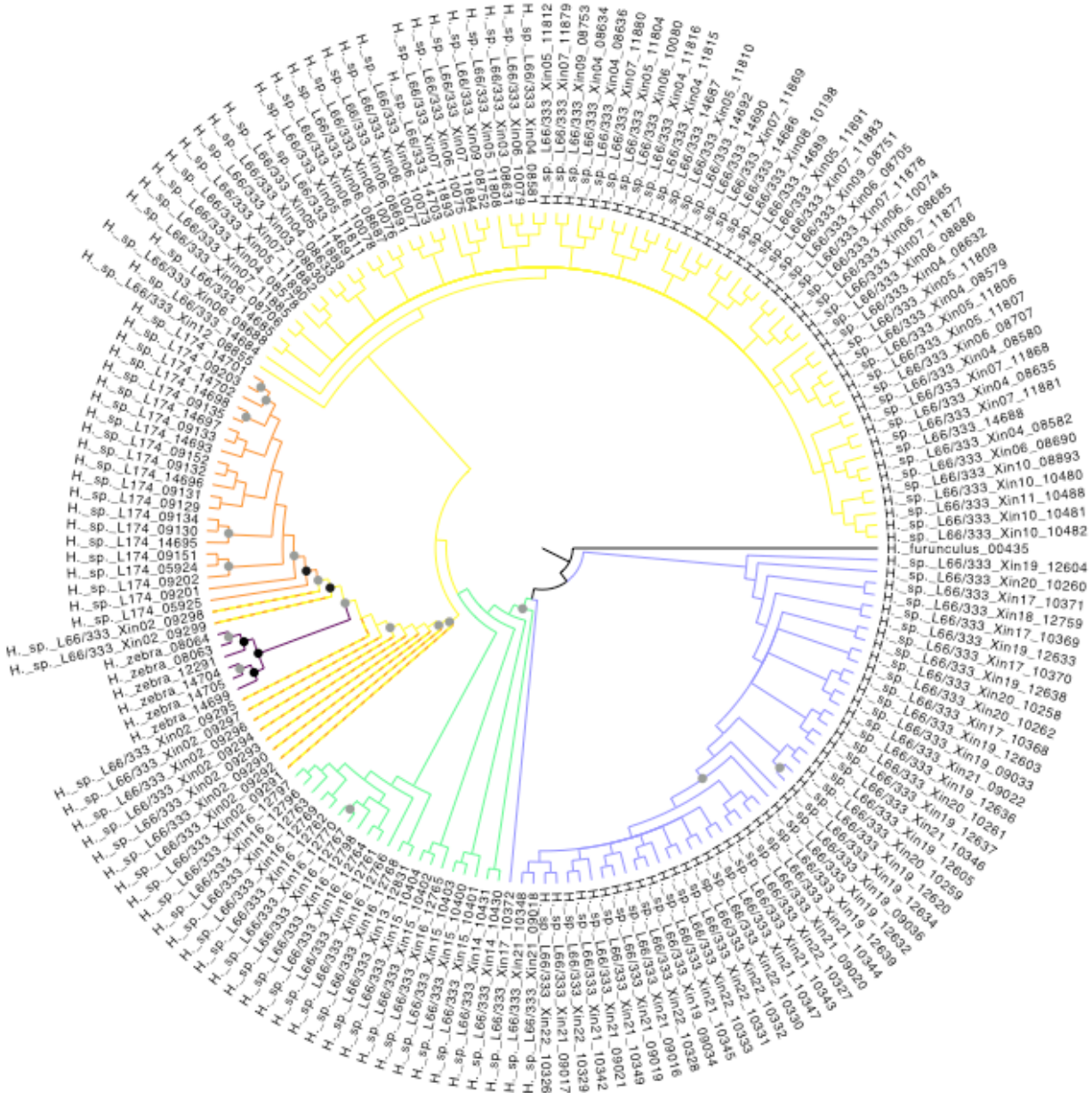


Figure 15. Phylogeny of *Hypancistrus* from the Xingu River based on coalescent method of SVDquartets analysis of 10984 unlinked SNPs including 181 individuals and four species (dataset 2, see Table 6). Circles indicate bootstrap support (black 85-100; grey 50-84).

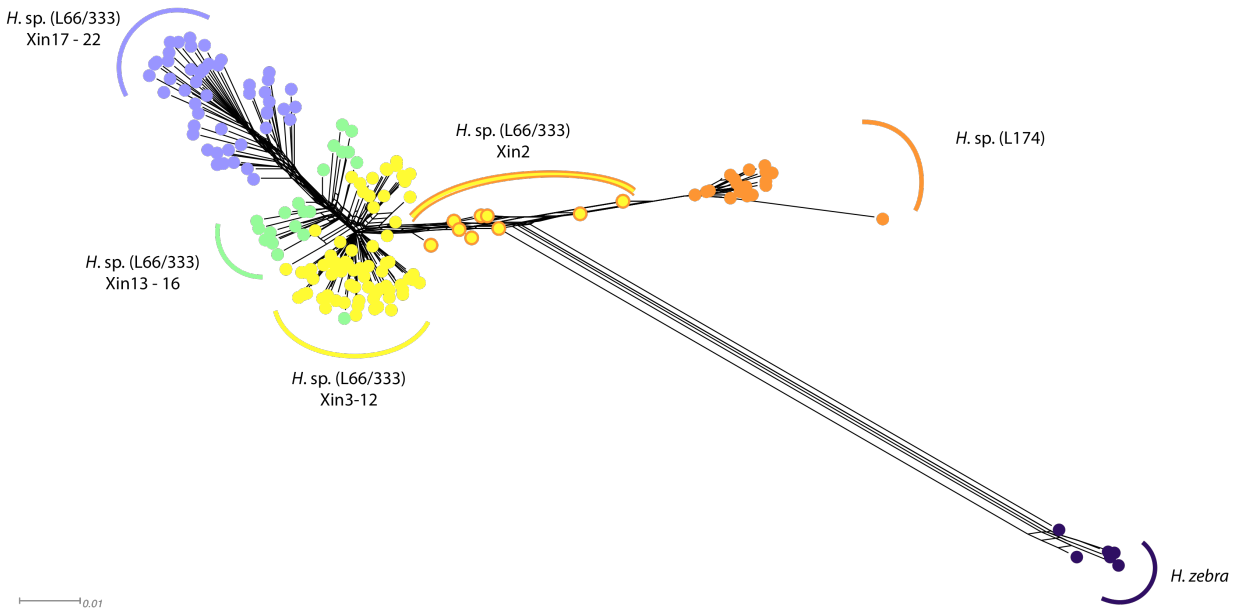


Figure 16. Network analysis of *Hypancistrus* from the Xingu River based on SplitsTree of 11013 unlinked SNPs including 180 individuals and three species.

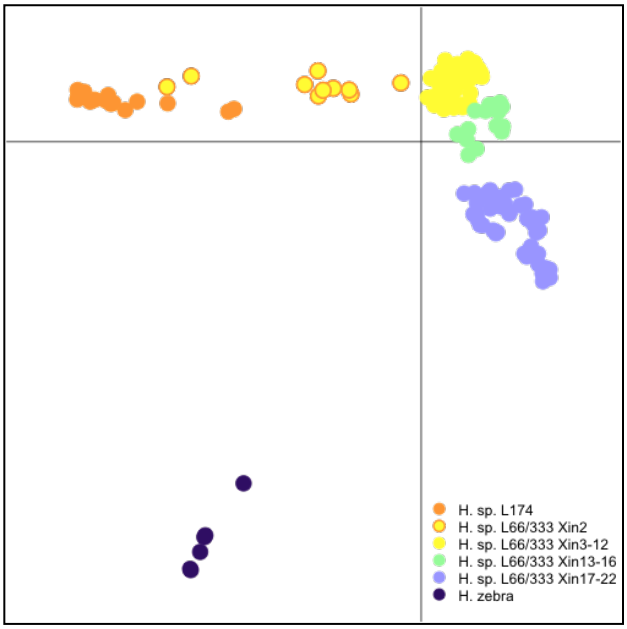


Figure 17. Principal components analysis of *Hypancistrus* from the Xingu River based on 11013 unlinked SNP loci including 180 individuals and three species.

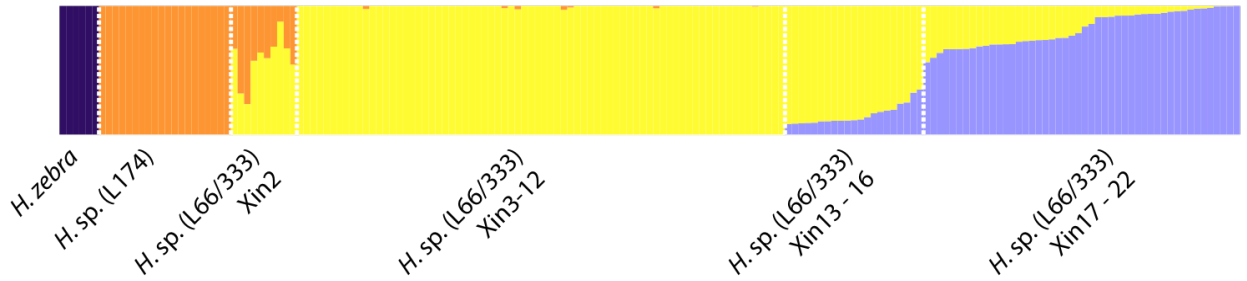


Figure 18. Population assignment plots for best k = 4 for FastStructure analyses of *Hypancistrus* from the Xingu River based on 11013 unlinked SNP loci including 180 individuals and three species.

BIBLIOGRAPHY

Albert, J.S. & Carvalho, T.P. (2011) Neogene assembly of modern faunas. *Historical biogeography of Neotropical freshwater fishes* (ed. by J.S. Albert and R.E. Reis), pp. 119-136. University of California Press, Berkeley.

Albert, J.S. & Reis, R.E. (2011a) *Historical biogeography of Neotropical freshwater fishes*. University of California Press, Berkeley.

Albert, J.S. & Reis, R.E. (2011b) Introduction to Neotropical freshwaters. *Historical biogeography of Neotropical freshwater fishes* (ed. by J.S. Albert and R.E. Reis), pp. 3-19. University of California Press, Berkeley.

Albert, J.S., Val, P. & Hoorn, C. (2018) The changing course of the Amazon River in the Neogene: center stage for Neotropical diversification. *Neotropical Ichthyology*, **16**

Aljanabi, S.M. & Martinez, I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692-4693.

Alter, S.E., Munshi-South, J. & Stiassny, M.L.J. (2017) Genomewide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River. *Molecular Ecology*, **26**, 1401-1419.

Anderson, E.P., Jenkins, C.N., Heilpern, S., Maldonado-Ocampo, J.A., Carvajal-Vallejos, F.M., Encalada, A.C., Rivadeneira, J.F., Hidalgo, M., Cañas, C.M., Ortega, H., Salcedo, N., Maldonado, M. & Tedesco, P.A. (2018) Fragmentation of Andes-to-Amazon connectivity by hydropower dams. *Science Advances*, **4**, eaao1642.

Andrade, M.C., Winemiller, K.O., Barbosa, P.S., Fortunati, A., Chelazzi, D., Cincinelli, A. & Giarrizzo, T. (2019) First account of plastic pollution impacting freshwater fishes in the

Amazon: Ingestion of plastic debris by piranhas and other serrasalmids with diverse feeding habits. *Environmental Pollution*, **244**, 766-773.

Arantes, C.C., Winemiller, K.O., Petrere, M., Castello, L., Hess, L.L. & Freitas, C.E.C. (2018) Relationships between forest cover and fish diversity in the Amazon River floodplain. *Journal of Applied Ecology*, **55**, 386-395.

Ardura, A., Gomes, V., Linde, A.R., Moreira, J.C., Horreo, J.L. & Garcia-Vazquez, E. (2013) The Meeting of Waters, a possible shelter of evolutionary significant units for Amazonian fish. *Conservation Genetics*, **14**, 1185-1192.

Armbruster, J.W. (2002) *Hypancistrus inspector*: a new species of suckermouth armored catfish (Loricariidae: Ancistrinae). *Copeia*, **2002**, 86-92.

Armbruster, J.W., Lujan, N.K. & Taphorn, D. (2007) Four new *Hypancistrus* (Siluriformes: Loricariidae) from Amazonas, Venezuela. *Copeia*, **2007**, 62-79.

Balian, E.V., Segers, H., Leveque, C. & Martens, K. (2008) The freshwater animal diversity assessment: an overview of the results. *Hydrobiologia*, **595**, 627-637.

Barber, I. (2010) From 'trash fish'to supermodel: the rise and rise of the three-spined stickleback in evolution and ecology. *Biologist*, **57**, 15-21.

Barletta, M., Jaureguizar, A.J., Baigun, C., Fontoura, N.F., Agostinho, A.A., Almeida-Val, V.M.F., Val, A.L., Torres, R.A., Jimenes-Segura, L.F., Giarrizzo, T., Fabré, N.N., Batista, V.S., Lasso, C., Taphorn, D.C., Costa, M.F., Chaves, P.T., Vieira, J.P. & Corrêa, M.F.M. (2010) Fish and aquatic habitat conservation in South America: a continental overview with emphasis on neotropical systems. *Journal of Fish Biology*, **76**, 2118-2176.

Beaumont, M.A., Zhang, W. & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025-2035.

Beheregaray, L.B., Cooke, G.M., Chao, N.L. & Landguth, E.L. (2015) Ecological speciation in the tropics: insights from comparative genetic studies in Amazonia. *Frontiers in Genetics*, **5**

Benone, N.L., Ligeiro, R., Juen, L. & Montag, L.F.A. (2018) Role of environmental and spatial processes structuring fish assemblages in streams of the eastern Amazon. *Marine and Freshwater Research*, **69**, 243-252.

Bernardi, G. (2013) Speciation in fishes. *Molecular Ecology*, **22**, 5487-5502.

Besnier, F. & Glover, K.A. (2013) ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers. *PLOS ONE*, **8**, e70651.

Bishop, P. (1995) Drainage rearrangement by river capture, beheading and diversion. *Progress in Physical Geography*, **19**, 449-473.

Blob, R.W., Bridges, W.C., Ptacek, M.B., Maie, T., Cediel, R.A., Bertolas, M.M., Julius, M.L. & Schoenfuss, H.L. (2008) Morphological selection in an extreme flow environment: body shape and waterfall-climbing success in the Hawaiian stream fish *Sicyopterus stimpsoni*. *Integrative and Comparative Biology*, **48**, 734-749.

Bloom, D.D. & Lovejoy, N.R. (2011) The biogeography of marine incursions in South America. *Historical biogeography of Neotropical freshwater fishes* (ed. by J.S. Albert and R.E. Reis), pp. 137-144. University of California Press, Berkeley.

Bloom, D.D. & Lovejoy, N.R. (2017) On the origins of marine-derived freshwater fishes in South America. *Journal of Biogeography*, **44**, 1927-1938.

Braga-Silva, A. & Galetti, P.M. (2016) Evidence of isolation by time in freshwater migratory fish *Prochilodus costatus* (Characiformes, Prochilodontidae). *Hydrobiologia*, **765**, 159-167.

Britton, J.R. & Orsi, M.L. (2012) Non-native fish in aquaculture and sport fishing in Brazil: economic benefits versus risks to fish diversity in the upper River Paraná Basin. *Reviews in Fish Biology and Fisheries*, **22**, 555-565.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A. & RoyChoudhury, A. (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular biology and Evolution*, **29**, 1917-1932.

Burridge, C.P., Craw, D. & Waters, J.M. (2006) River capture, range expansion, and cladogenesis: the genetic signature of freshwater vicariance. *Evolution*, **60**, 1038-1049.

Butlin, R., Debelle, A., Kerth, C., Snook, R.R., Beukeboom, L.W., Cajas, R.F.C., Diao, W., Maan, M.E., Paolucci, S., Weissing, F.J., van de Zande, L., Hoikkala, A., Geuverink, E., Jennings, J., Kankare, M., Knott, K.E., Tyukmaeva, V.I., Zoumadakis, C., Ritchie, M.G., Barker, D., Immonen, E., Kirkpatrick, M., Noor, M., Garcia, C.M., Schmitt, T., Schilthuizen, M. & Network, M.C.S. (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27-39.

Byers, K.J.R.P., Xu, S. & Schlüter, P.M. (2017) Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology*, **26**, 277-290.

Camargo, M., Giarrizzo, T. & Isaac, V. (2004) Review of the geographic distribution of fish fauna of the Xingu river basin, Brazil. *Ecotropica*, **10**, 123-147.

Campbell, K.E., Frailey, C.D. & Romero-Pittman, L. (2006) The Pan-Amazonian Ucayali Peneplain, late Neogene sedimentation in Amazonia, and the birth of the modern Amazon River system. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **239**, 166-219.

Cardoso, A.L., Carvalho, H.L., Benathar, T.C., Serrao, S.M., Nagamachi, C.Y., Pieczarka, J.C., Sousa, L.M., Ready, J.S. & Noronha, R.C. (2016) Integrated cytogenetic and mitochondrial DNA analyses Indicate that two different phenotypes of *Hypancistrus* (L066 and L333) belong to the same species. *Zebrafish*, **13**, 209-16.

Cardoso, Y.P. & Montoya-Burgos, J.I. (2009) Unexpected diversity in the catfish *Pseudancistrus brevispinis* reveals dispersal routes in a Neotropical center of endemism: the Guyanas Region. *Molecular Ecology*, **18**, 947-964.

Carvalho Júnior, J.R., Carvalho, N.A.S.S., Nunes, J.L.G., Camões, A., Bezerra, M.F.C., Santana, A.R. & Nakayama, L. (2009) Sobre a pesca de peixes ornamentais por comunidades do rio Xingu, Pará-Brasil: Relato de Caso. *Boletim do Instituto de Pesca*, **35**, 521-530.

Castello, L., McGrath, D.G., Hess, L.L., Coe, M.T., Lefebvre, P.A., Petry, P., Macedo, M.N., Renó, V.F. & Arantes, C.C. (2013) The vulnerability of Amazon freshwater ecosystems. *Conservation Letters*, **6**, 217-229.

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

Chifman, J. & Kubatko, L. (2015) Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology*, **374**, 35-47.

Cooke, G.M., Chao, N.L. & Beheregaray, L.B. (2012) Marine incursions, cryptic species and ecological diversification in Amazonia: the biogeographic history of the croaker genus *Plagioscion* (Sciaenidae). *Journal of Biogeography*, **39**, 724-738.

Cooke, G.M., Landguth, E.L. & Beheregaray, L.B. (2014) Riverscape genetics identifies replicated ecological divergence across an Amazonian ecotone. *Evolution*, **68**, 1947-1960.

Crotti, M., Barratt, C.D., Loader, S.P., Gower, D.J. & Streicher, J.W. (2019) Causes and analytical impacts of missing data in RADseq phylogenetics: insights from an African frog (*Afrixalus*). *Zoologica Scripta*, **48**, 157-167.

da Silva, M., Ribeiro, E.D., Matoso, D.A., Sousa, L.M., Hrbek, T., Py-Daniel, L.R. & Feldberg, E. (2014) Chromosomal polymorphism in two species of *Hypancistrus* (Siluriformes: Loricariidae): an integrative approach for understanding their biodiversity. *Genetica*, **142**, 127-39.

Dagosta, F.C.P. & Pinna, M.D. (2019) The fishes of the Amazon: distribution and biogeographical patterns, with a comprehensive list of species. *Bulletin of the American Museum of Natural History*, **2019**, 1-163, 163.

de Araújo, J.G., dos Santos, M.A.S., Rebello, F.K. & Isaac, V.J. (2017) Cadeia comercial de peixes ornamentais do Rio Xingu, Pará, Brasil. *Boletim do Instituto de Pesca*, **43**, 297-307.

de Sousa, A.L.P., Maciel, L.A.M., Reginaldo, L. & Rodrigues, R. (2018) Estudo da comercialização de peixes ornamentais da família Loricariidae (Siluriformes) em Santarém/PA.

Dray, S. & Dufour, A.-B. (2007) The ade4 package: implementing the duality diagram for ecologists. *2007*, **22**, 1-20.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, **7**, 214.

Durand, E.Y., Patterson, N., Reich, D. & Slatkin, M. (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239-2252.

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359-361.

Eaton, D. & Overcast, I. (2019) *ipyrad: Interactive assembly and analysis of RAD-seq data sets. Available at* Available at: https://github.com/dereneaton/ipyrad (accessed October 18 2019).

Escobar, M.D.L., Andrade-López, J., Farias, I.P. & Hrbek, T. (2015) Delimiting evolutionarily significant units of the fish, *Piaractus brachypomus* (Characiformes: Serrasalmidae), from the Orinoco and Amazon River basins with insight on routes of historical connectivity. *Journal of Heredity*, **106**, 428-438.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Evers, H.-G., Pinnegar, J.K. & Taylor, M.I. (2019) Where are they all from? – sources and sustainability in the ornamental freshwater fish trade. *Journal of Fish Biology*, **94**, 909-916.

Falush, D., van Dorp, L. & Lawson, D.J. (2016) A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. *bioRxiv*, 066431.

Farias, I.P., Torrico, J.P., Garcia-Davila, C., Santos, M.D.F., Hrbek, T. & Renno, J.F. (2010) Are rapids a barrier for floodplain fishes of the Amazon basin? A demographic study of the keystone floodplain species *Colossoma macropomum* (Teleostei: Characiformes). *Molecular Phylogenetics and Evolution*, **56**, 1129-1135.

Fausch, K.D., Torgersen, C.E., Baxter, C.V. & Li, H.W. (2002) Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience*, **52**, 483-498.

Fearnside, P.M. (2009) Hydroelectric dams in the Brazilian Amazon as sources of 'greenhouse' gases. *Environmental Conservation*, **22**, 7-19.

Fitzgerald, D.B., Sabaj Perez, M.H., Sousa, L.M., Gonçalves, A.P., Rapp Py-Daniel, L., Lujan, N.K., Zuanon, J., Winemiller, K.O. & Lundberg, J.G. (2018) Diversity and community structure of rapids-dwelling fishes of the Xingu River: Implications for conservation amid large-scale hydroelectric development. *Biological Conservation*, **222**, 104-112.

Flouri, T., Jiao, X., Rannala, B. & Yang, Z. (2018) Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution*, **35**, 2585-2593.

Flouri, T., Jiao, X., Rannala, B. & Yang, Z. (2019) A Bayesian implementation of the multispecies coalescent model with introgression for comparative genomic analysis. *bioRxiv*, 766741.

Francis, R.M. (2017) pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, **17**, 27-32.

Frederico, R.G., Olden, J.D. & Zuanon, J. (2016) Climate change sensitivity of threatened, and largely unprotected, Amazonian fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **26**, 91-102.

Freitas, F.L.M., Sparovek, G., Berndes, G., Persson, U.M., Englund, O., Barretto, A. & Mörtberg, U. (2018) Potential increase of legal deforestation in Brazilian Amazon after Forest Act revision. *Nature Sustainability*, **1**, 665-670.

Fry, J.D. (2009) Laboratory experiments on speciation. *Experimental evolution: concepts, methods, and applications of selection experiments* (ed. by T. Garland and M.R. Rose), pp. 631-656. University of California Press, Berkeley.

Gibson, G. (2005) The synthesis and evolution of a supermodel. *Science*, **307**, 1890-1891.

Glenn, T.C., Bayona-Vásquez, N.J., Kieran, T.J., Pierson, T.W., Hoffberg, S.L., Scott, P.A., Bentley, K.E., Finger, J.W., Watson, P.R., Louha, S., Troendle, N., Díaz-Jaimes, P., Mauricio, R. & Faircloth, B.C. (2017) Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1USD per Sample (3RAD). *bioRxiv*, 205799.

Gonçalves, A.P. (2011) *Ecologia e etnoecologia de Hypancistrus zebra (Siluriformes: Loricariidae) no rio Xingu, Amazônia brasileira*. Universidade Federal do Pará, Belém-PA.

Goudet, J. (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184-186.

Gruber, B., Georges, A., Berry, O. & Georges, A. (2019) *dartR: importing and analysing SNP and Silicodart data generated by genome-wide restriction fragment analysis. Available at https://rdrr.io/cran/dartR/*.

Grummer, J.A., Beheregaray, L.B., Bernatchez, L., Hand, B.K., Luikart, G., Narum, S.R. & Taylor, E.B. (2019) Aquatic landscape genomics and environmental effects on genetic variation. *Trends in Ecology & Evolution*,

Henderson, P.A. & Crampton, W.G.R. (2009) A comparison of fish diversity and abundance between nutrient-rich and nutrient-poor lakes in the Upper Amazon. *Journal of Tropical Ecology*, **13**, 175-198.

Herrera, S. & Shank, T.M. (2015) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. *bioRxiv*,

Hey, J., Chung, Y., Sethuraman, A., Lachance, J., Tishkoff, S., Sousa, V.C. & Wang, Y. (2018) Phylogeny estimation by integration over isolation with migration models. *Molecular Biology and Evolution*, **35**, 2805-2818.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**, 518-522.

Hoorn, C. & Wesselingh, F.P. (2010) Introduction: Amazonia, landscape and species evolution. *Amazonia – landscape and species evolution: a look into the past* (ed. by C. Hoorn and F.P. Wesselingh), pp. 1-6. Wiley-Blackwell, Chichester, UK ; Hoboken, NJ.

Hoorn, C., Bogotá-A, G.R., Romero-Baez, M., Lammertsma, E.I., Flantua, S.G.A., Dantas, E.L., Dino, R., do Carmo, D.A. & Chemale Jr, F. (2017) The Amazon at sea: onset and stages of the Amazon River from a marine record, with special reference to Neogene plant turnover in the drainage basin. *Global and Planetary Change*, ***in press***

Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartin, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri,

F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Sarkinen, T. & Antonelli, A. (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927-931.

Höppner, N., Lucassen, F., Chiessi, C.M., Sawakuchi, A.O. & Kasemann, S.A. (2018) Holocene provenance shift of suspended particulate matter in the Amazon River basin. *Quaternary Science Reviews*, **190**, 66-80.

Hrbek, T., Meliciano, N.V., Zuanon, J. & Farias, I.P. (2018) Remarkable geographic structuring of rheophilic fishes of the lower Araguaia River. *Frontiers in genetics*, **9**, 295-295.

Huang, H. & Knowles, L.L. (2016) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, **65**, 357-365.

Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, **23**, 254.

Irion, G. & Kalliola, R. (2010) Long−term landscape development processes in Amazonia. *Amazonia – landscape and species evolution: a look into the past* (ed. by C. Hoorn and F.P. Wesselingh), pp. 185-197. Wiley-Blackwell, Chichester, UK ; Hoboken, NJ.

Isbrücker, I. & Nijssen, H. (1991) *Hypancistrus zebra*, a new genus and species of uniquely pigmented ancistrine loricariid fish from the Rio Xingu, Brazil (Pisces: Siluriformes: Loricariidae). *Ichthyological Explorations of Freshwaters*, **1**, 345-350.

Jackson, N.D., Morales, A.E., Carstens, B.C. & O'Meara, B.C. (2017) PHRAPL: Phylogeographic Inference Using Approximate Likelihoods. *Systematic Biology*, **66**, 1045-1053.

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

Jombart, T. & Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070-3071.

Kano, Y., Nishida, S. & Nakajima, J. (2012) Waterfalls drive parallel evolution in a freshwater goby. *Ecology and Evolution*, **2**, 1805-1817.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179-1191.

Labonne, J. & Hendry, A.P. (2010) Natural and sexual selection giveth and taketh away reproductive barriers: models of population divergence in guppies. *American Naturalist*, **176**, 26-39.

Laraque, A., Lopez, J.L., Yepez, S. & Georgescu, P. (2019) Water and sediment budget of casiquiare channel linking Orinoco and Amazon catchments, Venezuela. *Water*, **11**, 2068.

Leaché, A.D., Harris, R.B., Rannala, B. & Yang, Z. (2013) The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology*, **63**, 17-30.

Linck, E. & Battey, C.J. (2018) Minor allele frequency thresholds dramatically affect population structure inference with genomic datasets. *bioRxiv*, 188623.

Lovejoy, N.R. & de Araújo, M.L.G. (2000) Molecular systematics, biogeography and population structure of Neotropical freshwater needlefishes of the genus *Potamorrhaphis*. *Molecular Ecology*, **9**, 259-268.

Lovejoy, N.R., Willis, S.C. & Albert, J.S. (2010) Molecular signatures of Neogene biogeographical events in the Amazon fish fauna. *Amazonia – landscape and species evolution: a look into the past* (ed. by C. Hoorn and F. Wesselingh), pp. 405-417. Wiley-Blackwell, Chichester, UK ; Hoboken, NJ.

Lovejoy, T.E. & Nobre, C. (2018) Amazon tipping point. *Science Advances*, **4**, eaat2340.

Lujan, N.K. & Armbruster, J.W. (2011) Two New Genera and Species of Ancistrini (Siluriformes: Loricariidae) from the Western Guiana Shield. *Copeia*, 216-225.

Lujan, N.K. & Conway, K.W. (2015) Life in the fast lane: a review of rheophily in freshwater fishes. *Extremophile Fishes: Ecology, Evolution, and Physiology of Teleosts in Extreme Environments* (ed. by R. Riesch, M. Tobler and M. Plath), pp. 107-136. Springer International Publishing, Cham.

Lujan, N.K., Armbruster, J.W., Lovejoy, N.R. & López-Fernández, H. (2015) Multilocus molecular phylogeny of the suckermouth armored catfishes (Siluriformes: Loricariidae) with a focus on subfamily Hypostominae. *Molecular Phylogenetics and Evolution*, **82, Part A**, 269-288.

Lujan, N.K., Cramer, C.A., Covain, R., Fisch-Muller, S. & López-Fernández, H. (2017) Multilocus molecular phylogeny of the ornamental wood-eating catfishes (Siluriformes, Loricariidae, Panaqolus and Panaque) reveals undescribed diversity and parapatric clades. *Molecular Phylogenetics and Evolution*, **109**, 321-336.

Lundberg, J.G., Marshall, L.G., Guerrero, J., Horton, B., Malabarba, M. & F., W. (1998) The stage for Neotropical fish diversification: a history of tropical South American rivers. *Phylogeny and Classification of Neotropical Fishes* (ed. by L.R. Malabarba, R.E. Reis, R.P. Vari, Z.M. Lucena and C.A.S. Lucena), pp. 13-48. EdiPUCRS, Porto Alegre.

Machado, V.N., Willis, S.C., Teixeira, A.S., Hrbek, T. & Farias, I.P. (2017) Population genetic structure of the Amazonian black flannelmouth characin (Characiformes, Prochilodontidae: *Prochilodus nigricans* Spix & Agassiz, 1829): contemporary and historical gene flow of a migratory and abundant fishery species. *Environmental Biology of Fishes*, **100**, 1-16.

Maisey, J.G. (2000) Continental break up and the distribution of fishes of Western Gondwana during the Early Cretaceous. *Cretaceous Research*, **21**, 281-314.

Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189-197.

Markert, J.A., Schelly, R.C. & Stiassny, M.L.J. (2010) Genetic isolation and morphological divergence mediated by high-energy rapids in two cichlid genera from the lower Congo rapids. *BMC Evolutionary Biology*, **10**

Marques, D.A., Meier, J.I. & Seehausen, O. (2019) A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, **34**, 531-544.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)* (ed by, pp. 1-8.

Moreau, M. & Coomes, O.T. (2007) Aquarium fish exploitation in western Amazonia: conservation issues in Peru. *Environmental Conservation*, **34**, 12-22.

Musilova, Z., Kalous, L., Petrtyl, M. & Chaloupkova, P. (2013) Cichlid fishes in the Angolan headwaters region: molecular evidence of the ichthyofaunal contact between the Cuanza and Okavango-Zambezi systems. *PLOS ONE*, **8**, e65047.

Mussmann, S.M., Douglas, M.R., Bangs, M.R. & Douglas, M.E. (2019) Comp-D: a program for comprehensive computation of D-statistics and population summaries of reticulated evolution. *Conservation Genetics Resources*,

Nadilo, B. (2012) World's third largest hydropower plant. *Gradevinar*, **64**, 243-252.

Narum, S.R., Powell, M.S., Evenson, R. & Sharp, B. (2006) Microsatellites reveal population substructure of Klickitat River native steelhead and genetic divergence from an introduced stock. *North American Journal of Fisheries Management*, **26**, 147-155.

Netto-Ferreira, A.L. & Moreira, C.R. (2018) New species of *Creagrutus* (Ostariophysi; Characiformes; Characidae) from the Rio Xingu drainage, Brazil. *Zootaxa*, **4375**, 250-256.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**, 268-274.

O'Leary, S.J., Puritz, J.B., Willis, S.C., Hollenbeck, C.M. & Portnoy, D.S. (2018) These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*, **27**, 3193-3206.

Oberdorff, T., Dias, M.S., Jézéquel, C., Albert, J.S., Arantes, C.C., Bigorne, R., Carvajal-Valleros, F.M., De Wever, A., Frederico, R.G., Hidalgo, M., Hugueny, B., Leprieur, F., Maldonado, M., Maldonado-Ocampo, J., Martens, K., Ortega, H., Sarmiento, J., Tedesco, P.A., Torrente-Vilara, G., Winemiller, K.O. & Zuanon, J. (2019) Unexpected fish diversity gradients in the Amazon basin. *Science Advances*, **5**, eaav8681.

Ochoa, L.E., Pereira, L.H.G., Costa-Silva, G.J., Roxo, F.F., Batista, J.S., Formiga, K., Foresti, F. & Oliveira, C. (2015) Genetic structure and historical diversification of catfish

Brachyplatystoma platynemum (Siluriformes: Pimelodidae) in the Amazon basin with implications for its conservation. *Ecology and Evolution*, **5**, 2005-2020.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C. & Samadi, S. (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450-459.

Paz-Vinas, I., Loot, G., Stevens, V.M. & Blanchet, S. (2015) Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology*, **24**, 4586-4604.

Pelicice, F.M., Pompeu, P.S. & Agostinho, A.A. (2015) Large reservoirs as ecological barriers to downstream movements of Neotropical migratory fish. *Fish & Fisheries*, **16**, 697-715.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLOS ONE*, **7**, e37135.

Pickrell, J.K. & Pritchard, J.K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, **8**, e1002967.

Pires, T.H.S., Borghezan, E.A., Cunha, S.L.R., Leitão, R.P., Pinto, K.S. & Zuanon, J. (2019) Sensory drive in colourful waters: morphological variation suggests combined natural and sexual selection in an Amazonian fish. *Biological Journal of the Linnean Society*, **127**, 351-360.

Pringle, C.M., Freeman, M.C. & Freeman, B.J. (2000) Regional effects of hydrologic alterations on riverine macrobiota in the New World: tropical-temperate comparisons: The massive scope of large dams and other hydrologic modifications in the temperate New World has resulted in distinct regional trends of biotic impoverishment. While neotropical rivers have fewer dams and limited data upon which to make regional generalizations, they are

ecologically vulnerable to increasing hydropower development and biotic patterns are emerging. *BioScience*, **50**, 807-823.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Puckett, E.E., Park, J., Combs, M., Blum, M.J., Bryant, J.E., Caccone, A., Costa, F., Deinum, E.E., Esther, A., Himsworth, C.G., Keightley, P.D., Ko, A., Lundkvist, Å., McElhinney, L.M., Morand, S., Robins, J., Russell, J., Strand, T.M., Suarez, O., Yon, L. & Munshi-South, J. (2016) Global population divergence and admixture of the brown rat (*Rattus norvegicus*). *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20161762.

R Development Core Team (2019) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rahel, F.J. (2007) Biogeographic barriers, connectivity and homogenization of freshwater faunas: it's a small world after all. *Freshwater Biology*, **52**, 696-710.

Raj, A., Stephens, M. & Pritchard, J.K. (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, **197**, 573-589.

Rannala, B. & Yang, Z. (2017) Efficient Bayesian species tree inference under the multispecies coalescent. *Systematic Biology*, **66**, 823-842.

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

Ribolli, J., Hoeinghaus, D.J., Johnson, J.A., Zaniboni-Filho, E., de Freitas, P.D. & Galetti, P.M. (2017) Isolation-by-time population structure in potamodromous Dourado *Salminus brasiliensis* in southern Brazil. *Conservation Genetics*, **18**, 67-76.

Roxo, F.F., Ochoa, L.E., Sabaj, M.H., Lujan, N.K., Covain, R., Silva, G.S.C., Melo, B.F., Albert, J.S., Chang, J., Foresti, F., Alfaro, M.E. & Oliveira, C. (2019) Phylogenomic reappraisal of the Neotropical catfish family Loricariidae (Teleostei: Siluriformes) using ultraconserved elements. *Molecular Phylogenetics and Evolution*, **135**, 148-165.

Rubin, B.E.R., Ree, R.H. & Moreau, C.S. (2012) Inferring Phylogenies from RAD Sequence Data. *PLOS ONE*, **7**, e33394.

Sabaj-Pérez, M. (2015) Where the Xingu bends and will soon break. *American Scientist*, **103**, 395-403.

Salzburger, W. & Meyer, A. (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften*, **91**, 277-290.

Sawakuchi, A., Hartmann, G., Sawakuchi, H., Pupim, F., Bertassoli, D., Parra, M., Antinao, J., Sousa, L., Sabaj Pérez, M. & Oliveira, P. (2015) The Volta Grande do Xingu: reconstruction of past environments and forecasting of future scenarios of a unique Amazonian fluvial landscape. *Scientific Drilling*, **20**, 21.

Schneider, C.H., Gross, M.C., Terencio, M.L. & Porto, J.I.R. (2012) Cryptic diversity in the mtDNA of the ornamental fish *Carnegiella strigata*. *Journal of Fish Biology*, **81**, 1210-1224.

Schwarzer, J., Misof, B. & Schliewen, U.K. (2012) Speciation within genomic networks: a case study based on *Steatocranus* cichlids of the lower Congo rapids. *Journal of Evolutionary Biology*, **25**, 138-148.

Schwarzer, J., Misof, B., Ifuta, S.N. & Schliewen, U.K. (2011) Time and origin of cichlid colonization of the Lower Congo rapids. *PLOS ONE*, **6**, e22380.

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

Silva-Oliveira, C., Sabaj, M.H., Ota, R.P. & Py-Daniel, L.H.R. (2019) *Bryconops rheorubrum* (Characiformes: Iguanodectidae), new species from the Rio Xingu Rapids, Brazil. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **166**, 1-21, 21.

Sobel, J.M., Chen, G.F., Watt, L.R. & Schemske, D.W. (2010) The biology of speciation. *Evolution*, **64**, 295-315.

Solís-Lemus, C., Bastide, P. & Ané, C. (2017) PhyloNetworks: a Package for phylogenetic networks. *Molecular Biology and Evolution*, **34**, 3292-3298.

Sousa, L.M., Chaves, M.S., Akama, A., Zuanon, J. & Sabaj, M.H. (2018) *Platydoras birindellii*, new species of striped raphael catfish (Siluriformes: Doradidae) from the Xingu Basin, Brazil. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **166**, 1-13, 13.

Stegmann, L.F., Leitão, R.P., Zuanon, J. & Magnusson, W.E. (2019) Distance to large rivers affects fish diversity patterns in highly dynamic streams of Central Amazonia. *PLOS ONE*, **14**, e0223880.

Stokes, M.F., Goldberg, S.L. & Perron, J.T. (2018) Ongoing river capture in the Amazon. *Geophysical Research Letters*, **45**, 5545-5552.

Storfer, A., Murphy, M., Evans, J., Goldberg, C., Robinson, S., Spear, S., Dezzani, R., Delmelle, E., Vierling, L. & Waits, L. (2006) Putting the 'landscape'in landscape genetics. *Heredity*, **98**, 128-142.

Swofford, D.L. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Tan, M. & Armbruster, J.W. (2016) Two new species of spotted *Hypancistrus* from the Rio Negro drainage (Loricariidae, Hypostominae). *Zookeys*, 123-35.

Terencio, M.L., Schneider, C.H. & Porto, J.I.R. (2012) Molecular signature of the D-loop in the brown pencilfish *Nannostomus eques* (Characiformes, Lebiasinidae) reveals at least two evolutionary units in the Rio Negro basin, Brazil. *Journal of Fish Biology*, **81**, 110-124.

Torrente-Vilara, G., Zuanon, J., Leprieur, F., Oberdorff, T. & Tedesco, P.A. (2011) Effects of natural rapids and waterfalls on fish assemblage structure in the Madeira River (Amazon Basin). *Ecology of Freshwater Fish*, **20**, 588-597.

Vitule, J.R.S., Agostinho, A.A., Azevedo-Santos, V.M., Daga, V.S., Darwall, W.R.T., Fitzgerald, D.B., Frehse, F.A., Hoeinghaus, D.J., Lima-Junior, D.P., Magalhães, A.L.B., Orsi, M.L., Padial, A.A., Pelicice, F.M., Petrere, M., Pompeu, P.S. & Winemiller, K.O. (2017) We need better understanding about functional diversity and vulnerability of tropical freshwater fishes. *Biodiversity and Conservation*, **26**, 757-762.

Vonhof, H.B. & Kaandorp, R.J. (2009) Climate variation in Amazonia during the Neogene and the Quaternary. *Amazonia – landscape and species evolution: a look into the past* (ed. by C. Hoorn and F.P. Wesselingh), pp. 199-210. Wiley-Blackwell, Chichester, UK ; Hoboken, NJ.

Waters, J.M., Craw, D., Youngson, J.H. & Wallis, G.P. (2001) Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. *Evolution*, **55**, 1844-1851.

Willis, S.C., Nunes, M., Montana, C.G., Farias, I.P., Orti, G. & Lovejoy, N.R. (2010) The Casiquiare river acts as a corridor between the Amazonas and Orinoco river basins: biogeographic analysis of the genus Cichla. *Molecular Ecology*, **19**, 1014-1030.

Willis, S.C., Winemiller, K.O., Montaña, C.G., Macrander, J., Reiss, P., Farias, I.P. & Ortí, G. (2015) Population genetics of the speckled peacock bass (*Cichla temensis*), South America's most important inland sport fishery. *Conservation Genetics*, **16**, 1345-1357.

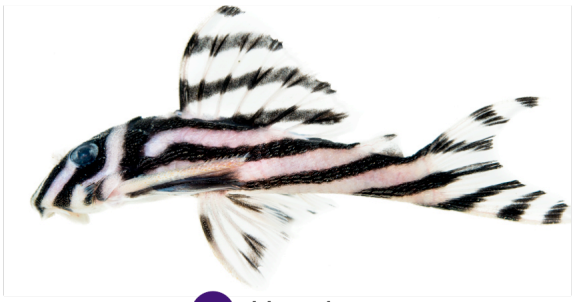
Winemiller, K.O. & Willis, S.C. (2011) The Vaupes Arch and Casiquiare Canal. *Historical biogeography of Neotropical freshwater fishes* (ed. by J.S. Albert and R.E. Reis), pp. 225-242. University of California Press, Berkeley.

Winemiller, K.O., Lopez-Fernandez, H., Taphorn, D.C., Nico, L.G. & Duque, A.B. (2008) Fish assemblages of the Casiquiare River, a corridor and zoogeographical filter for dispersal between the Orinoco and Amazon basins. *Journal of Biogeography*, **35**, 1551-1563.

Zuanon, J.A.S. (1999) *História natural da ictiofauna de corredeiras do rio Xingu, na região de Altamira, Pará*. Thesis, Universidade Estadual de Campinas, Campinas.

APPENDIX

H. sp. (L174)



H. zebra

H. sp. (L66/333) Xin2

H. sp. (L66/333) Xin3

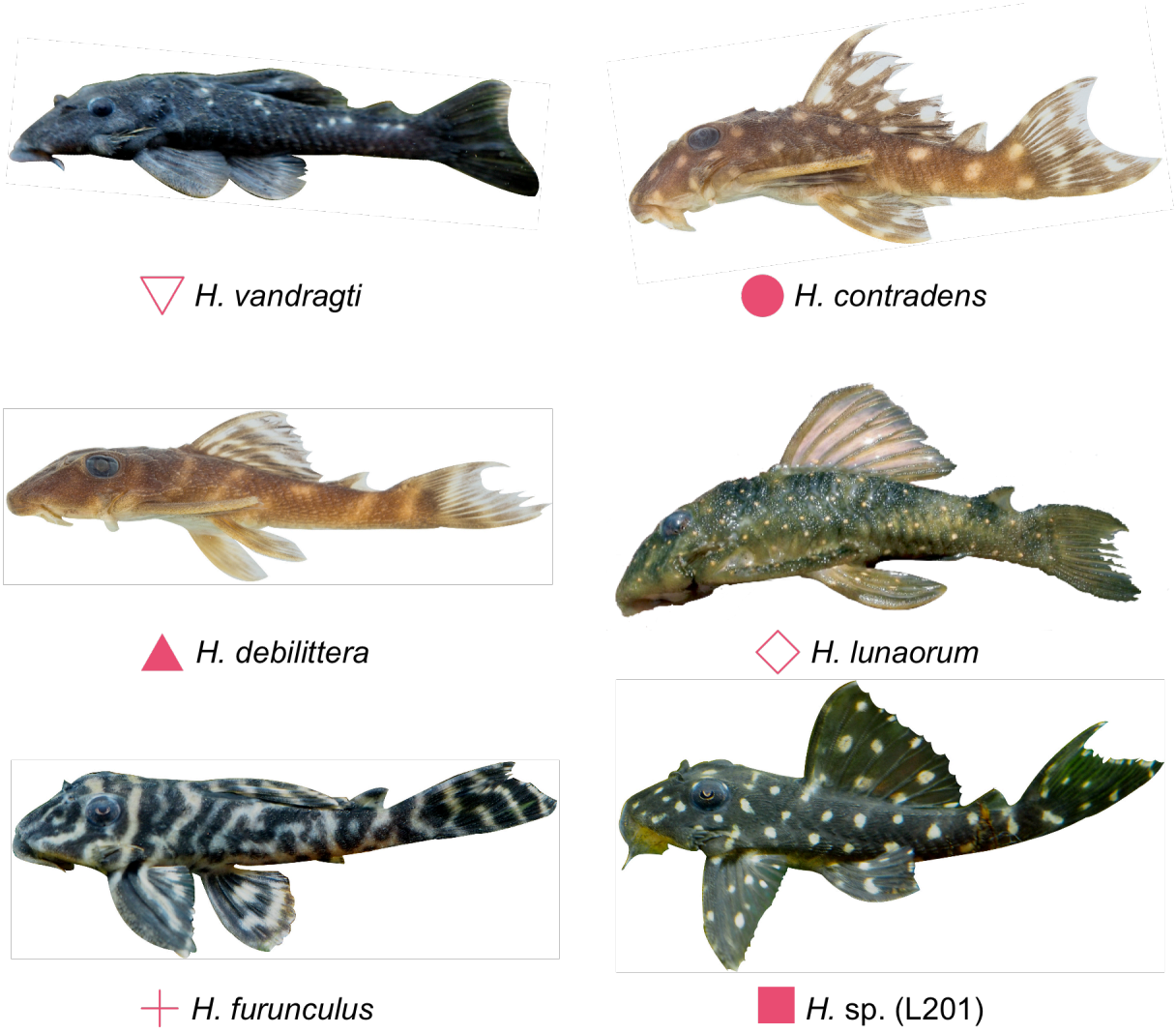
H. sp. (L66/333) Xin6

H. sp. (L66/333) Xin10

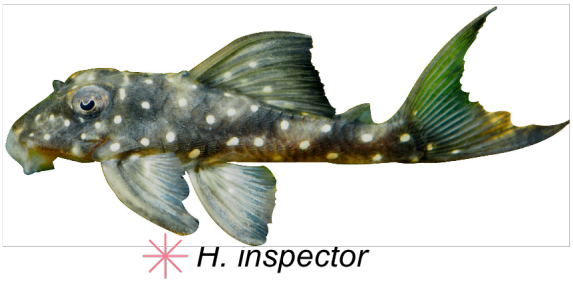
H. sp. (L66/333) Xin16

H. sp. (L66/333) Xin19

Hypancistrus from the Xingu River. *Hypacistrus* sp. (L174) and *H. zebra* were photographed in the field while alive. *Hypancistrus* sp. (L66/333) are photographs of preserved specimens from the Ichthyology Collection of the Academy of Natural Sciences of Drexel University – ANSP.



Hypancistrus from the Orinoco Basin. *Hypacistrus vandragti*, *H. furunculus*, and *H*. sp.(L201) were photographed in hobby aquariums while alive (Photos by Haakon Haagensen). *Hypancistrus debilittera* and *H. contradens* are photographs of preserved specimens from the Ichthyology Collection of the Academy of Natural Sciences of Drexel University – ANSP. *Hypancistrus lunaorum* photograph modified from Ambruster et al. (2007).



H. inspector

H. sp. (L136)

H. sp. (Jacareacanga)

H. sp. (L270)

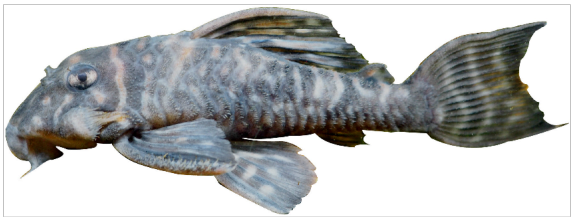
H. sp. (L260)

H. sp. (L316)

H. sp. (L411)

H. sp. (L499)

H. margaritatus



H. sp. (L500)

Hypancistrus from the Amazon Basin. *Hypacistrus* sp. (Jacareacanga) and *H. margaritatus* were photographed in the field while alive (Photos by Mark Sabaj). All others were photographed in hobby aquariums while alive (Photos by Haakon Haagensen).

VITA

MARCELLA GONÇALVES SANTOS

The University of Mississippi, Shoemaker Hall, University, MS 38677

E-mail: marcella_gs@yahoo.com.br

Education

B.S. **2008**

Biology – Universidade de Brasília Brasília, Brazil

Fully licensed teacher **2009**

Biology – Universidade de Brasília Brasília, Brazil

M.Sc. **2011**

Ecology – Universidade de Brasília, Brasília, Brazil

Research Experience: Projects

Research for academic credit **2005-2007**

Universidade de Brasília, Biology Institute, Brasília, Brazil

Principal Investigator: Dr. Sônia Báo

Research topic: Sperm ultrastructure and systematics of snakes

Research for academic credit **2007**

Universidade de Brasília, Center of Primatology, Brasília, Brazil

Principal Investigator: Dr. Maria Clotilde Henriques Tavares

Research topic: Investigation of the perception of vision illusions by humans and monkeys (*Cebus appela*)

Undergraduate research fellow **2008-2009**

Universidade de Brasília, Biology Institute, Brasília, Brazil

Principal Investigator: Dr. Guarino Colli

Research topic: Scale microstructure and phylogeography of the Cerrado gecko *Gymnodactylus amarali*

M.Sc. Student **2009-2011**

Universidade de Brasília, Biology Institute, Brasília, Brazil

Principal Investigator: Dr. Guarino Colli

Research topic: Phylogeography of the Cerrado endemic lizard *Micrablepharus atticolus*

Research Fellow **2011-2012**

Universidade de Brasília, Biology Institute, Brasília, Brazil

Principal Investigator: Dr. Guarino Colli

Research topic: Phylogeography of the Cerrado endemic anole *Norops meridionalis*

Ph.D. Student **2014-2019**

University of Mississippi, Department of Biology, University, MS

Principal Investigator: Dr. Brice Noonan

Research topic: Population genomics of catfishes of the genus *Hypancistrus* in the Xingu River, Brazil

Awards and Fellowships

Undergraduate research scholarship, Universidade de Brasília, Brazil	2008-2009
Masters research scholarship, CAPES, Brazil	2009-2011
Science Without Borders scholarship, CNPq, Brazil	2015-present
Biology Graduate Student Society award, University of Mississippi	2018

Publications

Santos, M.G., Nogueira, C., Giugliano, L.G. & Colli, G.R. 2014. Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae) an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, 41, 1506-1519.

Domingos, F.M.C.B., Bosque, R.J., Cassimiro, J., Colli, G.R., Rodrigues, M.T., **Santos, M.G**. & Beheregaray, L.B. 2014. Out of the deep: Cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution*, 80, 113-124.

Guarnizo, C.E., Werneck, F.P., Giugliano, L.G., **Santos, M.G.**, Fenker, J., Sousa, L., D'Angiolella, A.B., dos Santos, A.R., Strüssmann, C., Rodrigues, M.T., Dorado-Rodrigues, T.F., Gamble, T. & Colli, G.R. 2016. Cryptic lineages and diversification of an endemic anole lizard (Squamata, Dactyloidae) of the Cerrado hotspot. *Molecular Phylogenetics and Evolution*, 94, Part A, 279-289.

Domingos, F.M.C.B., Arantes, I.C., Bosque, R.J., **Santos, M.G.** 2017. Nesting in the lizard *Phyllopezus pollicaris* (Squamata: Phyllodactylidae) and a phylogenetic perspective on communal nesting in the family. *Phyllomedusa***,** 16, 255-267.

Teaching Assistantships

Biology Institute, Universidade de Brasília, Brasília, Brazil

Histology	2005
Evolution	2008
Biology of Reptiles	2008/2009
Conservation Genetics	2010
Statistics Applied to Ecology	2010
General Biology	2010

Department of Biology, University of Mississippi, Oxford, Mississippi

Genetics	2014/2015/2017/2019
Inquiry Into Life Laboratory: Human Biology	2015/2017
Introductory Biological Sciences	2018/2019

Academic Advising

Samantha Irwin, Undergrad	2018