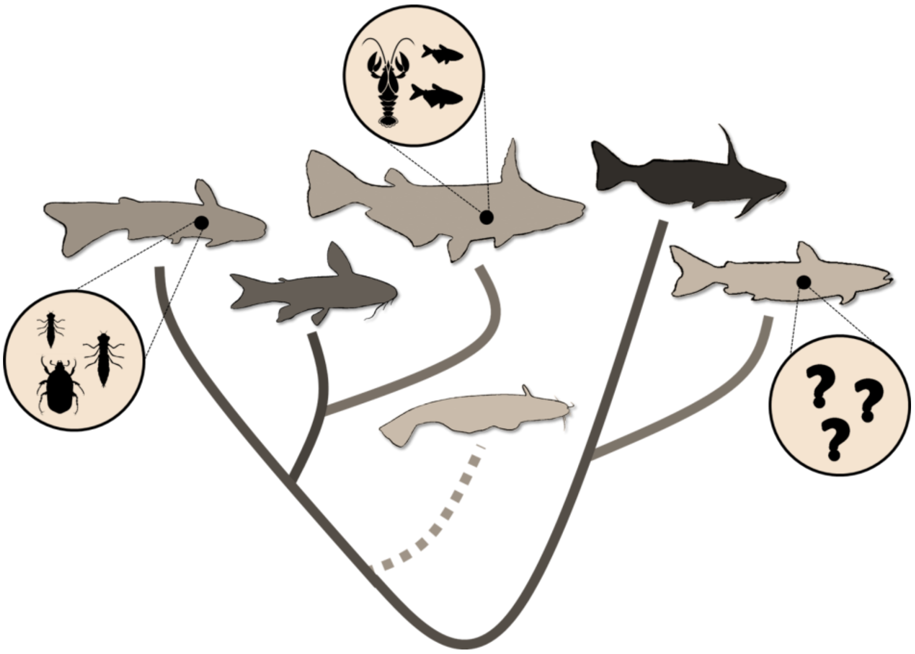
UNIVERSIDADE FEDERAL DO PARÁ
EMBRAPA AMAZÔNIA ORIENTAL
PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA
CURSO DE DOUTORADO EM ECOLOGIA

TIAGO MAGALHÃES DA SILVA FREITAS

LACUNAS DE CONHECIMENTO E EVOLUÇÃO DA DIETA DE PEIXES AUCHENIPTERIDEOS (TELEOSTEI: SILURIFORMES)

BELÉM
2019

TIAGO MAGALHÃES DA SILVA FREITAS



LACUNAS DE CONHECIMENTO E EVOLUÇÃO DA DIETA DE PEIXES AUCHENIPTERIDEOS (TELEOSTEI: SILURIFORMES)

Tese apresentada ao Programa de Pós-graduação em Ecologia, do Convênio Universidade Federal do Pará e EMBRAPA Amazônia Oriental, como requisito parcial para obtenção do título de Doutor em Ecologia.

Área de concentração: Ecologia.

Linha de Pesquisa: Ecologia de Comunidades

Orientador: Dr. Luciano Fogaça de Assis Montag
Co-orientador: Dr. Paulo De Marco Jr.

BELÉM
2019

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)

F862l Freitas, Tiago Magalhães da Silva
Lacunas de conhecimento e evolução da dieta de peixes auchenipterídeos (Teleostei: Siluriformes) / Tiago Magalhães da Silva Freitas. — 2020.
116 f. : il. color.

Orientador(a): Prof. Dr. Luciano Fogaça de Assis Montag Coorientador(a): Prof. Dr. Paulo de Marco Junior
Tese (Doutorado) - Programa de Pós-Graduação em Biodiversidade e Biotecnologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, 2020.

1. Diversidade de peixes. 2. Repositório online. 3. taxonomia. 4. filogenia. 5. reconstrução ancestral. I. Título.

CDD 574.524

TIAGO MAGALHÃES DA SILVA FREITAS

LACUNAS DE CONHECIMENTO E EVOLUÇÃO DA DIETA DE PEIXES AUCHENIPTERIDEOS (TELEOSTEI: SILURIFORMES)

Tese apresentada ao Programa de Pós-graduação em Ecologia, do Convênio Universidade Federal do Pará e EMBRAPA Amazônia Oriental, como requisito parcial para obtenção do título de Doutor em Ecologia.

COMISSÃO JULGADORA

Dr. Luciano Fogaça de Assis Montag (Presidente)
Universidade Federal do Pará (UFPA)

Dr. Leandro Juen
Universidade Federal do Pará (UFPA)

Dr. Jose Alexandre F. Diniz Filho
Universidade Federal de Goiás (UFG)

Dr. Jansen Alfredo Sampaio Zuanon
Instituto Nacional de Pesquisas da Amazônia (INPA)

Dra. Geiziane Tessarolo
Universidade Estadual de Goiás (UEG)

Dr. Fabricio Barreto Teresa
Universidade Estadual de Goiás (UEG)

Dr. Richard James Ladle
Universidade Federal de Alagoas (UFAL)

Dr. Cástor Guisande González
Universidad de Vigo (Espanha)

Aprovada em: 27 de dezembro de 2019.
Local de defesa: Sala SAT4, Instituto de Ciências Biológicas, UFPA.

Dedico esse trabalho a minha esposa *Mel Duarte*, meus pais (*Nádia* e *Francisco*), irmãos (*Bruno* e *Nívia*) e avô *Nadir*. Sem vocês nada disso seria possível!

AGRADECIMENTOS

Agradeço imensamente às seguintes pessoas e instituições que colaboraram no desenvolvimento deste trabalho:

Ao Orientador, **Prof. Dr. Luciano Montag** (Miúdo) por mais essa empreitada vencida. Obrigado pelos incentivos nesses mais de quinze anos de formação pessoal e acadêmica! Minha trajetória se deve muito a você.

Ao **Prof. Dr. Paulo De Marco**, Co-orientador, que me recebeu no TheMetaLand (UFG), me propiciando quatro anos de muita alegria. Levo comigo: "*piano, piano, se va lontano*".

Ao **Prof. Dr. Joaquin Hortal**, supervisor de Doutorado Sanduíche. Foi um enorme prazer concretizar esse trabalho com você! Obrigado por me acolher em Madri (Espanha). Essa oportunidade foi um privilégio. Mudou minha vida!

Aos colaboradores e amigos **Dra. Juliana Stropp**, **Dr. Lucas Jardim**, **Dra. Barbara Calegari**, **Dr. Joaquín Calatayud** e **Dr. Fabricio Villalobos**. Sem vocês, tudo seria muito mais difícil. Obrigado!

À Profa. **Dra. Maria Auxiliadora**, que foi minha orientadora no começo do Doutorado, e que muito contribuiu até o exame de Qualificação.

Aos professores **Dra. Cecília Leal**, **Dr. Leandro Juen** e **Dra. Thaisa Michelan**, pelas valiosas contribuições na Qualificação deste trabalho.

À **CAPES** pela concessão da bolsa de estudo do Doutorado Sanduíche, realizado no *Museo Nacional de Ciencias Naturales* em Madri, Espanha.

Ao curso de **Pós-Graduação em Ecologia** (UFPA/EMBRAPA) pela oportunidade.

Á Universidade Federal do Pará, em especial ao **Campus Universitário do Marajó-Breves**, que me concedeu a licença para cursar o Doutorado.

À minha esposa **Mel Duarte**, que além do amor e companheirismo, formou comigo uma linda família com **Nala** e **Dama** (em memória). Amo muito vocês!

Aos meus pais **Nádia e Francisco**, meus irmãos **Bruno** e **Nívia** e à minha avó **Nadir**, pelo amor, incentivo e apoio incondicional.

À minha segunda família, **Ivaina**, **Vinícius**, **Jedai, Taís, Benjamin** e **Écio**, muito obrigado pelo carinho!

Aos amigos do **TheMetaLand** e **UFG**, em especial: **André**, **Bruna**, **Daisy**, **Danira**, **Letícia**, **Luciano**, **Ludimila**, **Marcelo**, **Nayara**, **Nicholas**, **Paula**, **Raíssa, Rhayane**, **Vini**, **Zander** e **Zé**. O tempo em Goiânia foi ainda melhor com vocês.

Aos amigos do **LABECO** e **UFPA**, em especial: **Bruno Prudente**, **Erival**, **Fernando**, **Gilberto**, **Giovanni**, **Lidia**, **Naraiana**, **Thiely** e **Tiago Begot**.

Por fim, a **todos** aqueles que contribuíram, seja no desenvolvimento desta Tese, ou em algum outro aspecto da minha vida durante o Doutorado. Palavras não seriam capazes de expressar tamanha gratidão. **MUITO OBRIGADO!**

Lacunas de conhecimento e evolução da dieta de peixes auchenipterídeos (Teleostei: Siluriformes)

RESUMO

A família Auchenipteridae é composta por 125 espécies de bagres distribuídas pela região Neotropical. Esses táxons apresentam uma incrível variabilidade morfológica e de nicho. Nos últimos anos esse grupo de peixes tem despertado grande interesse para pesquisa, como revisões taxonômicas e relações evolutivas, além de estudos autoecológicos (p. ex. aspectos alimentares e reprodutivos). A disponibilidade de informações sobre esses peixes faz da família Auchenipteridae um valioso caso de estudo para investigar diferentes questões ecológicas e evolutivas. Aqui, usamos um abrangente banco de dados taxonômico, distribuição geográfica, dados filogenéticos e de hábitos alimentares das espécies de auchenipterídeos. Realizamos uma série de estatísticas descritivas e analíticas para avaliar: a precisão das identificações taxonômicas desses peixes em repositórios on-line (Seção I), déficit em diferentes áreas do conhecimento (Seção II) e conservação filogenética do nicho trófico (Seção III). Na primeira sessão, evidenciamos um grande nível de imprecisões taxonômicas nos registros de repositórios online. Ressaltamos que tais bancos de dados devem ser utilizados com cautela, pois muitos problemas de imprecisão taxonômica podem estar presentes, o que pode levar os pesquisadores a fornecer uma perspectiva incompleta ou até equivocada das variações no mundo natural. Na segunda seção, nossos resultados mostraram que uma proporção substancial de táxons ainda deve ser descrita; também evidenciamos que os bagres auchenipterídeos permanecem subamostrados ao longo do Neotrópico. Por outro lado, sugere-se uma forte tendência a uma hipótese filogenética robusta dessa família em função do bom acumulo de informações. Já em relação ao atual conhecimento sobre hábitos alimentares dos auchenipterídeos, observamos lacunas notáveis tanto ao nível de gênero quanto de espécies. Por fim, na terceira seção, observamos que a maioria dos hábitos alimentares dos auchenipterídeos é conservada ao longo da filogenia. Porém a reconstrução da dieta do ancestral permaneceu incerta, mas sugerimos que se tratava de um peixe de hábitos onívoros oportunistas. Enfim, destacamos que apesar de inúmeros avanços no conhecimento dos peixes auchenipterídeos, ainda há muito trabalho a ser feito. Reduzir esses déficits exigirá um esforço conjunto de taxonomistas, ecólogos e biogeógrafos. Os resultados obtidos nessa Tese lançaram novas luzes sobre diferentes áreas de conhecimento de peixes auchenipterídeos, uma linhagem relevante entre a fauna de peixes neotropicais.

Palavras-chave: diversidade de peixes, repositório online, taxonomia, filogenia, reconstrução ancestral.

Knowledge shortfalls and diet evolution in auchenipterid fishes (Teleostei: Siluriformes)

ABSTRACT

The family Auchenipteridae harbors 125 catfish species distributed across the Neotropical region. Those taxa show an incredible morphological and niche variability and have been a great deal of interest in recent years, such as taxonomical reviews, evolutionary relationships, and autecological studies. Thus, the availability of information on these fishes makes the auchenipterid family a useful study case to investigate different ecological and evolutionary issues. Here we use a comprehensive taxonomic, geographic distribution, phylogenetic and dietary database of auchenipterid species. We performed a series of descriptive and analytic statistics to assess: the accuracy of nomenclature classification in online repositories (Section I), knowledge shortfalls (Section II), and the phylogenetic trophic niche conservatism (Section III). In the 1st section, we evidenced a large taxonomic inconsistency among records and that taxonomic information obtained from repositories should be used with caution. Many inaccuracy issues may be embedded in the biodiversity databases' records, which could lead researchers to provide an incomplete or even mistaken perspective of the variations in the natural world. In the 2nd section, our results showed that a substantial proportion of taxa may remain to be described; also that auchenipterid catfishes remain under-collected. On the other hand, we suggested a great tendency toward a robust phylogenetic hypothesis of this family. Noteworthy biases in the current knowledge of feeding habits were also identified across auchenipterid genera and species. Lastly, in the 3rd section, regarding the niche conservatism, our findings supported that most of the feeding habits of auchenipterid fishes are conserved across the phylogeny and the extant diet of their ancestor remains uncertain but was most likely a an opportunistic omnivorous feeder. Finally, there is still much work to be done and to reduce the ongoing shortfalls will require a concerted effort of taxonomists, ecologists, and biogeographers. Although, our results shed new light on different knowledge areas of auchenipterid fishes, which is a relevant lineage among the Neotropical fish fauna.

Keywords: fish diversity, online repository, taxonomy, phylogeny, ancestral reconstruction.

SUMÁRIO

1. APRESENTAÇÃO

É com muito prazer que torno público minha Tese de Doutorado. Primeiro, gostaria de convidar o leitor a conhecer um pouco mais dos bastidores desta pesquisa. Ao longo da minha graduação e mestrado, sempre orientado pelo Prof. Dr. Luciano Montag (também orientador desta Tese), dediquei atenção aos estudos de aspectos ecológicos de peixes amazônicos, como alimentação e reprodução, especialmente da família Auchenipteridae. No doutorado queria seguir a mesma linha de pesquisa, mas não tive uma ideia atrativa. Foi quando, em 2013, durante um congresso, um pesquisador me sugeriu investigar a ecologia evolutiva dos peixes dessa família. Aquela ideia me pareceu brilhante. Assim, alguns anos depois, tudo começou.

Em 2016, já iniciado o doutorado, conheci o Prof. Dr. Paulo De Marco Jr. (co-orientador desta Tese; Universidade Federal de Goiás), que me apresentou à temática “conservação filogenética do nicho”. Com essa abordagem, poderíamos explorar a ecologia trófica dos peixes sob uma perspectiva evolutiva. Partindo dessa ideia, alguns questionamentos surgiram: “estamos interessados em investigar a evolução da dieta dos auchenipterídeos, mas e quanto aos outros conhecimentos dessa família? O que sabemos sobre taxonomia e distribuição geográfica? Quanto nós sabemos sobre relações filogenéticas e hábitos alimentares do grupo?”. Iniciamos, então, o estudo das lacunas no conhecimento. Estudar essa temática me propiciou uma das melhores experiências da minha vida profissional e pessoal. Fui desenvolver o estudo de lacunas no *Museo Nacional de Ciencias Naturales*, em Madri, sob a supervisão do Prof. Dr. Joaquín Hortal. Por fim, no andamento da pesquisa das lacunas, nos chamou a atenção a quantidade de imprecisões taxonômicas presentes em um banco de dados de registros de ocorrência de auchenipterídeos e resolvemos investigar mais a fundo.

Contado a história de como essa pesquisa foi construída, vamos à versão científica dos fatos. Esta Tese é composta por três sessões, redigidas na língua inglesa, e formatada como artigos científicos de acordo com normas de revistas apropriadas para as respectivas temáticas. A primeira sessão foi intitulada “**How reliable are species identifications in biodiversity big data? Evaluating the records of a Neotropical fish family in online repositories**”, e teve como coautores Luciano Montag, Paulo De Marco e Joaquin Hortal. Nesta sessão, nós avaliamos a acurácia das identificações taxonômicas de registros de auchenipterídeos em dois repositórios online, *GBIF* e *SpeciesLink*. Nós evidenciamos, em ambos os repositórios, uma grande quantidade de imprecisões quanto à nomenclatura utilizada nesses bancos de dados, e ressaltamos que tais dados devem ser utilizados com cautela. Essas imprecisões podem levar os pesquisadores a fornecer uma perspectiva incompleta ou até equivocada de suas pesquisas. Esse estudo foi submetido à revista *Systematics and Biodiversity* (Qualis A2, Fator de Impacto = 2,215) e aceito para publicação em novembro de 2019.

A segunda sessão, intitulada "**Shortfalls in the knowledge on Neotropical Auchenipteridae fishes**", contou com a participação da Juliana Stropp (Universidade Federal de Alagoas), Bárbara Calegari (Pontifícia Universidade Católica do Rio Grande do Sul), Joaquín Calatayud (*Umea University*, Suécia), Paulo De Marco Jr., Luciano Montag e Joaquín Hortal. Esse estudo teve como objetivo quantificar e avaliar lacunas de conhecimento sobre os peixes Auchenipteridae no contexto taxonômico (déficit Linneano), distribuição geográfica (déficit Wallaceano), relações evolutivas (déficit Darwiniano) e hábitos alimentares (déficit Raunkiærano). Entre os principais resultados, ressaltamos: (i) ainda há uma proporção substancial de espécies por ser formalmente descrita; (ii) bagres auchenipterídeos são sub-amostrados ao longo do Neotrópico; (iii) os estudos mais recentes acerca das relações evolutivas das espécies são semelhantes entre si, sugerindo uma tendência a uma hipótese filogenética robusta para a família; e (iv) menos da metade das espécies válidas tinha alguma informação publicada sobre dieta, mas esse conhecimento vem sendo acumulado. Esse estudo está formatado e será submetido para a revista *Systematic Biology* (Qualis A1, Fator de Impacto = 10,266).

Por fim, a terceira sessão foi intitulada "**Niche conservatism in the diet of auchenipterid catfishes**", e contou com a colaboração de Lucas Jardim (Universidade Federal de Goiás), Bárbara Calegari (PUCRS), Fabricio Villalobos (*Instituto de Ecología*, México), Paulo De Marco Jr., Joaquín Hortal e Luciano Montag. Entre os principais resultados, evidenciamos que a maioria dos hábitos alimentares dos peixes auchenipterídeos é conservada, ou seja, há um visível efeito da história evolutiva do grupo na dieta. Ademais, através da reconstrução do estado ancestral, sugerimos que o ancestral dos auchenipterídeos apresentava um hábito alimentar onívoro oportunista. Essa pesquisa está formatada e será submetida para a revista *Journal of Evolutionary Biology* (Qualis A1, Fator de Impacto = 2,541).

2. SESSÃO I

How reliable are species identifications in biodiversity big data? Evaluating the records of a Neotropical fish family in online repositories

A primeira sessão desta tese foi aceita para publicação na revista *Systematics and Biodiversity*.

How reliable are species identifications in biodiversity big data? Evaluating the records of a Neotropical fish family in online repositories

Tiago M. S. Freitas[1,2], Luciano F. A. Montag[3], Paulo De Marco Jr.[4], Joaquín Hortal[4,5]

[1]Faculdade de Ciências Naturais, Campus Universitário do Marajó-Breves, Universidade Federal do Pará, Alameda IV, 3418, Parque Universitário, 68800-000, Breves, PA, Brazil. freitastms@gmail.com [2]Programa de Pós-graduação em Ecologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Rua Augusto Corrêa, 01, Guamá, 66075-110, Belém, PA, Brazil. [3]Laboratório de Ecologia e Conservação, Instituto de Ciências Biológicas, Universidade Federal do Pará, Rua Augusto Corrêa, 01, Guamá, 66075-110, Belém, PA, Brazil. lfamontag@gmail.com [4]Departamento de Ecologia, ICB Universidade Federal de Goiás. Avenida Esperança, 74001-970, Goiânia, GO, Brazil. pdemarcojr@gmail.com [5]Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/José Gutiérrez Abascal 2, 28006, Madrid, Spain. jhortal@mncn.csic.es

Taxonomic reliability in online repositories

2.1. ABSTRACT

The increase of free and open online biodiversity databases is of paramount importance for current research in ecology and evolution. However, little attention is paid to using updated taxonomy in these “biodiversity big data” repositories and the quality of their taxonomic information is often questioned. Here we assess how reliable is the current use of nomenclatural classification in the distributional information available from two biodiversity information networks GBIF and the Brazilian SpeciesLink. We use as a study case the records of Auchenipteridae, a Neotropical fish family that has been subject to recent taxonomical reviews. A data filtering procedure was applied to identify and quantify the inaccuracies in the taxonomical status of the records in three steps: assessment of identification accuracy at the family, genus or species level; current validity of species name; and assignation of inaccurate species records to different categories of classification quality. Synonyms, nonexistent combinations, and outdated combinations were reassigned to currently valid species. A total of 9,148 records of Auchenipteridae fishes were analyzed, of which 4,165 were from GBIF and 4,983 from speciesLink, deriving from 46 and 31 sources, respectively. After correcting all possible records following the taxonomic data filtering steps, 6,988 records (76.4% of the original) were adequate for describing species distributions, while 2,160 remained inaccurate. The most inaccurate records at the species level were due to the use of outdated nomenclatures, resulting in non-valid combinations of species and genus, and synonymy. Our results evidence a large taxonomic inconsistency among records, and, most importantly, that taxonomic information obtained from repositories should be used with caution. Many inaccuracy issues may be embedded in the biodiversity databases’ records, which could lead researchers to provide an incomplete or even mistaken perspective of the variations in the natural world.

***Keywords*:** Auchenipteridae fishes, biological collections, data quality, GBIF, Linnean shortfall, Neotropics, SpeciesLink, taxonomy

2.2. INTRODUCTION

Most biodiversity information is concentrated on scientific collections from institutions in developed countries (Edwards, Lane, & Nielsen, 2000; Penn, Cafferty, & Carine, 2018). As a result, the original vouchers for the first biodiversity data of a region in the Southern Hemisphere, are more likely to be stored in European or North American capitals rather than in their country of origin (Edwards et al., 2000; Vanzolini, 1996; von Staden, Raimondo, & Dayaram, 2012). Nevertheless, the increasing globalization of the world through the internet has made this information available worldwide. Now, the availability of biodiversity data in global information networks (e.g. Global Biodiversity Information Facility – GBIF. and NIH Genetic Sequence Database – GenBank) increase interoperability in natural sciences. Indeed, the growing information availability, together with the continuous advance in biodiversity informatics (e.g. development of mapping software and statistical analysis packages; Soberón & Peterson, 2004), has enabled large-scale analyses and eased the interpretation of biodiversity data (Maldonado et al., 2015).

The increase of free and open online databases dedicated to biodiversity is noteworthy. The vast amount of compiled information (hereafter referred to as biodiversity big data or big data for short) is of paramount importance in the current scenario of large-scale research and synthesis. Biodiversity big data is currently used by ecologists and evolutionary biologists to assess and unravel global patterns, considering a variety of groups of living beings. These data are gradually becoming essential to enable effective decision-making concerning biodiversity conservation (Canhos et al., 2015; Frederico, De Marco, & Zuanon, 2014; Soberón & Peterson, 2004). For instance, big data can be used to identify knowledge gaps (e.g. Hortal, Lobo, & Jimenez-Valverde, 2007; Hortal et al., 2015; Stropp et al., 2016), estimate total numbers of existing species using species description rates (Beck, Böller, Erhardt, & Schwanghart, 2014; Costello & Wilson, 2011; Maldonado et al., 2015), model species distributions to evaluate distributional gaps and propose priority sites for conservation (Frederico, Zuanon, & De Marco, 2018; Guisan et al., 2013; Ramirez-Villegas et al., 2014), and propose evolutionary hypotheses using genetic sequence databases (Dong et al., 2018; Faircloth, Sorenson, Santini, & Alfaro, 2013; Foley, Springer, & Teeling, 2016).

However, big data are not flawless. In fact, there are large concerns about biases, errors, and misuse of current big data on biodiversity research (Hortal et al., 2015). In fact, the available data on species occurrences typically lacks enough quality and coverage so as to provide adequate descriptions of the spatial-temporal dynamics of biodiversity distribution (Boakes et al., 2010; Hortal et al., 2007; Hortal, Jiménez-Valverde, Gómez, Lobo, & Baselga, 2008; Rocchini et al., 2011; Sousa-Baena, Garcia, & Peterson, 2014; Stropp et al., 2016). In this sense, Faith et al. (2013) proposed a list of key recommendations for GBIF data providers in order to improve the quality of records and avoid temporal and geographical biases. Indeed, several studies have pointed to the existence of important problems in GBIF data, such as errors and incorrect species nomenclature usage (Evenhuis, 2007; Smith, Johnston, & Lücking, 2016; Vecchione et al., 2000). Those studies paid superficial attention to the use of updated taxonomic classification, and the consequences of such neglect emerge as a noteworthy issue for different fields of biology, from genomics to ecology (Kennedy, Kukla, & Paterson, 2005).

The use of big data does not solve several significant issues that affect biodiversity studies. In fact, we have limited knowledge on how many species exist on Earth. The so-called "Linnean Shortfall" states that most of the biodiversity have not been described and cataloged (Brown & Lomolino, 1998). This shortfall is of paramount importance for ecological and evolutionary research (Hortal et al., 2015), as it is hard to achieve a full understanding of the consequences of environmental change and degradation if we cannot individually recognize each of the species that make up each assemblage (Mace, 2004, Valan, Makonyi, Maki, Vondráček, & Ronquist, 2019). Furthermore, the Linnean shortfall has important consequences for conservation since we will be unable to design conservation strategies for species that we do not even know about (Raposo et al., 2017).

Knowledge on fish is no exception to the Linnean shortfall. Assessing the correct taxonomic status of a certain taxon has been shown to have important effects on the study of fish biodiversity (Vecchione et al., 2000). According to these authors, several sources of uncertainty can affect the accuracy of species identifications, including: literature availability (e.g. older references are usually less accessible online); level of expertise of the people making the taxonomic identification – as some groups are more difficult to identify, and experience is always required to recognize the features that are adequate for species identification; and the condition of preserved specimens in museums or herbaria. All these mechanisms can lead to a common end, the failure in identifying a species.

With all this in mind, in the present study we seek to evaluate the taxonomic identification accuracy of the data in two biodiversity information networks –GBIF and the Brazilian SpeciesLink– using as a case study the records of a Neotropical fish family (Auchenipteridae). More precisely, we aim to assess how reliable is the nomenclatural classification usage in the fish records in these two databases. Herein, we interpret as reliable those records with complete (*Genus* + *epithet*) and accurate (i.e. currently valid) species identification. Auchenipterid fishes have been the subject of recent taxonomical reviews (Ferraris, Vari, & Raredon, 2005; Ribeiro, Rapp Py-Daniel, &Walsh, 2017; Walsh, Ribeiro & Rapp Py-Daniel, 2015) and phylogenetic studies (Birindelli, 2014; Calegari, Vari, & Reis, 2019). However, further revisionary and phylogenetic work is needed to ensure that the taxonomic and evolutionary status of all the species from this fish family is well established. As a result, it is reasonable to assume that failures and biases on species identification exist within the records currently available in the online repositories. Although GBIF and SpeciesLink databases potentially contain duplicate data within and between them, rather than comparing them we intended to present a state of the art of these two databases, one globally used (GBIF) and another one of Brazilian initiative (SpeciesLink).

2.3. MATERIALS AND METHODS

2.3.1. Data survey

We compiled a database on the taxonomic status of all records belonging to the Auchenipteridae family in the two main repositories of biodiversity data available for Brazil: GBIF (Global Biodiversity Information Facility; https://www.gbif.org/), and SpeciesLink (*Sistema de Informação Distribuído para Coleções Biológicas*; http://splink.cria.org.br/). We gathered from both databases all records assigned as Auchenipteridae in the field “Family”. Data were accessed and downloaded on 8 May 2017. The taxa names assigned to all records obtained from online repositories were checked in the Catalog of Fishes for the most current nomenclature (Fricke, Eschmeyer, & van der Laan, 2019). We applied three data filtering steps to identify and quantify the inaccuracies in the records concerning the taxonomical status (Fig. 1). First, record identifications were evaluated for different taxonomic levels, family, genus, or species. Second, records that were identified to the species level were classified as valid when no taxonomic error or inaccuracy was found; species’ name validity was checked against Catalog of Fishes (Fricke, Eschmeyer, & van der Laan, 2019). Finally, inaccurate species records were categorized and subdivided into seven categories (Table 1), as follows: 1) *inaccurate*: presence of arguments such as "aff.", "cf.", "gr."; 2) *nonexistent combination*: uses of an nonexistent binomial combination of genus + species, where either a genus was misattributed to a specific epithet or vice versa – note that in this case, it is possible to correct the combination; 3) *nonexistent species*: non-existent combination that was not possible to associate with spelling mistakes or misattribution of species or genus; 4) *not Auchenipteridae*: assigned to a species that does not belong to the Auchenipteridae Family; 5) *outdated combination*: a given taxonomic designation that no longer belongs to the cited genus. This occurs when a taxonomic revision revalidates a genus or describes a new one, relocating existing species; 6) *species inquirenda*: species of doubtful identity needing further investigation (according to the International Code of Zoological Nomenclature); and 7) *synonym*: when a given taxon has two or more distinct nomenclature proposed by different authors – in this case, the first published name is considered valid. After verification, records classified as ‘synonym’ and ‘outdated combination’ were corrected and became feasible for use. In our classification, the remaining inaccurate species records were impossible to correct, and therefore were not used for further analyses.

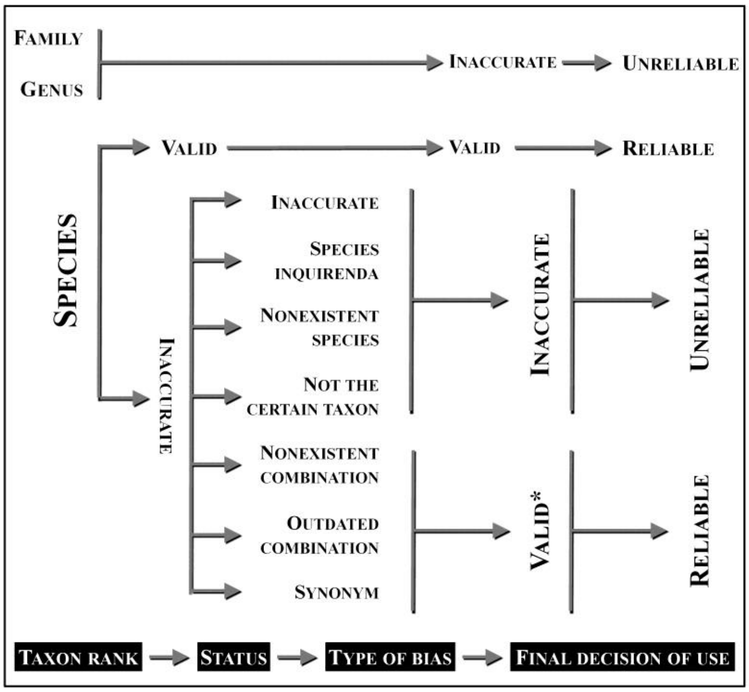


Fig. 1. Diagram of data filtering of the records of the Auchenipteridae fishes in online repositories. *After verification and validation, records were corrected and became reliable for use.

Table 1. Taxonomic bias in records of the Auchenipteridae fishes in online repositories.

Taxon rank	**Status**	**Type of bias**	**Description**
Family	Inaccurate	*Inaccurate*	Inaccurate identification at the family level.
Genus	Inaccurate	*Inaccurate*	Inaccurate identification at the genus level.
		Synonym	When the genus becomes synonymous with a more current one.
Species	Valid	*Misspelled species*	Any change in the spelling of a valid species.
		No error	No record errors.
	Inaccurate	*Inaccurate*	When the records present arguments such as "aff.", "cf.", "gr." for the identification of a species.
		Nonexistent combination	When the combination used was not found in the Catalog of Fishes, where either a genus could be safely assigned either to a specific species epithet or vice-versa.
		Nonexistent species	When the binomial combination is non-existent, and it is not possible to associate with possible spelling mistakes or misattribution of species or genus.
		Not Auchenipteridae	Assigned to a species that does not belong to the Auchenipteridae, although it is designated in the records.
		Outdated combination	When a record presents a species that no longer belongs to the cited genus. This occurs when a taxonomic revision revalidates a genus or describes a new one, relocating existing species.
		Species inquirenda	A Latin term meaning a species of doubtful identity needing further investigation.
		Synonym	When the species' name has become synonymous with one that preceded it.

2.3.2. Data analysis

We performed a series of descriptive statistics to evaluate the overall features of the data present in the online repositories based on the list of representative institutions providing records, number of records identified at family, genus and species level, number of inaccurate and valid species identifications, and number of records per categories of inaccurate records. We applied linear regression models (Zar, 2010) to evaluate the relationships between 1) the numbers of records identified at the genus level and the number of valid species per genus to verify if the records assigned at genus level belong to speciose genus or not, and the number of valid species per genus was checked in the Catalog of Fishes (Fricke, Eschmeyer, & van der Laan, 2019); and 2) the number of inaccurate records and the total number of records over time. For the last model, the y-intercept was equaled to zero in order to obtain the proportion of inaccurate records over the total number of records from the slope of the regression. A preliminary analysis using normal probability plots showed that all regressions produced normal residuals.

2.4. RESULTS

A total of 9,148 records of Auchenipteridae fishes were obtained from GBIF (4,165 records) and SpeciesLink (4,983 records), which were originally provided by 46 and 31 different scientific sources, respectively (Fig. 2). The National Institute of Amazonian Research (INPA; *Instituto Nacional de Pesquisas da Amazônia* – State of Amazonas, Brazil) was the largest record provider for both databases, with more than a thousand vouchers in each repository, accounting for nearly a quarter of all fish records of this family. Other providers such as University of São Paulo (USP, São Paulo, Brazil), Academy of Natural Sciences (ANSP, Pennsylvania, USA), and Pontifical Catholic University of Rio Grande do Sul (PUCRS, Porto Alegre, Brazil) also contributed with more than five hundred fish records each (Fig. 2).

Regarding species identification, records recognized at the species level summed up to 3,422 in GBIF and 3,854 in SpeciesLink, representing 82.2% and 77.3% of their respective totals (Fig. 3). Identification only at the genus level totaled 725 records (17.4%) for GBIF and 1,112 records (22.3%) for SpeciesLink. Less than twenty records (< 1%) identifying only at the family level were registered in each repository. Within records identifying at species level in GBIF, 2,934 were considered valid (85.7%) and 488 inaccurate (14.3%) (Fig. 3.1). Yet SpeciesLink showed 3,105 valid records (80.6%) and 749 with some taxonomic inaccuracy (19.4%) (Fig. 3.2). For GBIF, these inaccurate records were composed by outdated combination (363 records; 74.4%), synonym (102; 20.9%), *species inquirenda* (22; 4.5%), and nonexistent combination (1; 0.2%) (Fig. 3.3). While in SpeciesLink, inaccurate species records were composed by outdated combination (357 records; 47.7%), inaccurate records (165; 22.0%), synonym (88; 11.7%), *species inquirenda* (72; 9.6%), nonexistent combination (38; 5.1%), not an Auchenipteridae taxon (28; 3.7%), and nonexistent species (1; 0.1%) (Fig. 3.4). After correcting all possible records following these taxonomic data filtering steps (i.e. records categorized as nonexistent combination, outdated combination, and synonym categories), the number of records considered taxonomically accurate enough for their use raised from the original 3,108 valid records (34.0%) to a total of 6,988 records (76.4% of the original), while 2,160 remained inaccurate (23.6%). The corrected database is available as Supplementary Material S1.

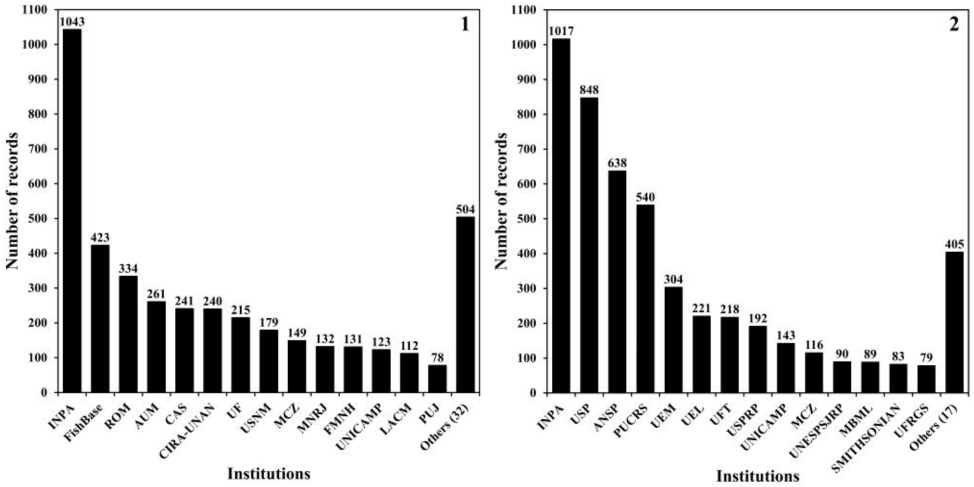


Fig. 2. Number of records of Auchenipterid fishes available per institution in the two Biodiversity Information Networks studied: **1**) Global Biodiversity Information Facility (GBIF), and **2**) Distributed Information System for Biological Collections (SpeciesLink). Source abbreviations follow Sabaj (2016).

Considering the 1,837 records with information available only at the genus level (GBIF and SpeciesLink, jointly), *Ageneiosus*, *Trachelyopterus,* and *Tatia* were the most abundant genera lacking proper species identifications (402, 360 and 312 records, respectively) (Fig. 4.1). A linear regression model shows a significant positive relationship between the number of valid species in a genus and the number of records with no valid species identification ($R^2 = 0.65$; $p < 0.001$; Fig. 4.2), evidencing a tendency of more incomplete identifications toward the most speciose genera. In addition, if we consider the two most relevant categories for records identified at species level (‘outdated combinations’ with 720 records and ‘synonymous’ with 190 records), we observed that *Parauchenipterus* (= *Trachelyopterus*) and *Ageneiosus* genera accounted for the highest outdated combinations records, with 395 and 174 records, respectively (Fig. 4.3); whereas the genus *Ageneiosus* presented 151 records classified as synonymous (Fig. 4.4).

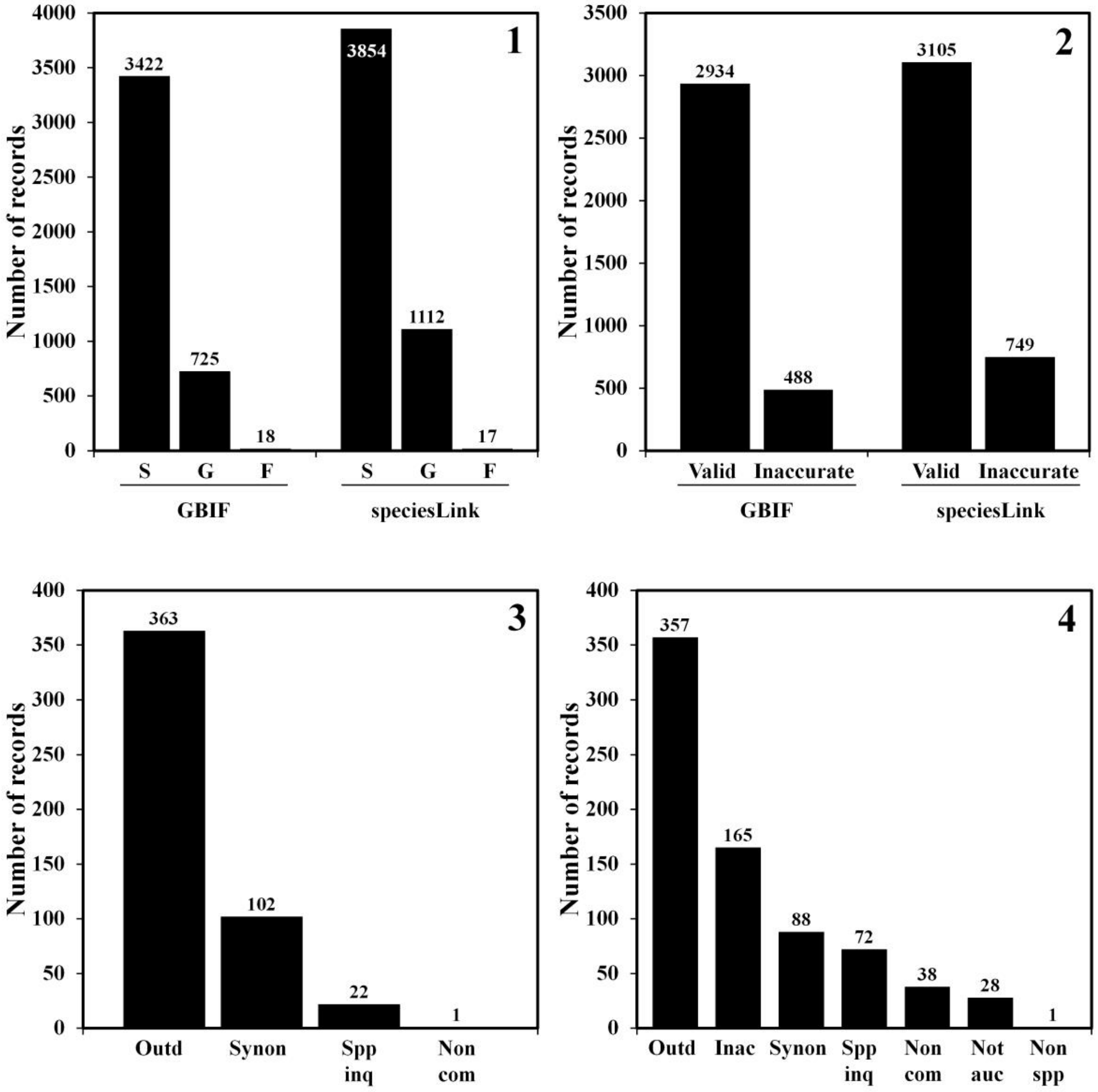


Fig. 3. Assessment of Auchenipteridae records obtained from Global Biodiversity Information Facility (GBIF) and Distributed Information System for Biological Collections (SpeciesLink); **1**) records identified at species (S), genus (G) and family level (F) for both repositories; **2**) valid or inaccurate records within records identified at species level; **3**) types of taxonomic bias of GBIF records; and **4**) types of taxonomic bias of speciesLink records. Type of bias: outdated combination (Outd), inaccurate (Inac), synonym (Synon), *species inquirenda* (Spp inq), nonexistent combination (Non com), not Auchenipteridae (Not auc), and nonexistent species (Non spp). The total number of records is shown above the bars.

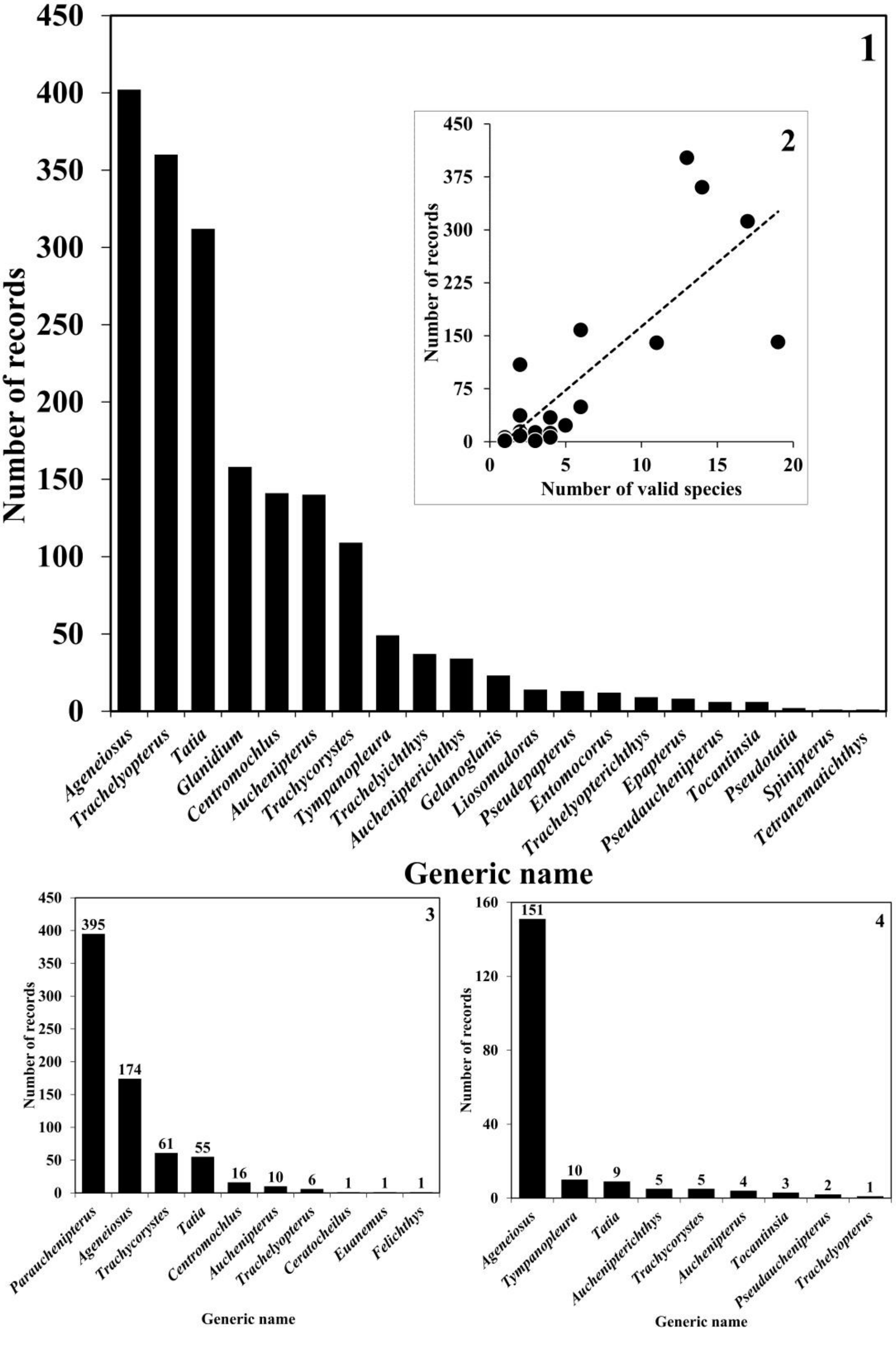


Fig. 4. Number of records identified at the genus level (**1**), and the tendency of doubtful taxonomy toward speciose genera (**2**); records with outdated combination (**3**) and synonyms names (**4**). The discontinuous line depicts the trend of higher number of records for the genera with most species identified by a linear regression model ($y = 18.079x - 17.556$, $N = 21$, $R^2 = 0.65$, $p < 0.001$).

The number of inaccurate records increases with the total number of records per year ($R^2 = 0.83$; $p < 0.001$; Fig. 5). The mean proportion of inaccurate records per year was 22.2%, as given by regression slope (b = 0.222). Interestingly, this proportion of erroneous records holds up through time; the numbers of both total and inaccurate records increase over the years at a relatively similar pace (Fig. 5).

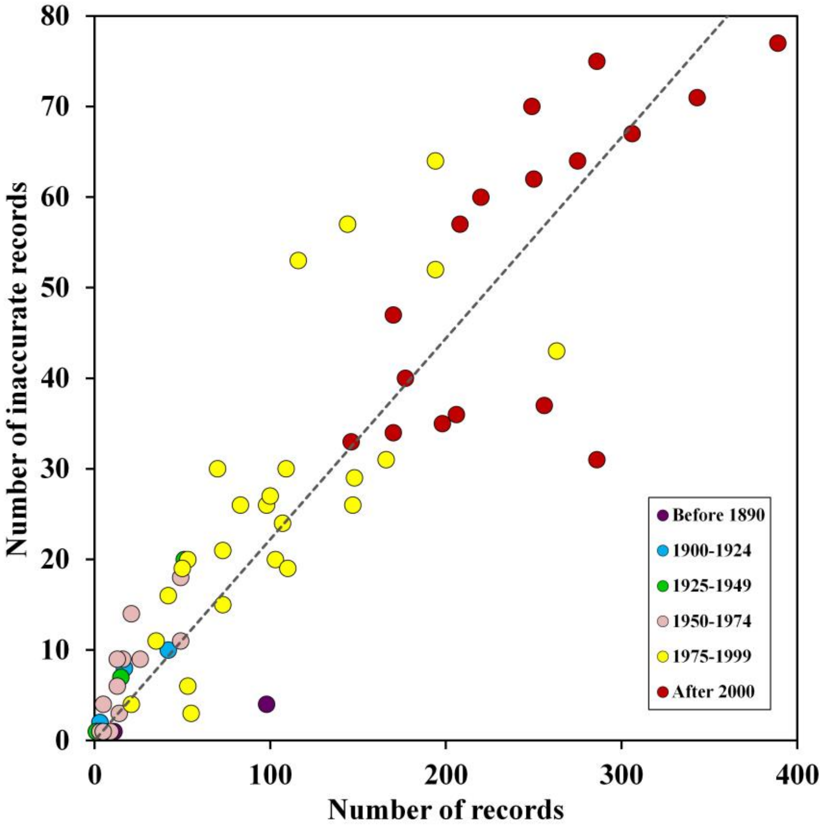


Fig. 5. Relationship between the number of inaccurate records and the total number of records of Auchenipteridae fishes per year. Circles represent years, colors different temporal periods, and the dotted line represents a linear regression model (y = 0.22x; N = 70, $R^2 = 0.83$, p = < 0.001).

2.5. DISCUSSION

Our application of taxonomic data filtering shows that more than 20% (> 2,000) of the total number of records of Auchenipterid fishes available online in GBIF and SpeciesLink biodiversity databases are taxonomically inaccurate. Most of the inaccurate records at the species level were due to the use of outdated nomenclatures such as the outdated combination of species and genus, and synonymy, which would be easy to correct by consulting online databases and scientific literature. Inaccurate records at the genus level, however, hold a more complex taxonomic deficiency.

2.5.1. Gaps of taxonomic knowledge in Auchenipteridae fish

The majority of taxonomic inaccuracies pertain to *Ageneiosus*, *Tatia*, and *Trachelyopterus*, three of the four most diverse genera in the number of species among Auchenipteridae (Fricke et al., 2019). *Trachelyopterus* (described by Cuvier and Valenciennes, 1840) presents a conflicting taxonomy with *Parauchenipterus* (described by Bleeker, 1862), considered as distinct genera for more than 140 years but synonymized with no additional comments by Ferraris Jr. (2003) in the ‘Check List of the Freshwater Fishes of South and Central America’. However, this synonymizing seems not to be broadly applied since *Parauchenipterus* has been continuously used in recent publications (Casatti, Pérez-Mayorg, Carvalho, Brejão, & Costa, 2016; Ferreira, Zuanon, Santos, & Amadio, 2011; Frota, Deprá, Petenucci, & Graça, 2016). Besides, no new *Trachelyopterus* (or *Parauchenipterus*) species has been described over the past 20 years. Therefore, a taxonomic revision is needed to better establish the nomenclatural status of these genera. Otherwise, confusing identifications will probably continue to occur, promoting disparity between databases of scientific collections.

In the case of *Tatia*, classification confusions are also frequently described in the literature (Sarmento-Soares, Lazzarotto, Rapp Py-Daniel, & Leitão, 2016; Vari & Calegari, 2014), which may result in misleading identifications in current records. In this case, uncertain taxonomy mainly occurs due to the difficulty in distinguishing the boundaries between *Centromochlus* and *Tatia*, two sister taxa within the subfamily Centromochlinae (Birindelli, 2014). Species reassignments between these genera are frequently reported (Sarmento-Soares & Birindelli, 2015; Sarmento-Soares & Martins-Pinheiro, 2008; Soares-Porto, 1998). This is more explicit in Sarmento-Soares et al. (2016), who diagnosed a new species of *Centromochlus* basically based on a *Tatia* species. The latest phylogenetic arrangement proposed for Auchenipteridae aimed to clarify the relationships between these taxa with a revised diagnosis of them, which resulted in the reorganization and splitting of both taxa into several new genera (Calegari, Vari, & Reis, 2019). This brand-new proposal of classification had no effect on the general patterns of taxonomic quality described here, as it has made minimal changes in species epithets, thereby increasing only the number of records with outdated combinations, which as discussed above is straightforward to correct.

The case of *Ageneiosus* is different because this genus has been recently reviewed (Ribeiro et al., 2017), with a description of four new species in the publication. However, before this revisionary study, *Tympanopleura* (previously assigned as synonymous to *Ageneiosus*) was re-elevated to genus status, and now includes some species previously placed within *Ageneiosus* (Walsh et al., 2015). This recent improvement in taxonomic knowledge still needs to be translated to the scientific collections, which need to update their databases as soon as possible after a voucher checking.

2.5.2. Impact of taxonomic inaccuracies on biodiversity research

Online repositories have been key to the recent development of macroecological studies (Ariño, Chavan, & Faith, 2013; Canhos et al., 2015; Costello, 2014). However, the widespread taxonomic inaccuracies that show up in our analysis demonstrate the importance of being up-to-date with current scientific literature, nomenclature, and species identification in biodiversity research. Interestingly, as the number of auchenipterid records increases over time, they also bring an increase in inaccurate records. This is contrary to our previous expectation of more improved taxonomy of the most recent records as, arguably, the taxonomic revisions made in the last years should have contributed to higher accuracy of identifications. In our view, the accumulation of more than 20% of erroneous records over time is a large and worrisome figure, and it highlights that it is crucial that taxonomists participate in fieldwork and/or that species identifications be made by trained personnel. Attention should be paid to the relative importance of these two sources of error, but the truth is that something backfired, and we need to know why.

Most of the records on auchenipterids are deposited in ichthyological collections in Brazilian institutions – which is, by the way, the country holding the highest diversity of fish species in South America (Reis et al., 2016). However, keeping records of a regularly updated scientific collection are a difficult task. Considering the remarkable fish diversity in South America, we highlight the relevance of taxonomic studies and argue that more taxonomists are needed, and more investments should be made in training them and in promoting taxonomic review projects. This presents a gloomy perspective if we consider the current state of scientific investment in megadiverse countries such as Brazil (Oliveira et al., 2016), which has been decreasing funding over the recent years (Angelo, 2017; Escobar, 2016; Fernandes et al., 2017).

The existence of mislabeled specimens in, arguably, well-curated scientific collections, reflects a potentially grave problem, which goes far beyond the walls of museums and herbaria (James et al., 2018). Although species identification is crucial for biodiversity science (Grieneisen, Zhan, Potter, & Zhang, 2015) – “and increasingly relevant to the very future of humankind”– taxonomy is underfunded and underappreciated (Drew, 2011; Ota, Message, Graça, & Pavanelli, 2015). Taxonomy is indeed a science under constant construction (Thomson et al., 2018). Our particular studied taxon, Auchenipteridae, is still considered to suffer from low sampling (Ota et al., 2015) despite the large increase in the knowledge of Neotropical fish diversity over recent years (Reis et al., 2016). No matter how widespread they are, these significant taxonomic gaps constitute a key component of the Linnean shortfall, hampering our perception of the actual biodiversity patterns (see Hortal et al., 2015). Therefore, the success of both research in ecology and evolution, and systematic conservation planning depend on the continuous investment in the description and classification of living beings (Thomson et al., 2018).

All in all, our results demonstrate that the taxonomic classification of Auchenipteridae fishes in collections should be constantly updated. It is possible –and indeed likely– that the evidence for the impact of an inaccurate nomenclature on the reliability of a significant proportion of species records presented here can be extrapolated to many other taxonomic groups (see, e.g., Stropp et al., 2016). In the current scientific panorama, the data stored in natural history collections and compiled in biodiversity information networks are used to investigate global biodiversity issues. Therefore, it is imperative that these "biodiversity big data" are used with caution because many issues of inaccuracy and or bias may be embedded within their records, which would lead researchers to inaccurate or biased results caused by an incomplete – or even mistaken– perspective of the variations of the natural world.

ACKNOWLEDGEMENTS

We wish to thank Juliana Stropp, Guilherme Dutra, and an anonymous reviewer for their help in improving this manuscript. The authors are grateful to the financial support provided by several grants from the Brazilian *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES - Finance Code 001; TMSF - 23038.042984/ 2008-30), and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq; PDM - 308694/2015-5; LFAM - 305017/2016-0; JH – PVE 314523/2014-6).

2.6. REFERENCES

Angelo, C. (2017). Brazil’s scientists battle to escape 20-year funding freeze. *Nature, 539*, 480. doi: 10.1038/nature.2016.21014

Ariño, A., Chavan, V., & Faith, D. (2013). Assessment of user needs of primary biodiversity data: Analysis, Concerns, and Challenges. *Biodiversity Informatics, 8*, 59–93. doi: 10.17161/bi.v8i2.4094

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions. *Ecological Informatics, 19*, 10–15. doi: 10.1016/j.ecoinf.2013.11.002

Birindelli, J. L. O. (2014). Phylogenetic relationships of the South American Doradoidea (Ostariophysi: Siluriformes). *Neotropical Ichthyology, 12,* 451–564. doi: 10.1590/1982-0224-20120027

Boakes, E. H., McGowan, P. J., Fuller, R. A., Chang-Qing, D., Clark, N. E., O’Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *Public Library of Science Biology, 8*, e1000385. doi: 10.1371/journal.pbio.1000385

Brown, J. H., & Lomolino, M. V. (1998). *Biogeography*. 2 ed. Massachussets: Sinauer Associates.

Calegari, B. B., Vari, R. P., & Reis, R. E. (2019). Phylogenetic systematics of the driftwood catfishes (Siluriformes: Auchenipteridae): a combined morphological and molecular analysis. *Zoological Journal of the Linnean Society*. doi: 10.1093/zoolinnean/zlz036

Canhos, D. A. L., Sousa-Baena, M. S., Souza, S., Maia, L. C., Stehmann, J. R., Canhos, V. P., ... Peterson, A. T.(2015). The importance of biodiversity e-infrastructures for megadiverse countries. *Public Library of Science Biology, 13,* e1002204. doi: 10.1371/journal.pbio.1002204

Casatti, L., Pérez-Mayorg, M. A., Carvalho, F. R., Brejão, G. L., & Costa, I. D. (2016). The stream fish fauna from the rio Machado basin, Rondônia State, Brazil. *Check List 9*, 1496–1504. doi: 10.15560/9.6.1496

Costello, M. J. (2014). Advancing online databases for global biodiversity conservation. *Biological Conservation, 173*, 65–165. doi: 10.1016/j.biocon.2014.04.008

Costello, M. J., & Wilson, S. P. (2011). Predicting the number of known and unknown species in European seas using rates of description. *Global Ecology and Biogeography, 20*, 319–330. doi: 10.1111/j.1466-8238.2010.00603.x

Dong, W. L., Wang, R. N., Zhang, N. Y., Fan, W. B., Fang, M. F., Li, Z. H. (2018). Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution. *International Journal of Molecular Sciences, 19*, 716. doi: 10.3390/ijms19030716

Drew, L. W. (2011). Are we losing the science of taxonomy? *BioScience, 61*, 942-946. doi: 10.1525/bio.2011.61.12.4

Edwards, J. L., Lane, M. A., & Nielsen, E. S. (2000). Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science, 289*, 2312–2314. doi: 10.1126/science.289.5488.2312

Escobar, H. (2016). Budget cap would stifle Brazilian science, critics say. *Science, 364*, 6446. doi: 10.1126/science.aal0276

Evenhuis, N. L. (2007). Helping solve the “other” taxonomic impediment: completing the Eight Steps to Total Enlightenment and Taxonomic Nirvana. *Zootaxa, 1407*, 3–12. doi: 10.11646/zootaxa.1494.1.3

Faircloth, B. C., Sorenson, L., Santini, F., & Alfaro, M. E. (2013). A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *Public Library of Science One, 8*, e65923. doi: 10.1371/journal.pone.0065923

Faith, D. P., Collen, B., Ariño, A. H., Koleff, P. O., Guinotte, J. M., Kerr, J. T., & Chavan, V. (2013). Bridging Biodiversity Data Gaps: Recommendations to Meet Users’ Data Needs. *Biodiversity Informatics, 8*, 41–58. doi: 10.17161/bi.v8i2.4126

Fernandes, G. W., Vale, M. M., Overbeck, G. E., Bustamante, M. M. C., Grelle, C. E. V., Bergallo, H. G. … Pillar, V. D. (2017). Dismantling Brazil’s science threatens global biodiversity heritage. *Perspectives in Ecology and Conservation, 15*, 239–243. doi: 10.1016/j.pecon.2017.07.004

Ferraris Jr., C. J. (2003). Family Auchenipteridae (Driftwood catfishes). In R. E. Reis, S. O. Kullander & C. J. Ferraris, Jr. (Eds.), *Checklist of the freshwater fishes of South and Central America* (pp. 470–482). Porto Alegre, RS: Edipucrs.

Ferraris Jr., C. J., Vari, R. P., & Raredon, S. J. (2005). Catfishes of the genus *Auchenipterichthys* (Osteichthyes: Siluriformes: Auchenipteridae): a revisionary study. *Neotropical Ichthyology, 3*, 89–106. doi: 10.1590/S1679-62252005000100005

Ferreira, E., Zuanon, J., Santos, G., & Amadio, S. (2011). A ictiofauna do Parque Estadual do Cantão, Estado do Tocantins, Brasil. *Biota Neotropica, 11*, 277–285. doi: 10.1590/S1676-06032011000200028

Foley, N. M., Springer, M. S., & Teeling, E. C. (2016). Mammal madness: is the mammal tree of life not yet resolved? *Philosophical Transactions of the Royal Society of London B, 371*, 20150140. doi: 10.1098/ rstb.2015.0140

Frederico, R. G, Zuanon, J., De Marco Jr., P. (2018). Amazon protected areas and its ability to protect stream-dwelling fish fauna. *Biological Conservation, 219*, 12–19. doi: 10.1016/j.biocon.2017.12.032

Frederico, R. G., De Marco Jr., P., & Zuanon, J. (2014). Evaluating the use of macroscale variables as proxies for local aquatic variables and to model stream fish distributions. *Freshwater Biology, 59*, 2303–2314. doi: 10.1111/fwb.12432

Fricke, R., Eschmeyer, W.N., van der Laan, R. (2019). Eschmeyer's Catalog of Fishes: Genera, Species, References. Retrieved from: https://www.calacademy.org/scientists/projects/catalog-of-fishes (assessed 3 June 2019)

Frota, A., Deprá, G. C., Petenucci, L. M., & Graça, W. J. (2016). Inventory of the fish fauna from Ivaí River basin, Paraná State, Brazil. *Biota Neotropica, 16*, e20150151. doi: b2sh

Grieneisen, M. L., Zhan, Y., Potter, D., & Zhang, M. (2015). Biodiversity, taxonomic infrastructure, international collaboration, and new species discovery. *BioScience, 64*, 322–332. doi: 10.1093/biosci/biu035

Guisan, A., Tingley, R., Baumgartiner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters, 16*, 1424–1435. doi: 10.1111/ele.12189. PMid:24134332.

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M. L., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset largescale knowledge on biodiversity. *Annual Review of Ecology, Evolution, and Systematics, 46*, 523–549. doi: annurev-ecolsys-112414-054400

Hortal, J., Jiménez-Valverde, A., Gómez, J. F., Lobo, J. M., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed realized niche of the species. *Oikos, 117*, 847–858. doi: 10.1111/j.0030-1299.2008.16434.x

Hortal, J., Lobo, J. M., & Jimenez-Valverde, A. (2007). Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conservation Biology, 21,* 853–863. doi: 10.1111/j.1523-1739.2007.00686.x

James, S. A., Soltis, P. S., Belbin, L., Chapman, A. D., Nelson, G., Paul, D. L., Collins, M. (2018). Herbarium data: Global biodiversity and societal botanical needs for novel research. *Applications in Plant Sciences, 6,* e1024. doi: 10.1002/aps3.1024

Kennedy, J., Kukla, R., & Paterson, T. (2005). Scientific names are ambiguous as identifiers for biological taxa: their context and definition are required for accurate data integration. In B. Ludäscher & L. Raschid (Eds.), *Data Integration in the Life Sciences*: Second International Workshop (pp. 80–95). California, CA: Springer Verlag.

Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London B, 359*, 711–719. doi: 10.1098/rstb.2003.1454

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., … Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography, 24,* 973–984. doi: 10.1111/geb.12326

Oliveira, U., Paglia, A. P., Brescovit, A. D., Carvalho, C. J. B., Silva, D. P., Rezende, D. T., ... Santos, A. J. (2016). The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. *Diversity and Distributions,* 1–13. doi: 10.1111/ddi.12489

Ota, R. B., Message, H. J., Graça, W. J., & Pavanelli, C. S. (2015). Neotropical Siluriformes as a Model for Insights on Determining Biodiversity of Animal Groups. *Public Library of Science One, 10*, e0132913. doi: 10.1371/journal.pone.0132913.

Penn, M. G., Cafferty, S., & Carine, M. (2018). Mapping the history of botanical collectors: spatial patterns, diversity, and uniqueness through time. *Systematics and Biodiversity, 16*, 1–13. doi: 10.1080/14772000.2017.1355854

Ramirez-Villegas, J., Cuesta, F., Devenish, C., Peralvo, M., Jarvis, A., & Arnillas, C. A. (2014). Using species distributions models for designing conservation strategies of Tropical Andean biodiversity under climate change. *Journal for Nature Conservation, 22*, 391–404. doi: 10.1016/j.jnc.2014.03.007

Raposo, M. A., Stopiglia, R., Brito, G. R. R., Bockmann, F. A., Kirwan, G. M., Gayon, J., & Dubois, A. (2017). What really hampers taxonomy and conservation? A riposte to Garnett and Christidis. *Zootaxa, 4317*, 179–184. doi: 10.11646/zootaxa.4317.1.10

Reis, R. E., Albert, J. S., Di Dario, F., Mincarone, M. M., Petry, P. & Rocha, L. A. (2016). Fish biodiversity and conservation in South America. *Journal of Fish Biology, 89*, 12–47. doi: 10.1111/jfb.13016.

Ribeiro, F. R. V., Rapp Py-Daniel, L.H., & Walsh, S. J. (2017). Taxonomic revision of the South American catfish genus *Ageneiosus* (Siluriformes: Auchenipteridae) with the description of four new species. *Journal of Fish Biology, 90*, 1388–1478. doi: 10.1111/jfb.13246

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G. & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography, 35*, 211–226. doi: 10.1177/0309133311399491

Sabaj, M. H. (2016). *Standard symbolic codes for institutional resource collections in herpetology and ichthyology*: an Online Reference. Retrieved from: http://www.asih.org/ (assessed 3 June 2019).

Sarmento-Soares, L. M, Lazzarotto, H., Rapp Py-Daniel, L., & Leitão, R. P. (2016). A new *Centromochlus* Kner, 1858 (Siluriformes: Auchenipteridae: Centromochlinae) from the transition between Amazon floodplain and Guiana shield, Brazil. *Neotropical Ichthyology, 14*, e160030. doi: 10.1590/1982-0224-20160030

Sarmento-Soares, L. M., & Birindelli, J. L. O. (2015). A new species of the catfish genus *Centromochlus* (Siluriformes: Auchenipeteridae: Centromochlinae) from the upper rio Paraná basin, Brazil. *Neotropical Ichthyology, 13,* 77–86. doi: 10.1590/1982-0224-20140042

Sarmento-Soares, L. M., & Martins-Pinheiro, R. F. (2008). A systematic revision of *Tatia* (Siluriformes: Auchenipteridae: Centromochlinae). *Neotropical Ichthyology, 6,* 495–542. doi: 10.1590/S1679-62252008000300022

Smith, B. E., Johnston, M. K., & Lücking, R. (2016). From GenBank to GBIF: Phylogeny-Based Predictive Niche Modeling Tests Accuracy of Taxonomic Identifications in Large Occurrence Data Repositories. *Public Library of ScienceOne, 11*, e0151232. doi: 10.1371/journal.pone.0151232

Soares-Porto, L. M. (1998). Monophyly and interrelationships of the Centromochlinae (Siluriformes, Auchenipteridae). In L. R. Malabarba, R. E. Reis, R. P. Vari, Z. M. S. Lucena & C. A. S. Lucena (Eds.), *Phylogeny and Classification of Neotropical fishes* (pp. 331–350). Porto Alegre, RS: Edipucrs.

Soberón, J., & Peterson, T. (2004). Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B, 359*, 689–698. doi: 10.1098/rstb.2003.1439

Sousa-Baena, M. S., Garcia, L. C., & Peterson, A. T. (2014). Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions, 20*, 369–381. doi: 10.1111/ddi.12136

Stropp, J., Ladle, R. J., Malhado, A. C. M., Hortal, J., Gaffuri, J., Temperley, W. H., Skøien, J. O., & Mayaux, P. (2016). Mapping ignorance: 300 years of collecting flowering plants in Africa. *Global Ecology and Biogeography, 25*, 1085–1096. doi: 10.1111/geb.12468

Thomson, S. A., Pyle, R. L., Ahyong, S. T., Alonso-Zarazaga, M., Ammirati, J., Araya, J. F., ... Zhou, H. Z. (2018). Taxonomy based on science is necessary for global conservation. *Public Library of Science Biology, 16,* e2005075. doi: 10.1371/journal.pbio.2005075

Valan, M., Makonyi, K., Maki, A., Vondráček, D., & Ronquist F. (2019). Automated Taxonomic Identification of Insects with Expert-Level Accuracy Using Effective Feature Transfer from Convolutional Networks. *Systematic Biology, syz014*. doi: 10.1093/sysbio/syz014

Vanzolini. P. E. (1996). *Brasil dos Viajantes: A contribuição Zoológica dos primeiros naturalistas viajantes no Brasil*. São Paulo: USP.

Vari, R. P. & Calegari, B. B. (2014). New species of the catfish genus *Tatia* (Siluriformes: Auchenipteridae) from the rio Teles Pires, upper rio Tapajós basin, Brazil. *Neotropical Ichthyology, 12*, 667–674. doi: 10.1590/1982-0224-20130193

Vecchione, M., Mickevich, M. F., Fauchald, K., Collette, B. B., Williams, A. B., Munroe, T. A., & Young, R. E. (2000). Importance of assessing taxonomic adequacy in determining fishing effects on marine biodiversity. *ICES Journal of Marine Science*, *57*, 677–681. doi: 10.1006/jmsc.2000.0707

von Staden L., Raimondo, D., & Dayaram, A. (2013). Taxonomic research priorities for the conservation of the South African flora. *South African Journal of Science, 109*, 1182. doi: 10.1590/sajs.2013/1182

Walsh, S. J., Ribeiro, F. R. V., & Rapp Py-Daniel, L. H. (2015). Revision of *Tympanopleura* Eigenmann (Siluriformes: Auchenipteridae) with description of two new species. *Neotropical Ichthyology*, *13*, 1–46. doi: 10.1590/1982-0224-20130220

Zar, J. H. (2010). *Biostatistical analysis*. 5th ed. New Jersey: Prentice Hall.

3. SESSÃO II

Shortfalls in the knowledge on Neotropical Auchenipteridae fishes

A segunda sessão desta tese será formatada conforme as normas da publicação científica *Systematic Biology*, disponível em: https://academic.oup.com/sysbio

Shortfalls in the knowledge on Neotropical Auchenipteridae fishes

Tiago M.S. Freitas[1,2], Juliana Stropp[3,4], Bárbara B. Calegari[5,6], Joaquín Calatayud[7],
Paulo De Marco Jr.[8], Luciano F.A. Montag[9], Joaquín Hortal[4,8]

[1]Faculdade de Ciências Naturais, Campus Universitário do Marajó-Breves, Universidade Federal do Pará, Alameda IV, 3418, Parque Universitário, 68800-000, Breves, PA, Brazil. freitastms@gmail.com. [2]Programa de Pós-graduação em Ecologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Rua Augusto Corrêa, 01, Guamá, 66075-110, Belém, PA, Brazil. [3]Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Av. Lourival Melo Mota, Tabuleiro do Martins, 57072-900, Maceió, AL, Brazil. justropp@gmail.com. [4]Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/José Gutiérrez Abascal 2, 28006, Madrid, Spain. jhortal@mncn.csic.es. [5]Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS); Laboratório de Sistemática de Vertebrados. Av. Ipiranga, 6681, CEP: 90619-900, Porto Alegre, RS, Brazil. [6]Smithsonian Institution; Department of Vertebrate Zoology, National Museum of Natural History, 20560, Washington, DC, USA. barbara.calegari@gmail.com. [7]Integrated Science Lab, Department of Physics, Umea University, SE-901 87 Umea, Sweden. j.calatayud.ortega@gmail.com. [8]Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Avenida Esperança, 74001-970, Goiânia, GO, Brazil. pdemarcojr@gmail.com. [9]Laboratório de Ecologia e Conservação, Instituto de Ciências Biológicas, Universidade Federal do Pará, Rua Augusto Corrêa, 01, Guamá, 66075-110, Belém, PA, Brazil. lfamontag@gmail.com.

Knowledge shortfalls on auchenipterid fishes

3.1. ABSTRACT

The Neotropics harbor the greatest diversity of freshwater fishes on Earth. Despite recent advances in characterizing this diverse group of vertebrates, key shortfalls in the knowledge of its total number of species, geographical distribution, evolution, and ecological traits, remain poorly explored. Quantifying the extent of each of these shortfalls is essential for planning future research and guiding conservation efforts. This study aims to quantify key aspects of four knowledge shortfalls in Auchenipteridae driftwood catfishes: taxonomical knowledge (Linnean shortfall); geographical distribution (Wallacean shortfall); evolutionary relationships (Darwinian shortfall); and feeding habits (Raunkiæran Shortfall). The family Auchenipteridae is widely distributed across the Neotropics and presents considerable morphological and niche variability. The current status of the Linnean shortfall was assessed by analyzing the historical accumulation of the 123 valid species names. To address the Wallacean shortfall, we compiled a dataset containing 10,336 records of Auchenipterid catfishes, out of which 5,393 were deemed suitable for our analysis. Using these data, we estimated inventory completeness for all Neotropical freshwater ecoregions and the cells from a one-degree spatial resolution grid. To quantify the Darwinian shortfall, we explored the dissimilarities among the different phylogenetic hypotheses proposed for Auchenipteridae through time. To address the Raunkiærian shortfall, we assessed the quality and completeness of current knowledge about the trophic ecology using published scientific literature. Our results show a steep increase in the historical accumulation of valid species names over time (1766-2018), suggesting that a substantial proportion of taxa may remain to be described. The Auchenipterid catfishes remain largely under-collected; only 45% of the ecoregions and less than three percent of the grid-cells of the Neotropics with occurrence data can be considered as reasonably sampled. The temporal variation in the accumulated number of grids for the most recorded species indicates that the yearly rate of increase in the number of new explorations is slower before 1980, increasing after this date. The topologies of recent phylogenies are more similar to each other than former ones, showing a clear tendency toward a robust phylogenetic hypothesis for this family. Current knowledge on feeding habits presents biases across genera and species and is still accumulating with every new study published. In sum, there is still much work to be done, as more species are expected to be described and most of the already described species lack reliable information regarding their geographical distribution. Reducing these gaps will require a concerted effort of taxonomists, ecologists, and biogeographers.

Keywords: Taxonomy; Geographic distribution; Evolution; Catfish; Neotropics; Linnean shortfall; Wallacean shortfall; Darwinian shortfall; Raunkiærian shortfall.

3.2. INTRODUCTION

Discovering and describing how many species there are on Earth remain as pivotal pursuits for biologists. Yet, acute changes in the Earth's environments like climate changes and widespread loss of natural habitat threaten many species that remain virtually invisible to science (Zhang 2011). The risk of losing species without knowing them is aggravated by decreasing investments for research (Troudet et al. 2017; Thomson et al. 2018) and increasing threats to natural habitats (Joppa et al. 2011). In this context, identifying and quantifying the current shortfalls of our knowledge on biodiversity can help to define priorities for effective ecological and evolutionary studies as well as for conservation-related decisions.

Knowledge of the distribution, ecology, and evolution of most taxa on Earth is even scarcer or absent (Hortal et al. 2015; Cornwell et al. 2019; Dagosta and Pinna 2019). Such knowledge paucity matters because when taxonomy is inaccurate it leads to inconsonant or doubtful definitions of the geographic range of species, their phylogenetic relationships, or estimations on species abundance and population dynamics (Lomolino 2004; Cardoso et al. 2011; Hortal et al. 2015). Further, the lack of knowledge on species distributions precludes the identification of species feeding habits and tolerance to environmental conditions (Cardoso et al. 2011). It follows that these so-called knowledge shortfalls are thus related to each other (Hortal et al. 2015; Cornwell et al. 2019). Yet, most studies quantifying deficiencies in the current knowledge of species focus on specific shortfalls (Lobo et al. 2007; Stropp et al. 2016; Troudet et al. 2017), a pragmatic approach that may, however, prevent from providing an overall picture of how scientific progress is filling in the knowledge gaps on biodiversity information.

The shortfalls in biodiversity knowledge can be defined as the gaps between realized/extant knowledge and the complete knowledge of any aspect of biodiversity within a biological domain at a given moment in time (Hortal et al. 2015). Given the limited resources for biodiversity research (Barber et al. 2014), it is crucial to develop scientifically sound criteria for quantifying the various knowledge shortfalls. These criteria can guide us to direct scientific effort to persistent and/or easy-to-fill knowledge gaps. In this context, seven knowledge shortfalls covering key aspects of biodiversity have been proposed, aiming to account for the lack of knowledge on the total species richness on Earth (Linnean); the geographical distribution of species (Wallacean); species abundance and population dynamics (Prestonian); the evolution and phylogenetic relationships of species or lineages (Darwinian); species tolerances to abiotic factors (Hutchinsonian); species traits and functions (Raunkiæran); and biotic interactions (Eltonian) (see Hortal et al., 2015 for a complete description).

In this work, we studied and quantified several knowledge gaps on the distribution, ecology, evolution, and feeding habit using the Neotropical fish family Auchenipteridae as a study case. Particularly, fishes are among vertebrates the most diverse group and also the one who contains the higher estimates of the total number of species. This is especially the case for the vast Neotropical ichthyofauna (Lundberg et al. 2000; Ota et al. 2015) since current estimates standing range between 8,000 and 9,000 species (Reis et al. 2016). The Auchenipteridae, composed of more than 120 species (Calegari et al. 2019) is one of the most representative in the order Siluriformes (catfishes). This family is widely-distributed along all major cis-Andean basins, with few species in the Caribbean's trans-Andean drainages and a single species in the Pacific slope (Ferraris Jr. 2003). Auchenipterids are unique among catfishes in at least one aspect of their reproductive biology: an internal insemination system which enables females to carry mature and unfertilized eggs, and packets of sperm inside her gonads (Ferraris Jr. 2003; Mazzoldi et al. 2007). In general, auchenipterid fishes are nocturnal and can be seen swimming just below the water surface in search of food items that fall from the adjacent forest (Ferraris Jr. 2003; Freitas et al. 2011). However, diurnal species and additional feeding habits (e.g. piscivorous, frugivorous, and planktonic) have been described for a handful of Auchenipteridae taxa (Ferraris Jr. 2003; Pouilly et al. 2004; Corrêa et al. 2009). Besides, the Auchenipteridae has been subject to several phylogenetic investigations in the recent decades (Britski 1972; Ferraris Jr. 1988; Curran 1989; Royero 1999; Akama 2004; Ribeiro 2010; Birindelli 2014; Calegari et al. 2019).

Although there is a significant amount of field information on the auchenipterid fishes, these data are plagued with taxonomic inaccuracies (Freitas et al. 2020 *in press*), and their representativeness has not been yet explored. Due to this, we conduct a joint assessment of the Linnean and Wallacean shortfalls, as well as on particular aspects of the Darwinian and Raunkiæran shortfalls. More specifically, we seek to evaluate the accumulated knowledge on species richness, coverage on the geographical distribution of species, the stability of phylogenetic hypotheses, and the accumulation of knowledge on feeding habitats. This study will provide a sound basis for future research in this group, and may also provide a general model for the assessment of knowledge advances in Neotropical Siluriformes and/or other taxa.

3.3. MATERIALS AND METHODS

3.3.1. Linnean shortfall

The current status of the Linnean shortfall on Auchenipteridae was assessed by analyzing the historical accumulation curve of valid species names. All names were checked for the most current nomenclature, according to Calegari et al. (2019) (see also Freitas et al. 2020 *in press*). This information was then used to estimate how many Auchenipteridae species remain to be described. To minimize the effect of changing rates of species description over time, we focused our analysis on the number of species described between the years 1974 and 2018, a period when species description follows a relatively homogeneous process (see Results section), and represents the recent advance in the taxonomy of the group. The year 1974 was chosen as the starting point because it contains the last abrupt leap in the description of Auchenipteridae species when nine species were described at once by Mees (1974). At this moment 79 species are considered as valid. To assess the increment in the description of species for the before-mentioned period, we used a piecewise regression (Zar 2010) to describe separately the temporal trends before and after any eventual breakpoint in the description rates (also applied by Ota et al., 2015). This kind of breakpoints in historical species accumulation curves can be associated with historical facts that may be related to the change in the rate of species description (such as the creation of scientific journals, investments in taxonomy and field expeditions; see Hortal et al., 2008). Further, we extrapolated the number of species that may remain to be discovered according to the description rates in the period 1974–2018. We did so through the second-order estimator Jackknife2 (hereafter, Jack2), using the number of species described per year as samples. We performed this analysis using the function *specpool* function of Vegan package version 2.5-5 (Oksanen et al. 2019) in R environment (R Core Team, 2015). We continued by determining the year in which the predicted species richness is likely to be achieved. To do this, we applied a density-dependent model through the formula (Gotelli 2008), $N_t = K/[1 + (K/N_0 - 1)e^{-rt}$, where K is the carrying capacity (here computed as the sum of Jack2 plus the number of species described before 1974), N is the number of described species at a time (t), N_0 is the total number of valid species, and r is a theoretical constant growth rate. To consider different scenarios of logistic growth models we used five theoretical constant growth rates (r = 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5) (see Ota et al., 2015).

3.3.2. Wallacean shortfall

To address the Wallacean shortfall, we first compiled a dataset containing 10,421 records of auchenipterids catfishes collected over ca. 180 years (from 1829 to December 2016). We retrieved these records from two online repositories: SpeciesLink (4,983 records), Global Biodiversity Information Facility (GBIF) (4,165 records) and included only preserved specimens and published observations of fishes from the Auchenipteridae (see Freitas et al. 2020 *in press* for further details). Besides, we included ichthyological records from the following Brazilian institutions, which were not available on online repositories: Museu Paraense Emílio Goeldi, Belém - Pará (1,108 records), Laboratório de Ictiologia de Altamira, Altamira - Pará (80 records), Universidade Federal da Bahia, Salvador - Bahia (44 records), Universidade Federal do Rio Grande do Norte, Natal - Rio Grande do Norte (35 records), and Universidade Federal da Paraíba, João Pessoa - Paraíba (six records). We obtained the records with the authorization of the respective collection curators.

All records were screened following a three-stage data filtering process. First, all duplicated records, i.e. records holding identical information on the fields “scientific collection” and “catalog number”. Second, we selected records with reliable taxonomic identification at the species level (see the procedure in Freitas et al. 2020 *in press*). For this step, we checked all taxa names in Calegari et al. (2019) for the most current nomenclature. Taxa names holding incorrect nomenclature and those that we were unable to correct were excluded. Third, we selected only records with information of latitude and longitude, and complete date of collection (i.e. day/month/year). The whole dataset is available in Supplementary Material S1. After data filtering, we used all records to calculate the species accumulation curves (SACs) and used the final slopes of these curves as a proxy for inventory completeness (Hortal et al. 2004, 2008, 2011; Yang et al. 2013). We calculated SACs for two sets of geographical features: Freshwater Ecoregions of the World (Abell et al. 2008); and geographic grid cells of one-degree resolution. We calculated the SACs and obtained the slopes only for ecoregions that contained at least 50 records, and grid cells with at least 20 records. All ecoregions and cells with 50 and 20 records or more, respectively, and slopes lower than or equal to 0.01 were considered well-sampled. In addition, ecoregions and cells with 50 and 20 records or more, respectively, and slope value between 0.01 and 0.05 were "reasonably sampled". Here, a slope of 0.01 means that for every 100 records added to the ecoregion or grid cell, one new species would be registered, while a slope equals 0.05 means the addition of 20 records to the sample, one new species would be recorded. Such reference values describe the relationship between sampling effort and species inventorying (Hortal et al. 2004; Yang et al. 2013).

Finally, we established historical species range accumulation curves (see Lobo et al., 2007) to explore the temporal evolution of sampling effort for the most recorded species. These curves assess the accumulation of occurrence records in new grid cells through time for the following species: *Trachelyopterus galeatus* (712 records), *Auchenipterichthys longimanus* (335 records), *Ageneiosus ucayalensis* (334 records), *Ageneiosus inermis* (312 records), *Centromochlus heckelii* (294 records), and *Auchenipterus nuchalis* (271 records).

3.3.3. Darwinian shortfall

The Darwinian shortfall was assessed by evaluating the temporal variations in the similarity of all available phylogenetic hypotheses. We do this, under the assumption that a higher degree of convergence (i.e. higher similarity) between phylogenies indicates a well-established set of evolutionary relationships among different lineages within the family. Increasing dissimilarity would, on the contrary, indicate incongruent phylogenetic patterns, and therefore an unstable (i.e. uncertain) knowledge on the evolutionary patterns of the group. To this end, we compiled all phylogenies available for Auchenipteridae (Britski 1972; Ferraris Jr. 1988; Curran 1989; Royero 1999; Akama 2004; Ribeiro 2010; Birindelli 2014; Calegari et al. 2019). These phylogenetic hypotheses involve different types of input data (e.g. morphological or molecular data), phylogenetic reconstruction methods (e. g. gradist, cladist, molecular), and coverage of taxa within the family.

We first obtained the phylogenies through their illustrations using the software TreeSnatcher Plus (Laubach et al. 2012). Then, we computed the similarities between each pair of phylogenies as the correlation between the patristic distances of common taxa present in both phylogenies. Given that branch length in a phylogenetic tree varies between phylogenies depending on the criteria and class of data applied, we first converted all branch lengths to one (i.e. patristic distances were based on the number of phylogenetic nodes separating two tips). The phylogenies include different taxa, but we only focused on patristic distances between genera. Notwithstanding, the number of common taxa between phylogenies varies, which might affect the correlation coefficients between topologies. To account for this potentially confounding effect, a resampling procedure was conducted, in which the lowest number of common genera observed for any phylogeny pair (n = 12) was used as a baseline. Then, for a phylogeny pair, ten shared genera were randomly selected to calculate the correlation between phylogenetic distances. This procedure was repeated ten thousand times per phylogeny pair, and we calculated the mean correlation coefficient. This approach yields a similarity matrix between phylogenies based on the topological congruence among shared genera. This matrix was used to classify the phylogenies according to a UPGMA analysis. We expect a higher degree of congruency within the most recent phylogenies. We computed all analyses in the *R* environment (R Core Team 2015), using the “ape” package (Paradis and Schliep 2019). All phylogenetic trees are available in Supplementary Material S2.

3.3.4. Raunkiærian shortfall

We assessed the quality and completeness of current knowledge about the trophic ecology of Auchenipteridae fishes. To do this, we searched scientific literature about the description of feeding habits in three online databases: ISI Web of Science (www.webofknowledge.com), Scielo (www.scielo.org), and Google Scholar (www.scholar.google.com). This survey considered only scientific literature published before December 2017. For this search, we used the keywords “diet”, “feeding”, “feeding ecology”, and “trophic ecology”, in combination with “Auchenipteridae” or the names of genera (e.g., "*Auchenipterus* AND feeding"). To cover the wide range of specialized scientific journals of international and national relevance, keywords were searched in Portuguese, English, Spanish, and French.

All publications with any information on feeding habits of auchenipterid species were then classified into eleven categories according to the dietary niche: aquatic insects, terrestrial insects, crustaceans, other invertebrates, fish, other vertebrates, fruits and seeds, vegetal matter, zooplankton, algae, and detritus. If a species *A* consumed two different food items (e.g. terrestrial insects and crustaceans), then we added two entries with feeding habitats for that species: “species *A* x terrestrial insects”, and “species *A* x crustaceans”, no matter whether these records were in two different publications or just one. The extent of the Raunkiærian shortfall in trophic habits was then evaluated by establishing a trophic habit accumulation curve using the combination of different feeding habits per species as a function of the cumulative number of published papers. This aimed to verify if new feeding habits are documented as new studies are published. In case there is a reasonable knowledge of feeding habits for the Auchenipteridae fishes, we expect that an asymptotic rarefaction curve would be obtained. We performed this analysis in EstimateS software version 9.1 (Colwell 2013). Finally, a diagram summarizing the knowledge shortfalls is provided.

3.4. RESULTS

3.4.1. Linnean shortfall

Within the time series 1766-2018, 123 currently valid species names were assigned for the Auchenipteridae (Fig. 1a). The first species was described in 1766, and until 1974, one new species was described every three years (0.4 species per year). Such a rate of species description contrasts with that from 1974 to 2018, with 1.2 new species per year on average (i.e., a 4-fold increase).

The piecewise regression conducted on species descriptions after 1974 estimated a breakpoint around the year 2000, when 93 species had already been described (Fig. 1b). The end of the historical species accumulation curve was therefore described by two quadratic polynomial functions in both intervals (“1974–2000”: 24 species described; $y = 0.009x2 - 35.31x + 34693$, $R^2 = 0.972$; and “2001–2018”: 29 species described; $y = 0.0795x2 - 317.84x + 317628$, $R^2 = 0.9918$). Considering only the period between 1974 and 2018 (53 species described), the estimator Jack2 predicted a total of 152 species for the Auchenipteridae family, which added to the 70 species discovered before 1974, renders a predicted richness of 222 species for the family. According to this estimation, 99 new species of Auchenipteridae fishes remain to be described (~ 45% of the total). The scenarios of future species description using the five theoretical constant growth rates forecast that all species in the family will be described by the years 2170 ($r = 0.05$), 2100 ($r = 0.1$), 2059 ($r = 0.2$), 2046 ($r = 0.3$), 2039 ($r = 0.4$) and 2035 ($r = 0.5$) (Fig. 1c).

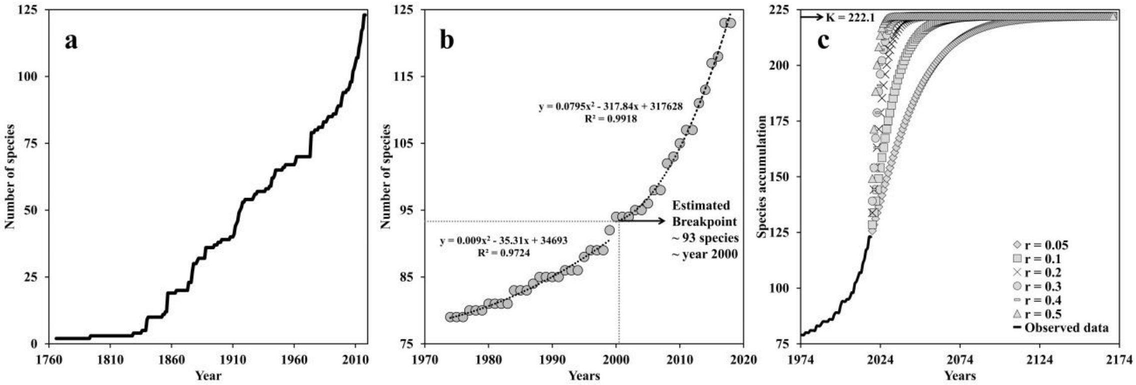


Figure 1. Linnean shortfall assessment for the Auchenipteridae family. (a) Historical accumulation curve of species description. (b) Piecewise regression on species description rates for the period 1974–2018, indicating the estimated breakpoint. (c) Predictions of the accumulated number of species described for five different scenarios of species description using theoretical constant growth rates (r).

3.4.2. Wallacean shortfall

Only 5,467 (52.5%) out of 10,421 records (and 469 entities) resulting in 117 valid species were deemed suitable for our analyses. Auchenipterid catfishes were recorded in 42 freshwater ecoregions in the Neotropics (which is composed by 50 ecoregions in total). The number of records per ecoregion ranges from 1 to 1,170 (median of 66.5 records; Table 1). Species richness ranged from 1 to 55 per ecoregion, and 12 species on average (with a median of six species). The richest ecoregion was Amazonas Lowlands with 1,170 records of 55 species, followed by Rio Negro (372 records for 43 species), Amazonas Estuary & Coastal Drainage (600 records for 36 species), and Orinoco Llanos (243 records for 34 species). We observed the lowest auchenipterid richness in the Amazonas High Andes (*Duringlanis perugiae*), Santa Maria (*Trachelyopterus amblops*), and Fluminense (*Trachelyopterus striatulus*) ecoregions, with one species each.

The slope of species accumulation curves were calculated for the 24 ecoregions that contained at least 50 records (57% of all ecoregions). Slope values ranged from less than 0.01 to over 0.99 (Table 1). SAC slopes were lower than 0.01 for seven ecoregions (i.e. 16.7% of the total): Amazonas Estuary & Coastal Drainages, Amazonas Lowlands, Guianas, Lower Uruguay, Paraguay, Upper Parana, and Xingu. These ecoregions were considered as well-sampled. Another 12 ecoregions (28.6% of ecoregions with occurrence data) can be considered reasonably sampled (slopes between 0.01 and 0.05); and five as poorly sampled (slopes larger than 0.05) (Table 1, Fig. 2).

We found that 1,353 cells out of 1,968 (68.8% of all grid cells) did not have a single record of auchenipterid fishes, whereas 615 cells (31.2% of all grid cells) contained at least one record. The number of records and species per cell ranged from 1 to 197 records and 1 to 28 species. When considering cells with at least one record, we found a median of four records and two species per cell. The richest cell showed 97 records for 28 species, whereas 231 cells contained only one species (records varying from 1 to 31). Seventy cells held at least 20 records, which represented only 11.4% of all cells with records. The final slopes of the species accumulation curves in these cells ranged from less than 0.01 to 0.42. From these, only three cells (0.5% of cells with records) can be considered as well-sampled (slope ≤ 0.01), and 18 (2.4% of cells with occurrence) were reasonably sampled (slope > 0.01 to ≤ 0.05) (Fig. 2).

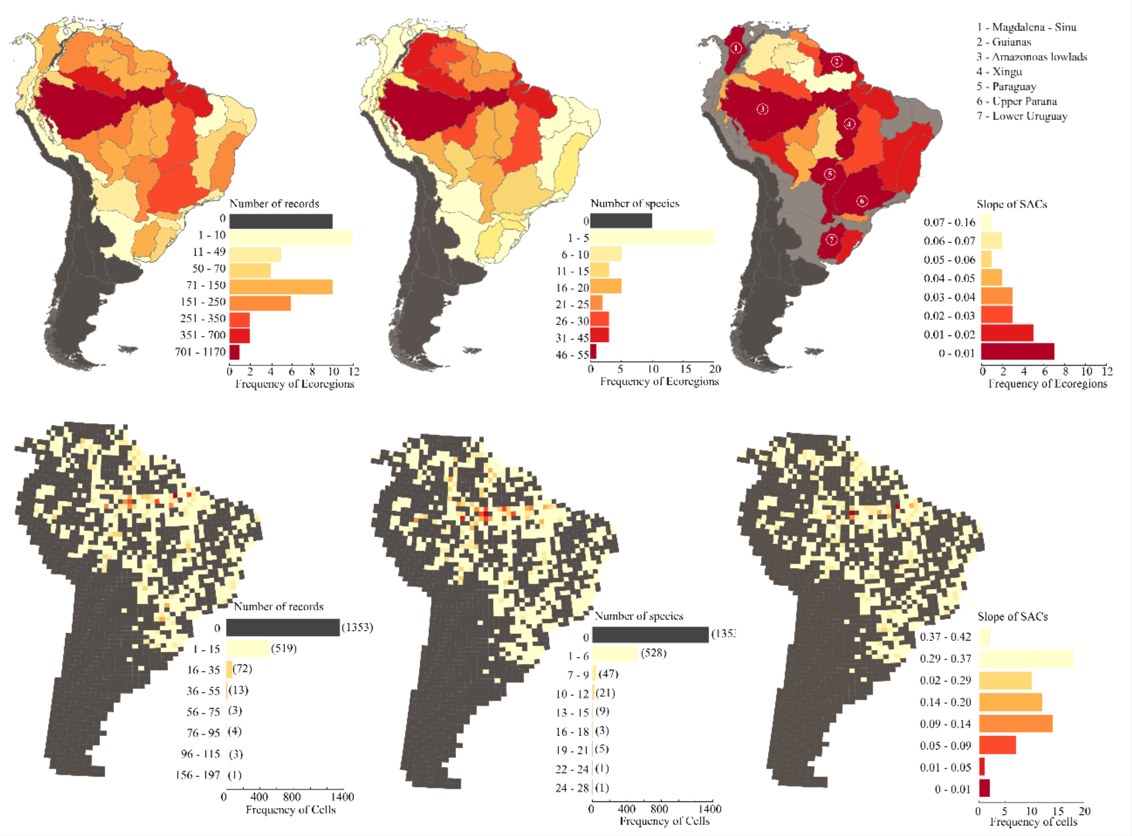


Figure 2. Inventory completeness for the Auchenipteridae across South America. Maps depict the number of records, species observed, and final slope of the species accumulation curves (SACs) for freshwater ecoregions (upper row) and one-degree resolution grid cells (lower row). Well-sampled ecoregions are numerically identified.

Table 1. Number of species occurrence records, species richness and final slope of the species accumulation curves for the 42 ecoregions with occurrence records of Auchenipteridae. Well-sampled ecoregions are indicated in bold. We also indicate the number of grid cells that are either well- or reasonably-sampled in each ecoregion (slopes ≤ 0.01 or >0.01 and ≤ 0.05, respectively).

Ecoregion	**Records**	**Species**	**Slope**	**Well-sampled cells**	**Reasonably sampled cells**
Amazonas Lowlands	**1170**	**55**	**< 0.01**	**-**	**2**
Amazonas Estuary & Coastal Drainages	**630**	**36**	**0.01**	**1**	**1**
Rio Negro	372	43	0.02	-	-
Tocantins-Araguaia	310	26	0.03	-	2
Upper Parana	**284**	**12**	**< 0.01**	**-**	**2**
Orinoco Llanos	243	34	0.06	1	2
Paraguay	**217**	**16**	**< 0.01**	**-**	**1**
Madeira Brazilian Shield	208	26	0.03	-	2
Essequibo	196	23	0.02	-	-
Northeastern Mata Atlantica	186	9	0.02	-	-
Mamore-Madre de Dios Piedmont	181	20	0.02	-	-
Xingu	**135**	**16**	**0.01**	**-**	**1**
Tapajos-Juruena	123	17	0.06	1	-
Iguassu	117	6	0.03	-	-
Lower Uruguay	**115**	**8**	**0.01**	**-**	**2**
Orinoco Delta & Coastal Drainages	112	15	0.04	-	1
Orinoco Guiana Shield	111	26	0.06	-	-
Guianas	**106**	**16**	**0.01**	**-**	**-**
Amazonas Guiana Shield	89	22	0.11	-	-
Magdalena - Sinu	73	3	> 0.99	-	-
Guapore-Itenez	71	15	0.04	-	1
Laguna dos Patos	62	4	0.02	-	-
Western Amazon Piedmont	61	9	0.05	-	1
São Francisco	56	5	0.02	-	-
Northeastern Caatinga & Coastal Drainages	48	3	-	-	-
Upper Uruguay	37	6	-	-	-
Paraiba do Sul	35	3	-	-	-
North Andean Pacific Slopes-Rio Atrato	23	3	-	-	-
Ribeira de Iguape	19	3	-	-	-
Chaco	13	5	-	-	-
Lower Parana	10	4	-	-	-
Parnaiba	9	2	-	-	-
Fluminense	8	1	-	-	-
Tramandai-Mampituba	7	3	-	-	-
South America Caribbean Drainages-Trinidad	6	3	-	-	-
Orinoco Piedmont	5	5	-	-	-
Maracaibo	5	3	-	-	-
Rio Tuira	5	2	-	-	-
Southeastern Mata Atlantica	5	2	-	-	-
Ucayali-Urubamba Piedmont	2	2	-	-	-
Amazonas High Andes	1	1	-	-	-
Santa Maria	1	1	-	-	-

The species *Trachelyopterus galeatus* showed the broadest geographic distribution, being recorded in 245 grid cells (39.8% of the cells with records), followed by *Ageneiosus inermis* (121 cells, 19.7%), *A. ucayalensis* (112 cells, 18.2%), *Auchenipterus nuchalis* (97 cells, 15.8%), *Centromochlus heckelii* (93 cells, 15.1%), and *Auchenipterichthys longimanus* (64 cells, 10.4%). Forty-one species were recorded in five or less grid cells, so one-third of the auchenipterid diversity has its known distribution restricted to very few areas. After over 180 years of sampling (1829–2016), occurrences of auchenipterid fishes are still accumulating as the sampling of new areas (grid cells) expands. That is, the known geographic distribution of auchenipterid fishes is still expanding, at least for the most conspicuous species (Fig. 3).

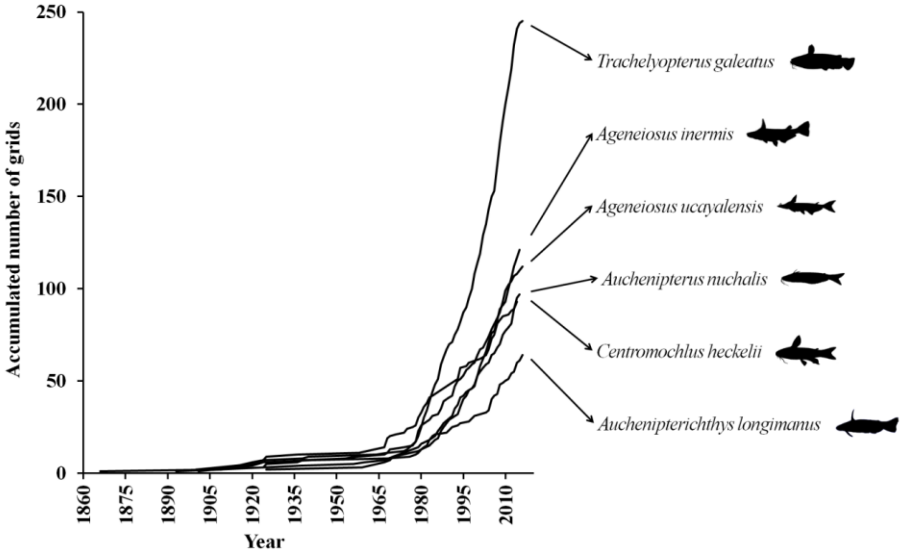


Figure 3. Temporal accumulation of the number of grid cells with occurrence records for the six most recorded Auchenipteridae species (period 1829-2016).

3.4.3. Darwinian shortfall

The analysis of dissimilarity between phylogenies showed that the topologies of Birindelli (2014) and Ribeiro (2010) were the least dissimilar (23.1% of dissimilarity), followed by the topologies of Akama (2004) and Calegari et al. (2019) (28.1% dissimilarity) (Fig. 4). Taking into account the current number of valid species, phylogenetic studies covered, on average, 46% of all valid species. Birindelli (2014) used the lowest proportion of species and the number of valid species (20%, or 23 species used out of 113). In the most current phylogenetic study (Calegari et al., 2019), 97 species out of the 124 valid names were used, representing 77% of all currently valid species.

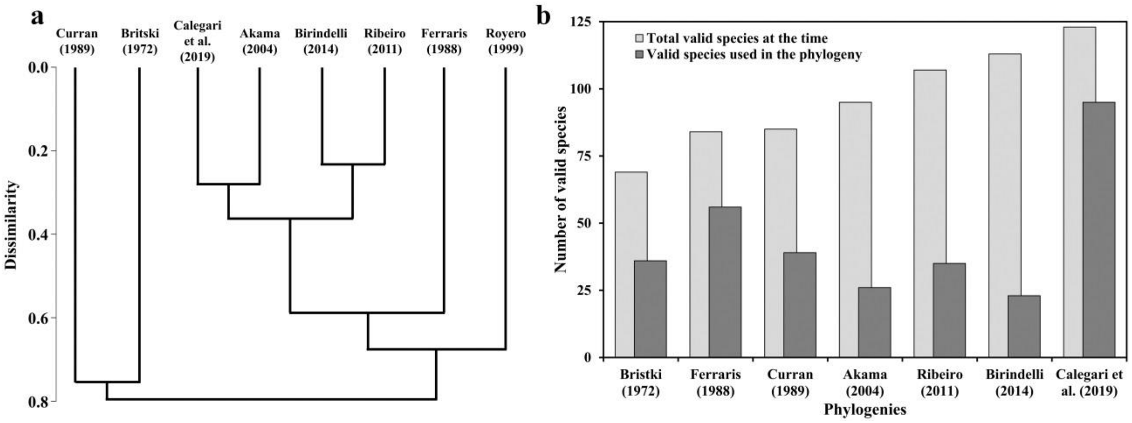


Figure 4. Dendrogram of dissimilarity between phylogenetic topologies of the family Auchenipteridae (a). The number of used species in the phylogenies and number of valid species at the time of reconstruction (b).

3.4.4. Raunkiærian shortfall

We compiled 67 published studies that had cited the feeding habits of 54 auchenipterid species. Among them, we registered 244 combinations of diet and species, out of which 142 unique combinations of diet and species. The final slope of the feeding habit accumulation curve shows that the dietary knowledge on Auchenipteridae catfishes is still expanding, at a nearly constant pace of 1.5 unique combinations of feeding habits and species per published study (Fig. 5a). The species with best-documented feeding habits was *Trachelyopterus galeatus* (40 feeding combinations, ten of them unique, described in 17 published papers), followed by *Ageneiosus ucayalensis* (13 combinations, four unique, in six papers), *Trachelyopterus striatulus* (14 combinations, five unique, in seven papers), and *Auchenipterus nuchalis* (13 combinations, four unique, in nine papers). For *Entomocorus gameroi*, we obtained six feeding habits and six unique combinations. From the 54 species with data on feeding habits, 14 species showed only a single record/single combination.

The distribution of species studied among genera shows that although the most speciose genera hold data for more species, there are some significant biases. While the two genera with more species (*Tatia* and *Trachelyopterus*) have data for around one-fourth of their species, the following ones are much better represented (in particular *Ageneiousus*) (Fig. 5b). Further, five genera had records of feeding habits for all their valid species (*Asterophysus*, *Auchenipterichthys*, *Epapterus*, *Tocantinsia*, and *Trachycorystes*), while seven genera did not have any feeding habit records documented for theirs (*Gephyromochlus*, *Liosomadoras*, *Pseudepapterus*, *Pseudotatia*, *Spinipterus*, *Trachelyichthys*, and *Trachelyopterichthys*). The complete list of feeding habits per species and the consulted literature is available in Supplementary Material S3. Lastly, an overall picture of the results, showing the overlap among the studied knowledge shortfalls is shown in Fig. 6 (see Supplementary Material S4 for the complete list of species and their data availability).

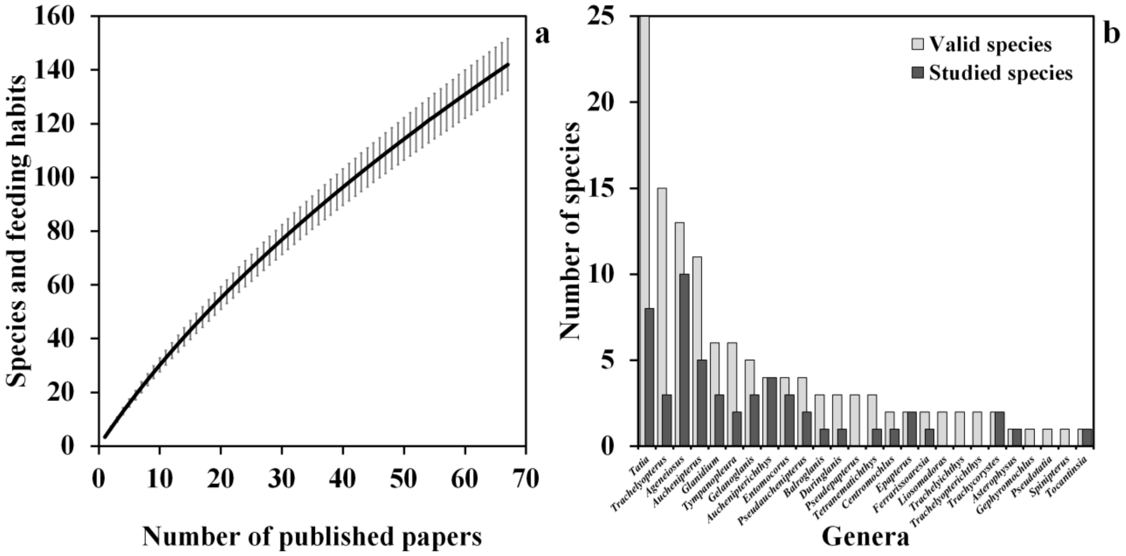


Figure 5. Current status on the knowledge of feeding habits for Auchenipteridae species. (a) Accumulation of observed feeding habits per species with an increasing number of publications. (b) Distribution of species with feeding habits among the currently valid genera, compared with the number of species in each one of these genera.

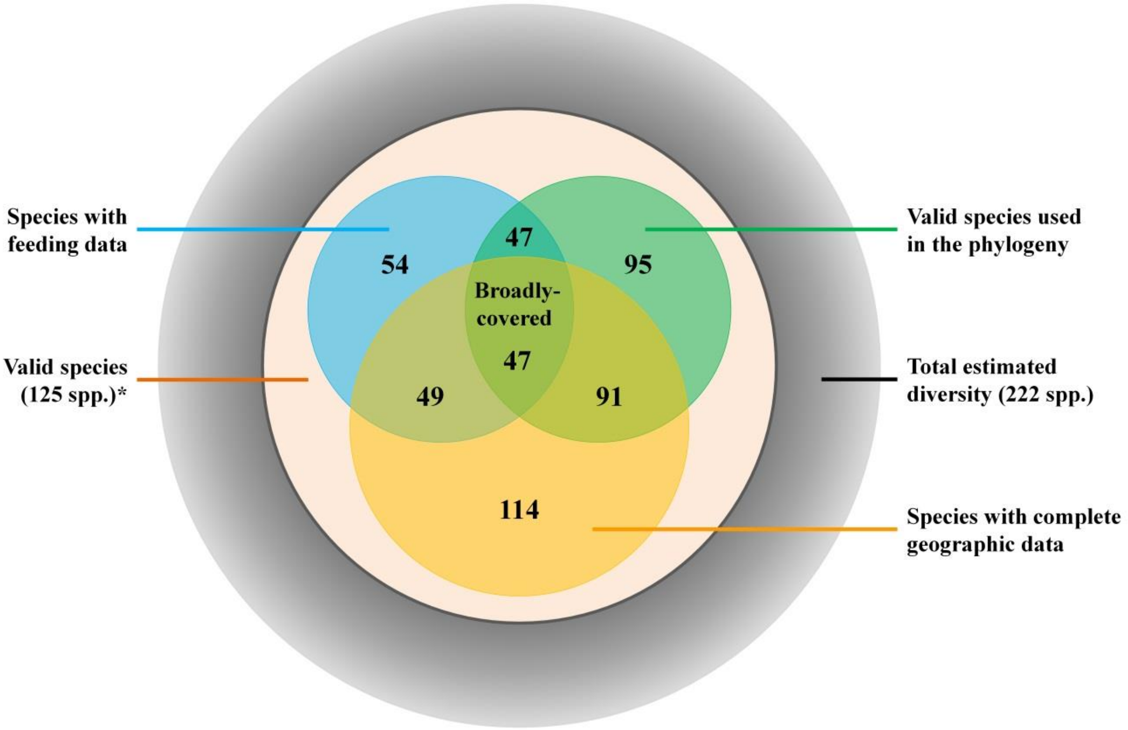


Figure 6. Shortfalls in the knowledge of the Auchenipteridae catfish family (order Siluriformes) at the species level. The Linnean shortfall (i.e. undescribed species) surrounding our current knowledge is represented in gray. The area outside each circle represents the species in lack of knowledge. Broadly covered species (shown in the overlap area) are those for which there is data on their distribution, phylogenetic relationships (from Calegari et al., 2019), and feeding habits.

3.5. DISCUSSION

The knowledge of taxonomy, geographic distribution, evolutionary relationships, and feeding habits of Auchenipteridae catfishes has increased in recent years. Our results indicate that the historical accumulation of species description showed a steep increase between the years 1974 and 2018, thereby suggesting that a sound knowledge about the total richness of the family is still not on the horizon. This implies that an increased effort in inventorying and describing auchenipterid catfishes along South American freshwater environments is needed in order to attain a fair knowledge of the overall diversity of this group. Despite such undersampling, the large similarities between the most recent phylogenies suggest that current evolutionary hypotheses for this family are fairly robust, at least in their most basal topology. This is not the case for the knowledge about the feeding habits and diet of auchenipterids, which is still in a phase of expansion. Our results highlight the need for further efforts in describing species and/or developing taxonomical reviews, field explorations and ecological investigations for this diverse group of fishes. Below we discuss the impact of each one of the studied shortfalls and pinpoint historical events that contributed to such temporal dynamics.

3.5.1. Linnean shortfall

Increasing taxonomic efforts typically lead to increments in species description rates, rather than to leveling-off the accumulation of new taxa. As a rule, the more taxonomy improves, the larger are the numbers of descriptions and estimated species richness (Troudet et al. 2017). This is the case of Auchenipteridae: our results show a remarkable increase in the description of auchenipterid species in the last decades. Indeed, a substantial proportion of taxa may remain to be described to science (almost 50% according to our estimations), corroborating that Auchenipteridae is a poorly sampled family (Ota et al. 2015).

The taxonomic history of auchenipterid fishes started in 1766 with the description of *Silurus inermis* and *S. galeatus* (currently valid as *Ageneiosus inermis* and *Trachelyopterus galeatus*, respectively) by Carolus Linnaeus in the twelfth edition of "*Systema Nature*" (see Calegari et al., 2019). But the first real improvement in the description of auchenipterid species was promoted by Achille Valenciennes in 1840, a French zoologist, who explored South America in the early nineteenth century and described species of *Ageneiosus*, *Auchenipterus*, and *Trachelyopterus*. In the late nineteenth century, the auchenipterid diversity increased substantially due to the studies of Rudolf Kner (1857) and Franz Steindachner (between 1881 and 1915), who described species that now are allocated in several genera within the family (see Calegari et al., 2019). After a period of intense taxonomical contribution by Carl H. Eigenmann and Alípio de Miranda Ribeiro in the early twentieth century, Gerloff Mees published in 1974 one of the most important monographs on auchenipterid systematics (our *breakpoint*), the book "*The Auchenipteridae and Pimelodidae of Suriname (Pisces, Nematognathi)*", where he described several new species of *Centromochlus* and *Tatia*, as well as the genera *Pseudotatia*, *Tocantinsia*, and *Trachelyichthys* (Mees 1974). This monograph set up the basis for the recent increase in taxonomic effort (1980-2017), thanks to the work of dozens of researchers around the world, especially by the North American Carl J. Ferraris Jr., Richard Vari and Steven Walsh, and the Brazilians José Birindelli, Luiza M. Sarmento-Soares, Frank Ribeiro and Bárbara B. Calegari. This high rate of description continues to date, with the recent description of *Tatia caudosignata* DoNascimiento, Albornoz-Garzón & García-Melo 2019 and *Spinipterus moijiri* Rocha, Rossoni, Akama & Zuanon 2019, which have not been included in our analyses.

The Linnean shortfall is made of unknown species from two distinct categories (Hortal et al. 2015): species yet to be sampled and those species that have been sampled but have not been described yet. The latter condition can be represented by either species already known as a new form by the specialist but not yet formally described, as well as unidentified specimens deposited in scientific collections, or particular cases of misidentification. The taxonomy of groups with many species in these situations often benefits from beta taxonomy (i.e. revisionary work). In Auchenipteridae, the revisionary studies published in the last decade had a great contribution to the description of new species. Around 40% of auchenipterid species described between 2001 and 2018 resulted from taxonomic revisions (e.g. Reis and Borges 2006; Sarmento-Soares and Martins-Pinheiro 2008; Ribeiro et al. 2017). In this sense, revisionary studies are responsible for re-identifications of material deposited in the collections including older and recent specimens sampled, also contributing to updating the database collection and the geographical distribution of species (Lundberg et al. 2000; von Staden et al. 2013).

Conversely, the taxonomic gap represented by the not-yet-sampled species depends on inventory surveys (mainly in poorly known regions), which are of paramount importance for registering species new to science (Wheeler 2004; Mora et al. 2008; Hortal et al. 2015). In the case of Auchenipteridae, 60% of the species described in the last decade resulted from field inventories (e.g. Akama and Ferraris 2011; Sarmento-Soares et al. 2013, 2016). The high number of species recently discovered in the field may be related to the inconspicuousness of Auchenipteridae catfishes. The development of new collecting techniques, new technologies, and exploration of new habitats can lead to large increases in the species description for a group of organisms (see Winston 1999). For instance, the scarcity of studies using molecular techniques for species delimitation in Auchenipteridae (except by Cooke et al. 2012; Calegari et al. 2019; Hashimoto et al. 2020) is noteworthy. The taxonomic knowledge of many Neotropical fish family is going all out with species delimitation by the molecular approach (Characidae - García-Melo et al. 2019; Curimatidae - Melo and Oliveira 2017; Loricariidae - Roxo et al. 2015).

We also highlight the importance to be cognizant of intrinsic features of the natural history of species to better understand what is contributing to the knowledge gaps. Auchenipteridae (popularly known as driftwood catfishes) is a family of catfishes known by its distinctive habit of using submerged trunks as diurnal shelters (in genera such as *Balroglanis, Liosomadoras, Tatia,* and *Trachelyichthys*) (Birindelli and Zuanon 2012), or by foraging at twilight and during night on the surface of the water (such as the miniature species of *Gelanoglanis*) (Calegari and Reis 2017). Because of these peculiarities, collecting fish from this family requires using diverse methods such as manual collecting inside trunks or ichthyoplankton nets.

3.5.2. Wallacean shortfall

The analysis of inventory completeness reveals that the geographical distribution of Auchenipteridae along South American freshwaters is largely under-sampled. Only 17% of the ecoregions with recorded presence of auchenipterid catfishes are well sampled for the family, and 29% are reasonably sampled according to our criteria. That is, more than half of the ecoregions are largely under-collected for those catfishes. This pattern is even more striking when looking at the inventories of one-degree grid cells: only three and 18 cells were either well or reasonably sampled, respectively. Such pattern of striking changes in inventory completeness with diminishing scale is typical of most biodiversity inventories (Lobo et al., 2018), as intensive local inventories are associated with specific research projects, long-term ecological studies, historically relevant localities repeatedly visited by taxonomists (Hortal et al. 2008; Sastre and Lobo 2009) and/or the proximity to research centers and taxonomists' home ranges (Dennis and Thomas 2000; Lobo and Martin-Piera 2002; Rodrigues et al. 2010). The latter is the case of the three anecdotally well-sampled grid cells, which are placed near two well-established scientific research centers, the *Instituto Nacional de Pesquisas da Amazonia* (INPA) and the *Museu Paraense Emilio Goeldi* (MPEG), in the Amazonian cities of Manaus and Belém (Brazil) respectively. Indeed, biodiversity occurrence-data are historically biased towards locations that offer easier logistics, such as roads, and nearby cities or villages (Hortal et al. 2007; Lobo et al. 2007; Meyer et al. 2015). This is particularly important in areas with access difficulties as high as the Amazon, where large efforts are needed to study local biodiversity patterns (see Avila-Pires et al. 2009; Jézéquel et al. 2019).

The temporal evolution in the accumulated number of grid cells with occurrences registered for the most recorded species shows that the yearly increase in geographic coverage was generally low before 1980, increasing conspicuously after this date. Such steep increase in the number of records from the 80s and 90s has also been observed for many biological groups around the world (e.g. angiosperms, Canhos et al. 2015; invertebrates, Lobo et al. 2007; Hortal et al. 2008; Isaac and Pocock 2015; birds, Amano et al. 2016; small mammals, Escribano et al. 2016), and is generally associated with the development of universities, scientific facilities, and conservation efforts. Note, however, that similar increase in inventory rates increases appeared earlier for better-studied groups (e.g. amphibians, Rodrigues et al. 2010) or may even diminish in many areas after historical colonial inventories (Stropp et al. 2016). In Brazil, several factors can be associated with the increase in the records of fauna and flora during this period, such as the creation of protected areas, and improvement of the environmental impact studies (Pott and Estrela 2017). These studies required the collection of biological material, which was deposited in scientific collections. Besides, investment for field expeditions increased around the 60s, favoring the creation of important scientific fish collections, such as the INPA fish collection (Rapp Py-Daniel et al. 2015), and the expansion of other already established ichthyological collections, such as the collections of MPEG (Silva et al. 2017) and the Museum of Zoology of the Universidade de São Paulo (MZUSP) (Menezes et al. 1997; Marinho et al. 2019).

3.5.3. Darwinian shortfall

The knowledge of the evolution of Auchenipteridae has advanced considerably since the first phylogenetic studies on this family and now seems to be largely settled. Not surprising, the four most recent phylogenies showed the highest similarity among them since all of them included several morphological characters in their analyses, originally described by Ferraris Jr. (1988) and Royero (1999). Indeed, the phylogenetic trees of Akama (2004), Ribeiro (2010), Birindelli (2014), and Calegari et al. (2019) show a remarkable congruence in the composition and relationships within the subfamily Auchenipterinae (*sensu* Calegari et al., 2019). The large consistency in the most recent phylogenies contrasts with older phylogenies, which in general were hampered by a poorer sampling of this family, as the earlier ones (Britski, 1972; Ferraris Jr., 1988; Curran, 1989) sought to establish the relationships among genera accounting for only a few auchenipterid representatives. Some recent auchenipterid phylogenetic studies resulted from revisionary studies focused on a particular genus (Royero, 1999; Akama, 2004; Ribeiro, 2010), increasing the number of taxa sampled.

Despite the increase in phylogenetic studies, a full understanding of the evolutionary relationships of these driftwood catfishes persists. The most important basal differences between these evolutionary hypotheses come from composition of the clade "(*Auchenipterichthys* + (*Trachelyopterichthys* + *Trachelyichthys*)", which is considered monophyletic by Akama (2004) and Calegari et al. (2019), while Ribeiro (2010) and Birindelli (2014) also include *Trachycorystes* and *Trachelyopterus* within it. These discrepancies occur because the different hypotheses have been based on incomplete sampling biased towards several genera, using also largely different sets of species. This incongruity is the biggest obstacle for the fulfillment of phylogenetic gaps (Diniz-Filho et al. 2013; Hortal et al. 2015; Assis 2018) and calls for an even more comprehensive survey of the family. The recent species-level phylogeny of Calegari et al. (2019) is the first to include a large proportion of the species of the family (97 of 124 valid species). Also, it is the first phylogeny including molecular data under a total evidence approach, in contrast to all former hypotheses based on morphological data. The increasing availability of molecular data is indeed providing a better understanding of the evolutionary relationships among living beings (Moritz 1995; Diniz-Filho et al. 2013), also rendering important insights into the description of biodiversity (Sheth and Thaker 2017).

3.5.4. Raunkiærian shortfall

Knowledge of fish diet offers a consistent approach to assessing interactive processes within aquatic communities (Winemiller 1989), enabling predictions about the roles of particular trophic groups and fish species in the system (López-Rodríguez et al. 2019). Such ecological data is desirable for developing conservation strategies and is, therefore, a key element in the protection of species and ecosystems (Abelha et al. 2001; Braga et al. 2012). Examining the amount, patterns, and trends of the study of species traits allows us to identify knowledge gaps and to guide future research strategies (Guerra et al. 2018). Despite such key importance, current knowledge on feeding habits presents significant biases across auchenipterid genera and species. The first feeding habit of an auchenipterid (*Ageneiosus militaris*) was published at the *Anales del Museo de Historia Natural de Montevideo* by Devincenzi and Teague (1942). After our compilation, we registered that less than 50% of the known species has some published information on diet, and information on the diet of auchenipterid fishes is still accumulating with every new study published. Special attention should be drawn to the genera which were underrepresented in our survey (e.g. *Tatia*, *Trachelyopterus*), as well as towards those with no available information on feeding habits (*Gephyromochlus*, *Liosomadoras*, *Pseudepapterus*, *Pseudotatia*, *Spinipterus*, *Trachelyichthys*, and *Trachelyopterichthys*). In general, we lack comprehensive knowledge of trophic ecology of fishes (Braga et al. 2012), as well as other vertebrates such as amphibians (Solé and Rödder 2000). The great fish diversity and the limited standardization of the methods applied hamper to fill this gap in this group (Bennemann et al. 2006), compared to other groups such as primates (Hanya and Chapman 2013).

A noteworthy aspect observed in our results was the majority of combinations of diet and species were unique, that is, registered a single time in the literature. This type of information highlights the punctual nature of the records obtained in the literature and indicates the need for more comprehensive studies (spatially and temporally) in order to have a more reliable snapshot of the auchenipterid's feeding habits. Although we are dealing here with one ecological trait (feeding habit), this shortfall is certainly vast for other important features that are even more difficult to measure.

3.5.5. Interactions between shortfalls within auchenipterid fishes

One way or another, all shortfalls are intimately related to each other, but all of them are necessarily tied to the Linnean shortfall (Hortal et al., 2015). The basic alpha-taxonomic process of the Auchenipteridae species description is far from complete. Failing to distinguish species – the basic units of ecological and evolutionary studies – prevents from attributing any characteristic to them accurately, whether it is geographic distribution, phylogenetic relatedness or a functional trait. Filling in the Linnean shortfall with a formal description of species allows assigning at least one location to a taxon, thus beginning to fill in the Wallacean shortfall. For instance, species are only known from the type-locality, such as the auchenipterid *Pseudotatia parva*, known for the lower course of the Rio São Francisco (Northeast Brazil) (Mees, 1974). As new biological inventories are carried out, more likely to record a species for other locations, as the case of *Spinipterus acsi*, for which a new record extended its geographic distribution to almost 2,000 km by river from the type locality (Calegari et al. 2018). Notwithstanding, the uncertainties in the Darwinian shortfall in Auchenipteridae can be reduced with the advances in taxonomic research, but they will remain limited by the large extent of the Linnean shortfall, as new species that have been described or are still to be described will not be included in phylogenetic hypotheses. This generalized interaction between shortfalls hampers the development of knowledge on evolutionary patterns (Hortal et al., 2015; Assis, 2018).

Wallacean shortfalls can also have pervasive impacts on Raunkiæran shortfalls (Hortal et al., 2015). Indeed the lack of geographical coverage can also affect the knowledge about within-species variation of trophic habits, as the frequency of feeding on different sources can be highly influenced by where fishes are caught. For instance, *Tocantinsia piresi* showed carnivorous and frugivorous habits in two different Amazonian river basins (compare Mérona et al. 2001 with Dary et al. 2017). Indeed, species with larger distribution areas (e.g. *Trachelyopterus galeatus* and *Auchenipterus nuchalis*) showed larger numbers of feeding studies and feeding combinations (not shown). Besides, widely distributed species are commonly more generalist concerning their feeding habits (Abelha et al. 2001), and since they are easily found, they are also broadly studied. On the other hand, the feeding habits of auchenipterid species with restricted distributions (e.g. *Ferrarissoaresia meridionalis*) or recently described (e.g. *Ageneiosus apiaka*, *Spinipterus acsi*) have been seldom studied. Darwinian shortfalls can also contribute to the Raunkiæran shortfalls, as a novel phylogenetic arrangement can create – or rather make evident – new gaps on the knowledge of traits. The current topology of the Auchenipteridae phylogeny recovered the genus *Gephyromochlus* as separated from *Glanidium* (Calegari et al., 2019), so from now on, no available information of feeding habit is available for this new genus (while before such information could be extrapolated from that of other *Glanidium* species). In this sense, well-resolved phylogenetic hypotheses may be crucial to make inferences about the relationships between traits and ecological functions, and/or whether they are the result of other evolutionarily correlated features (Hortal et al. 2015; Thompson et al. 2015).

3.5.6. Concluding remarks

The compilation of biodiversity data at regional and global scales has shown large advances in recent years (Hortal et al. 2015; Cornwell et al. 2019). The growing availability of information and the advance in informatics have enabled large-scale analyzes and eased the interpretation of biodiversity data (Soberón and Peterson 2004; Maldonado et al. 2015). Many initiatives have provided data that is being used in ecological and evolutionary studies, such as the Global Biodiversity Information Facility (GBIF) or NIH genetic sequence database (GenBank) (Canhos et al. 2015). These global database projects are essential for reducing the extent of the biodiversity knowledge shortfalls (Hortal et al., 2015). On the one hand, they mobilize information previously hidden and/or inaccessible in grey literature. But on the other, they can serve as a way of identifying which are the biases in our current knowledge and where are the gaps in the data, allowing identify specific regions, taxa or clades in need of further work. But there is a limit to the value of currently available data; as our results show, in general, the diversity of most living groups is largely unknown due to limited taxonomic knowledge. Here, the continuous growth in the number of formally described species suggests promising scientific advances for the coming years. Such growth implies that there is still much work to be done on describing different aspects of biodiversity, for more species are expected to be described in the future, and most of the already described species lack reliable information regarding their geographical distribution, evolutionary relationships, functional traits, and other key ecological aspects. Filling these gaps is crucial to hold a reliable picture of the broad-scale biodiversity pattern and process (Hortal et al., 2015). Reducing these gaps will require a concerted effort of taxonomists, ecologists, biogeographers, using evaluations of knowledge shortfalls such as the one conducted here to guide field, experimental and lab work towards covering specific knowledge gaps.

ACKNOWLEDGEMENTS

The authors are grateful to the financial support provided by several grants from the Brazilian *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES - Finance Code 001; TMSF - 23038.042984/ 2008-30), and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq; PDM - 308694/2015-5; LFAM - 305017/2016-0; JH – PVE 314523/2014-6, JS - Post-Doctoral fellowship 434391/2016-6). LFAM also thanks to PROCAD (UFPA/UFG/UNESP: Ecologia, Genética e Conservação do Cerrado, Amazônia e Mata Atlântica).

3.6. REFERENCES

Abelha M.C.F., Agostinho A.A., Goulart F. 2001. Plasticidade trófica em peixes de água doce. Acta Sci. 23:425–434.

Abell R., Thieme M.L., Revenga C., Bryer M., Kottelat M., Bogutskaya N., Coad B., Mandrak N., Balderas S.C., Bussing W., Stiassny M.L.J., Skelton P., Allen G.R., Unmack P., Naseka A., Ng R., Sindorf N., Robertson J., Armijo E., Higgins J. V., Heibel T.J., Wikramanayake E., Olson D., López H.L., Reis R.E., Lundberg J.G., Sabaj Pérez M.H., Petry P. 2008. Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. Bioscience. 58:403–414.

Akama A. 2004. Revisão sistemática dos gêneros Parauchenipterus Bleeker, 1862 e Trachelyopterus Valenciennes, 1840 (Siluriformes, Auchenipteridae). Unpublished D. Phil. Dissertation, Universidade de São Paulo.

Akama A., Ferraris C.J. 2011. Spinipterus , a new genus of small, spiny catfish (Siluriformes: Auchenipteridae) from the Peruvian Amazon. Zootaxa. 2992:52–60.

Amano T., Lamming J.D.L., Sutherland W.J. 2016. Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. Bioscience. 66:393–400.

Assis L.C.S. 2018. Revisiting the Darwinian shortfall in biodiversity conservation. Biodivers. Conserv. 27:2859–2875.

Avila-Pires T.C.S., Hoogmoed M.S., Rocha W.A. da. 2009. Notes on the Vertebrates of northern Pará, Brazil: a forgotten part of the Guianan Region, I. Herpetofauna. Bol. do Mus. Para. Emílio Goeldi Ciências Nat. 5:13–112.

Barber P.H., Ablan-Lagman M.C.A., Berlinck R.G., Cahyani D., Crandall E.D., Ravago-Gotanco R., Juinio-Meñez M.A., Mahardika I.N., Shanker K., Starger C.J., Toha A.H.A., Anggoro A.W., Willette D.A. 2014. Advancing biodiversity research in developing countries: the need for changing paradigms. Bull. Mar. Sci. 90:187–210.

Bennemann S.T., Casatti L., Oliveira D.C. de. 2006. Alimentação de peixes: proposta para análise de itens registrados em conteúdos gástricos. Biota Neotrop. 6.

Birindelli J.L.O. 2014. Phylogenetic relationships of the South American Doradoidea (Ostariophysi: Siluriformes). Neotrop. Ichthyol. 12:451–564.

Birindelli J.L.O., Zuanon J. 2012. Systematics of the Jaguar catfish genus Liosomadoras Fowler, 1940 (Auchenipteridae: Siluriformes). Neotrop. Ichthyol. 10:1–11.

Braga R.R., Bornatowski H., Vitule J.R.S. 2012. Feeding ecology of fishes: an overview of worldwide publications. Rev. Fish Biol. Fish. 22:915–929.

Britski H.A. 1972. Sistemática e evolução dos Auchenipteridae e Ageneiosidae (Teleostei, Siluriformes). Unpublished D. Phil. Thesis, Universidade de São Paulo.

Calegari B.B., Akama A., Reis R.E. 2018. First record of the driftwood catfish Spinipterus acsi Akama &amp; Ferraris, 2011 (Siluriformes, Auchenipteridae) for Brazil, Juru&aacute; River, Amazon basin. Check List. 14:693–697.

Calegari B.B., Reis R.E. 2017. New species of the miniature genus Gelanoglanis

(Siluriformes: Auchenipteridae) from the Tocantins River basin (Brazil) and osteological description of G. nanonocticolus. J. Fish Biol. 90:1702–1716.

Calegari B.B., Vari R.P., Reis R.E. 2019. Phylogenetic systematics of the driftwood catfishes (Siluriformes: Auchenipteridae): a combined morphological and molecular analysis. Zool. J. Linn. Soc. 187:661–773.

Caley M.J., Fisher R., Mengersen K. 2014. Global species richness estimates have not converged. Trends Ecol. Evol. 29:187–188.

Canhos D.A.L., Sousa-Baena M.S., de Souza S., Maia L.C., Stehmann J.R., Canhos V.P., De Giovanni R., Bonacelli M.B.M., Los W., Peterson A.T. 2015. The Importance of Biodiversity E-infrastructures for Megadiverse Countries. PLOS Biol. 13:e1002204.

Cardoso P., Erwin T.L., Borges P.A.V., New T.R. 2011. The seven impediments in invertebrate conservation and how to overcome them. Biol. Conserv. 144:2647–2655.

Colwell R.K. 2013. EstimateS: Statistical estimation of speciesrichness and shared species from samples. User’s Guide andapplication available online athttp://purl.oclc.org/estimates

Cooke G.M., Chao N.L., Beheregaray L.B. 2012. Five Cryptic Species in the Amazonian Catfish Centromochlus existimatus Identified Based on Biogeographic Predictions and Genetic Data. PLoS One. 7:e48800.

Cornwell W.K., Pearse W.D., Dalrymple R.L., Zanne A.E. 2019. What we (don’t) know about global plant diversity. Ecography (Cop.). 42:1819–1831.

Corrêa C.E., Petry A.C., Hahn N.S. 2009. Influência do ciclo hidrológico na dieta e estrutura trófica da ictiofauna do Rio Cuiabá, Pantanal Mato-Grossense. Iheringia. Série Zool. 99:456–463.

Curran D.J. 1989. Phylogenetic Relationships among the Catfish Genera of the Family Auchenipteridae (Teleostei: Siluroidea). Copeia. 1989:408–419.

Dagosta F.C.P., Pinna M. De. 2019. The Fishes of the Amazon: Distribution and Biogeographical Patterns, with a Comprehensive List of Species. Bull. Am. Museum Nat. Hist. 2019:1.

Dary E.P., Ferreira E., Zuanon J., Röpke C.P. 2017. Diet and trophic structure of the fish assemblage in the mid-course of the Teles Pires River, Tapajós River basin, Brazil. Neotrop. Ichthyol. 15.

Dennis R.L.H., Thomas C. 2000. Bias in Butterfly Distribution Maps: The Influence of

Hot Spots and Recorder's Home Range. J. Insect Conserv. 4:73–77.

Devincenzi G.J., Teague G.W. 1942. Ictiofauna del Rio Uruguay medio. An. del Mus. Hist. Nat. Montevideo. 5:1–103.

Diniz-Filho J.A.F., Loyola R.D., Raia P., Mooers A.O., Bini L.M. 2013. Darwinian shortfalls in biodiversity conservation. Trends Ecol. Evol. 28:689–695.

Escribano N., Ariño A.H., Galicia D. 2016. Biodiversity data obsolescence and land uses changes. PeerJ. 4:e2743.

Ferraris Jr. C.J. 1988. The Auchenipteridae: putative monophyly and systematics, with a classification of the neotropical doradoid catfishes. Unpublished D. Phil. Thesis, City University of New York.

Ferraris Jr. C.J. 2003. Family Auchenipteridae. In R. E. Reis, S. O. Kullander & C. J. Ferraris, Jr. (Eds.), Checklist of the freshwater fishes of South and Central America (pp. 470–482). Porto Alegre, RS: Edipucrs.

Freitas T.M. da S., Almeida V.H. da C., Valente R. de M., Montag L.F. de A. 2011. Feeding ecology of *Auchenipterichthys longimanus* (Siluriformes: Auchenipteridae) in a riparian flooded forest of Eastern Amazonia, Brazil. Neotrop. Ichthyol. 9:629–636.

Freitas T.M.S., Montag L.F.A., De Marco P., Hortal J. 2020. How reliable are species identifications in biodiversity big data? Evaluating the records of a Neotropical fish family in online repositories. Syst. Biodivers. *in press*.

García-Melo J.E., Oliveira C., Da Costa Silva G.J., Ochoa-Orrego L.E., Garcia Pereira L.H., Maldonado-Ocampo J.A. 2019. Species delimitation of neotropical Characins (Stevardiinae): Implications for taxonomy of complex groups. PLoS One. 14:e0216786.

Gotelli N.J. 2008. A primer of ecology. Vermont: Sinauer Associates.

Guerra V., Llusia D., Gambale P.G., Morais A.R. de, Márquez R., Bastos R.P. 2018. The advertisement calls of Brazilian anurans: Historical review, current knowledge and future directions. PLoS One. 13:e0191691.

Hanya G., Chapman C.A. 2013. Linking feeding ecology and population abundance: a review of food resource limitation on primates. Ecol. Res. 28:183–190.

Hashimoto S., Py-Daniel L.H.R., Batista J.S. 2020. A molecular assessment of species diversity in Tympanopleura and Ageneiosus catfishes (Auchenipteridae: Siluriformes). J. Fish Biol. 96:14–22.

Hortal J., de Bello F., Diniz-Filho J.A.F., Lewinsohn T.M., Lobo J.M., Ladle R.J. 2015.

Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annu. Rev. Ecol. Evol. Syst. 46:523–549.

Hortal J., Diniz-Filho J.A.F., Bini L.M., Rodríguez M.Á., Baselga A., Nogués-Bravo D., Rangel T.F., Hawkins B.A., Lobo J.M. 2011. Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. Ecol. Lett. 14:741–748.

Hortal J., Garcia-Pereira P., García-Barros E. 2004. Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. Ecography (Cop.). 27:68–82.

Hortal J., Jiménez-Valverde A., Gómez J.F., Lobo J.M., Baselga A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos. 117:847–858.

Hortal J., Lobo J.M., Jiménez-Valverde A. 2007. Limitations of Biodiversity Databases: Case Study on Seed-Plant Diversity in Tenerife, Canary Islands. Conserv. Biol. 21:853–863.

Isaac N.J.B., Pocock M.J.O. 2015. Bias and information in biological records. Biol. J. Linn. Soc. 115:522–531.

Jézéquel C., Tedesco P.A., Maldonado-Ocampo J.A., Prada-Pedreros S., Ortega H., Hildalgo M., Martens K., Wever A. De, Zuanon J., Torrente-Vilara G., Sousa L.M., Guiarrizzo T., Valdiviezo J., Sarmiento J., Maldonado M., Carvajal-Vallejos F.M., Yunoki T., Nascimiento C. Do, Agudelo E., Mojica J.I., Herrera R.R., Villa F., Oberdorff T. 2019. Metadata description of the AMAZONFISH database. Freshw. Metadata J. 43:1–9.

Joppa L.N., Roberts D.L., Myers N., Pimm S.L. 2011. Biodiversity hotspots house most undiscovered plant species. Proc. Natl. Acad. Sci. 108:13171–13176.

Laubach T., von Haeseler A., Lercher M.J. 2012. TreeSnatcher plus: capturing phylogenetic trees from images. BMC Bioinformatics. 13:1–5.

Lobo J.M., Baselga A., Hortal J., Jiménez-Valverde A., Gómez J.F. 2007. How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? Divers. Distrib. 13:772–780.

Lobo J.M., Martin-Piera F. 2002. Searching for a Predictive Model for Species Richness of Iberian Dung Beetle Based on Spatial and Environmental Variables. Conserv. Biol. 16:158–173.

Lomolino M. V. 2004. Conservation biogeography. In: Lomolino M. V., Heaney L.R., editors. In Frontiers of Biogeography: New Directions in the Geography of Nature. Sunderland: Sinauer. p. 293–296.

López-Rodríguez A., Silva I., de Ávila-Simas S., Stebniki S., Bastian R., Massaro M.V., Pais J., Tesitore G., Teixeira de Mello F., D’Anatro A., Vidal N., Meerhoff M., Reynalte-Tataje D.A., Zaniboni-Filho E., González-Bergonzoni I. 2019. Diets and Trophic Structure of Fish Assemblages in a Large and Unexplored Subtropical River: The Uruguay River. Water. 11:1–26.

Lundberg J.G., Kottelat M., Smith G.R., Stiassny M.L.J., Gill A.C. 2000. So Many Fishes, So Little Time: An Overview of Recent Ichthyological Discovery in Continental Waters. Ann. Missouri Bot. Gard. 87:26.

Maldonado C., Molina C.I., Zizka A., Persson C., Taylor C.M., Albán J., Chilquillo E., Rønsted N., Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? Glob. Ecol. Biogeogr. 24:973–984.

Marinho M.M.F., Peixoto L., Datovo A., Pinna M. de, Britski H., Menezes N.A., Figueiredo J.L. de, Oyakawa O.T., Gianeti M., Loboda T. 2019. MZUSP The ichthyological Collection at Museu de Zoologia da Universidade de São Paulo. Bol. Soc. Bras. Ictiol. 119:94–101.

Mazzoldi C., Lorenzi V., Rasotto M.B. 2007. Variation of male reproductive apparatus in relation to fertilization modalities in the catfish families Auchenipteridae and Callichthyidae (Teleostei: Siluriformes). J. Fish Biol. 70:243–256.

Mees G.F. 1974. The Auchenipteridae and Pimelodidae of Suriname (Pisces, Nematognathi). Zool. Verh. 132:1–256.

Melo B.F., Oliveira C. 2017. Three new species of Curimatopsis (Characiformes: Curimatidae) from the Amazon basin. J. Fish Biol. 91:528–544.

Menezes N.A., Figueiredo J.L., Britski H.A. 1997. lchthyological Collection Building at the Museu de Zoologia da Universidade de Sao Paulo, Brazil. Am. Soc. Ichthyol. Herpetol.:561–565.

Mérona B. de, Santos G.M. dos, Almeida R.G. de. 2001. Short term effects of Tucuruí Dam (Amazonia, Brazil) on the trophic organization of fish communities. Environ. Biol. Fishes. 60:375–392.

Meyer C., Kreft H., Guralnick R., Jetz W. 2015. Global priorities for an effective information basis of biodiversity distributions. Nat. Commun. 6:8221.

Mora C., Tittensor D.P., Myers R.A. 2008. The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. Proc. R. Soc. B Biol. Sci. 275:149–155.

Moritz C. 1995. Uses of molecular phylogenies for conservation. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 349:113–118.

Oksanen J., Blanchet F.G., Friendly M., Kindt R., Legendre P., McGlinn D., Minchin P.R., O’Hara R.B., Simpson G.L., Solymos P., Stevens M.H.H., Eduard S., Wagner H. 2019. Vegan: Community Ecology Package. .

Ota R.R., Message H.J., da Graça W.J., Pavanelli C.S. 2015. Neotropical Siluriformes as a Model for Insights on Determining Biodiversity of Animal Groups. PLoS One. 10:e0132913.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Pott C.M., Estrela C.C. 2017. Histórico ambiental: desastres ambientais e o despertar de um novo pensamento. Estud. Avançados. 31:271–283.

Pouilly M., Yunoki T., Rosales C., Torres L. 2004. Trophic structure of fish assemblages from Mamoré River floodplain lakes (Bolivia). Ecol. Freshw. Fish. 13:245–257.

R Core Team R. 2015. R: A language and environment for statistical computing.

Rapp Py-Daniel L.H., Ito P.M.M., Ota R.P., Soares I.M., Bastos D.A., Hashimoto S. 2015. No 30 anos da maior coleção de peixes amazônicos: A Coleção de Peixes do INPA. Bol. Soc. Bras. Ictiol. 116:4–16.

Reis R.E., Albert J.S., Di Dario F., Mincarone M.M., Petry P., Rocha L.A. 2016. Fish biodiversity and conservation in South America. J. Fish Biol. 89:12–47.

Reis R.E., Borges T.A.K. 2006. The South American Catfish Genus Entomocorus (Ostariophysi: Siluriformes: Auchenipteridae), with the Description of a New Species from the Paraguay River Basin. Copeia. 2006:412–422.

Ribeiro F.R. V. 2010. Sistemática do gênero Ageneiosus La Cépède (Siluriformes; Auchenipteridae). Unpublished PhD Dissertation, Instituto Nacional de Pesquisas da Amazônia.

Ribeiro F.R. V., Rapp Py-Daniel L.H., Walsh S.J. 2017. Taxonomic revision of the South American catfish genus Ageneiosus (Siluriformes: Auchenipteridae) with the description of four new species. J. Fish Biol. 90:1388–1478.

Rodrigues A.S.L., Gray C.L., Crowter B.J., Ewers R.M., Stuart S.N., Whitten T., Manica A. 2010. A Global Assessment of Amphibian Taxonomic Effort and

Expertise. Bioscience. 60:798–806.

Roxo F.F., Ochoa L.E., Costa-Silva G.J., Oliveira C. 2015. Species delimitation in Neoplecostomus (Siluriformes: Loricariidae) using morphologic and genetic approaches. DNA Barcodes. 3:110–117.

Royero R. 1999. Studies on the systematics and phylogeny of the catfish family Auchenipteridae (Teleostei: Siluriformes). Unpublished PhD Thesis, University of Bristol.

Sarmento-Soares L.M., Cabeceira F.G., Carvalho L.N., Zuanon J., Akama A. 2013. Centromochlus meridionalis, a new catfish species from the southern Amazonian limits, Mato Grosso State, Brazil (Siluriformes: Auchenipteridae). Neotrop. Ichthyol. 11:797–808.

Sarmento-Soares L.M., Lazzarotto H., Py-Daniel L.R., Leitão R.P. 2016. A new Centromochlus Kner, 1858 (Siluriformes: Auchenipteridae: Centromochlinae) from the transition between Amazon floodplain and Guiana shield, Brazil. Neotrop. Ichthyol. 14.

Sarmento-Soares L.M., Martins-Pinheiro R.F. 2008. A systematic revision of Tatia (Siluriformes: Auchenipteridae: Centromochlinae). Neotrop. Ichthyol. 6:495–542.

Sastre P., Lobo J.M. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. Biol. Conserv. 142:462–467.

Sheth B.P., Thaker V.S. 2017. DNA barcoding and traditional taxonomy: an integrated approach for biodiversity conservation. Genome. 60:618–628.

Silva T.M. da, Santos J.C. dos, Ferreira V.A.V., Cruz Ramos L.A. da, Wosiacki W.B., Sousa M.P.A. de. 2017. Data from the ichthyological collection of the Museu Paraense Emílio Goeldi. Zookeys. 687:89–99.

Soberón J., Peterson T. 2004. Biodiversity informatics: managing and applying primary biodiversity data. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 359:689–698.

Solé M., Rödder D. 2000. Dietary assessments of adult amphibians. In: Dodd Junior C.K., editor. Amphibian ecology and conservation: a handbook of techniques. Oxford: Oxford University Press. p. 167–184.

Stropp J., Ladle R.J., M. Malhado A.C., Hortal J., Gaffuri J., H. Temperley W., Olav Skøien J., Mayaux P. 2016. Mapping ignorance: 300 years of collecting flowering plants in Africa. Glob. Ecol. Biogeogr. 25:1085–1096.

Thompson P.L., Davies T.J., Gonzalez A. 2015. Ecosystem Functions across Trophic Levels Are Linked to Functional and Phylogenetic Diversity. PLoS One.

10:e0117595.

Thomson S.A., Pyle R.L., Ahyong S.T., Alonso-Zarazaga M., Ammirati J., Araya J.F., Ascher J.S., Audisio T.L., Azevedo-Santos V.M., Bailly N., Baker W.J., Balke M., Barclay M.V.L., Barrett R.L., Benine R.C., Bickerstaff J.R.M., Bouchard P., Bour R., Bourgoin T., Boyko C.B., Breure A.S.H., Brothers D.J., Byng J.W., Campbell D., Ceríaco L.M.P., Cernák I., Cerretti P., Chang C.-H., Cho S., Copus J.M., Costello M.J., Cseh A., Csuzdi C., Culham A., D'Elía G., d'Udekem d'Acoz C., Daneliya M.E., Dekker R., Dickinson E.C., Dickinson T.A., van Dijk P.P., Dijkstra K.-D.B., Dima B., Dmitriev D.A., Duistermaat L., Dumbacher J.P., Eiserhardt W.L., Ekrem T., Evenhuis N.L., Faille A., Fernández-Triana J.L., Fiesler E., Fishbein M., Fordham B.G., Freitas A.V.L., Friol N.R., Fritz U., Frøslev T., Funk V.A., Gaimari S.D., Garbino G.S.T., Garraffoni A.R.S., Geml J., Gill A.C., Gray A., Grazziotin F.G., Greenslade P., Gutiérrez E.E., Harvey M.S., Hazevoet C.J., He K., He X., Helfer S., Helgen K.M., van Heteren A.H., Hita Garcia F., Holstein N., Horváth M.K., Hovenkamp P.H., Hwang W.S., Hyvönen J., Islam M.B., Iverson J.B., Ivie M.A., Jaafar Z., Jackson M.D., Jayat J.P., Johnson N.F., Kaiser H., Klitgård B.B., Knapp D.G., Kojima J., Kõljalg U., Kontschán J., Krell F.-T., Krisai-Greilhuber I., Kullander S., Latella L., Lattke J.E., Lencioni V., Lewis G.P., Lhano M.G., Lujan N.K., Luksenburg J.A., Mariaux J., Marinho-Filho J., Marshall C.J., Mate J.F., McDonough M.M., Michel E., Miranda V.F.O., Mitroiu M.-D., Molinari J., Monks S., Moore A.J., Moratelli R., Murányi D., Nakano T., Nikolaeva S., Noyes J., Ohl M., Oleas N.H., Orrell T., Páll-Gergely B., Pape T., Papp V., Parenti L.R., Patterson D., Pavlinov I.Y., Pine R.H., Poczai P., Prado J., Prathapan D., Rabeler R.K., Randall J.E., Rheindt F.E., Rhodin A.G.J., Rodríguez S.M., Rogers D.C., Roque F. de O., Rowe K.C., Ruedas L.A., Salazar-Bravo J., Salvador R.B., Sangster G., Sarmiento C.E., Schigel D.S., Schmidt S., Schueler F.W., Segers H., Snow N., Souza-Dias P.G.B., Stals R., Stenroos S., Stone R.D., Sturm C.F., Štys P., Teta P., Thomas D.C., Timm R.M., Tindall B.J., Todd J.A., Triebel D., Valdecasas A.G., Vizzini A., Vorontsova M.S., de Vos J.M., Wagner P., Watling L., Weakley A., Welter-Schultes F., Whitmore D., Wilding N., Will K., Williams J., Wilson K., Winston J.E., Wüster W., Yanega D., Yeates D.K., Zaher H., Zhang G., Zhang Z.-Q., Zhou H.-Z. 2018. Taxonomy based on science is necessary for global conservation. PLOS Biol. 16:e2005075.

Troudet J., Grandcolas P., Blin A., Vignes-Lebbe R., Legendre F. 2017. Taxonomic bias in biodiversity data and societal preferences. Sci. Rep. 7:9132.

von Staden L., Raimondo D., Dayaram A. 2013. Taxonomic research priorities for the conservation of the South African flora. S. Afr. J. Sci. 109:1–10.

Wheeler Q.D. 2004. Taxonomic triage and the poverty of phylogeny. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 359:571–583.

Winemiller K.O. 1989. Patterns of variation in life history among South American fishes in seasonal environments. Oecologia. 81:225–241.

Winston J.E. 1999. Describing Species: A Practical Taxonomic Procedure for Biologists. Cambridge: Cambridge University Press.

Yang W., Ma K., Kreft H. 2013. Geographical sampling bias in a large distributional database and its effects on species richness-environment models. J. Biogeogr. 40:1415–1426.

Zar J.H. 2010. Biostatistical Analysis. New Jersey: Prentice Hall.

Zhang Z.-Q. 2011. Animal biodiversity: An introduction to higher-level classification and taxonomic richness. Zootaxa. 3148:7–12.

4. SESSÃO III

Niche conservatism in the diet of auchenipterid catfishes

A terceira sessão desta tese será formatada conforme as normas da publicação científica *Journal of Evolutionary Biology*, disponível em: https://onlinelibrary.wiley.com/journal/14209101

Niche conservatism in the diet of auchenipterid catfishes

Tiago M.S. Freitas[1,2], Lucas Jardim[3], Bárbara B. Calegari[4,5], Fabricio Villalobos[3,6], Paulo De Marco Jr.[3], Joaquín Hortal[3,7], Luciano F.A. Montag[8]

[1]Faculdade de Ciências Naturais, Campus Universitário do Marajó-Breves, Universidade Federal do Pará, Alameda IV, 3418, Parque Universitário, 68800-000, Breves, PA, Brazil. freitastms@gmail.com. [2]Programa de Pós-graduação em Ecologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Rua Augusto Corrêa, 01, Guamá, 66075-110, Belém, PA, Brazil. [3]Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Avenida Esperança, 74001-970, Goiânia, GO, Brazil. LJ – lucas.ljardim@gmail.com; PDM - pdemarcojr@gmail.com. [4]Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS); Laboratório de Sistemática de Vertebrados. Av. Ipiranga, 6681, CEP: 90619-900, Porto Alegre, RS, Brazil. [5]Smithsonian Institution; Department of Vertebrate Zoology, National Museum of Natural History, 20560, Washington, DC, USA. barbara.calegari@gmail.com. [6]Laboratorio de Macroecología Evolutiva, Red de Biología Evolutiva, Instituto de Ecología, A.C. Xalapa, Veracruz, Mexico. fabricio.villalobos@gmail.com [7]Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales (MNCN-CSIC), C/José Gutiérrez Abascal 2, 28006, Madrid, Spain. jhortal@mncn.csic.es. [8]Laboratório de Ecologia e Conservação, Instituto de Ciências Biológicas, Universidade Federal do Pará, Rua Augusto Corrêa, 01, Guamá, 66075-110, Belém, PA, Brazil. lfamontag@gmail.com.

Niche conservatism in auchenipterid fishes

4.1. ABSTRACT

Diet plays a preeminent role in the evolution of animal lineages, and is generally assumed to be largely conserved across the animal tree of life. However, knowledge on the evolution and conservatism of fish diet is far from complete. Our aim was to map and quantify the variations in diet of Auchenipteridae, a Neotropical catfish family. The variety of dietary habits combined with a well-supported phylogenetic hypothesis allows investigating phylogenetic trophic niche conservatism in this family. We used a comprehensive database on auchenipterid species feeding habits and the most recent Bayesian inference topology for the family, to answer two main questions: (i) how did feeding habits evolve across the phylogeny? and (ii) are feeding habits conserved over time? Feeding habits showed different probabilities of being gained or lost along evolutionary time. Food sources like aquatic and terrestrial insects, crustaceans and fish showed constant (or even increasing) chances to appear in the diet throughout the phylogeny, and also showed balanced transition rates between being present or absent in the diet. On the contrary, fruits, vegetal matter, plankton, and detritus showed low probabilities to be retained across the phylogeny. Based on the phylogenetic signal, we detected a trend towards conservatism for five feeding habits: terrestrial insects, fish, fruit, plankton, and detritus. This contrasts with the lack of phylogenetic structure of feeding upon aquatic insects and vegetal matter, which were randomly distributed across the phylogeny. Importantly, stronger diet overlaps among species occur at shorter phylogenetic distances, supporting phylogenetic niche conservatism for diet preferences. These results are essential to understand the process of character evolution and ecological diversification in auchenipterid fishes.

Keywords: phylogenetic niche conservatism; ancestral reconstruction; character evolution; feeding habit; Neotropical fishes.

4.2. INTRODUCTION

Diet is a fundamental aspect of an animal's life. Knowledge about what an animal eats is of paramount importance to understand its biology (Gerking, 1994; Garvey & Whiles, 2016). Compared to the ecological implications of an animal's diet (Bels & Herrel, 2019), the role of diet and feeding on the evolution of lineages is less well understood, particularly at large phylogenetic scales (i.e. whole clades; Román-Palacios et al. 2019). Until now, dietary information has been used to study different evolutionary aspects of animals such as the marine–freshwater transitions in fishes (Davis *et al.*, 2012), the evolution of the chemosensory system of squamates (Baeckens *et al.*, 2017), the relationship between body size and jaw shape in small mammals (Zelditch *et al.*, 2017), as well as aspects related to their geographical distributions and diversity patterns (Litsios *et al.*, 2012; Olalla-Tárraga *et al.*, 2017). However, our knowledge about large-scale evolutionary patterns of animals' diet, like the ancestral diet of a whole clade, its conservatism or lability throughout a clade's history, remains scarce (Román-Palacios *et al.*, 2019).

More recently, several studies have suggested that diet influences evolutionary diversification in both vertebrates (Price *et al.*, 2012; Burin *et al.*, 2016) and invertebrates (Wiens *et al.*, 2015), and is also conserved across the animal tree of life (Román-Palacios *et al.*, 2019). The conservatism of ecological traits over evolutionary time has been intensively discussed among biologists and ecologists (Blomberg *et al.*, 2003; Fritz & Purvis, 2010) and has been a great deal of interest in recent years (Crisp & Cook, 2012; Anderson & Wiens, 2017; Floeter *et al.*, 2018). This idea remarks to Darwin's "*On the origin of species*", and implies that closely related taxa show a clear tendency to share a similar ecology (Peterson *et al.*, 1999; Webb *et al.*, 2002; Wiens & Graham, 2005). The resultant pattern of similarity in ecological niches amongst related species has been termed as phylogenetic niche conservatism (hereafter PCN; Losos, 2008).

Despite being a hot topic in the current scientific literature, the precise patterns of evolution of ecological traits still suffer some important knowledge gaps (Diniz-Filho *et al.*, 2013a; Hortal *et al.*, 2015). For instance, as noted by Olalla-Tárraga *et al.* (2017), studies of niche conservatism commonly focus on the Grinnellian niche (i.e., the environmental conditions defining species geographic distributions; e.g. Hawkins *et al.*, 2006; Soberón, 2007; Culumber & Tobler, 2016) and not on the Eltonian niche (i.e., species interactions and resource-consumer dynamics; e.g. Soberón, 2007; Manlick *et* *al.*, 2019). However, knowledge of species' Eltonian niches is crucial for understanding the link between biodiversity and ecosystem functioning (Dehling & Stouffer, 2018). In this case, the lack of studies investigating the phylogenetic conservatism of Eltonian niches results from the limited availability of basic data on species life histories such as their dietary preferences and foraging behavior (Rosado *et al.*, 2016; Dehling & Stouffer, 2018). In addition, such an investigation on Eltonian PNC is further limited by the lack of understanding about the phylogenetic relationships among species (Diniz-Filho *et al.*, 2013a; Assis, 2018). These limitations have been particularly important to hinder our understanding of the evolutionary patters of animal diet (Abrams, 2000; Svanbänk & Persson, 2004).

Fishes comprise more than half of all vertebrate species, showing a wide variety of body forms and sizes (Nelson, 2016). This large species diversity is accompanied by a remarkable diversity of feeding habits (Gerking, 1994; Wootton, 1998; Abelha *et al.*, 2001). There are fish representatives of almost every trophic category, from specialized grazing herbivores to species that feed by filtering tiny organisms, to large carnivores or even decomposers (Moyle & Cech-Jr., 2004; Helfman *et al.*, 2009; Hickman *et al.*, 2012). Despite recent efforts to understand the phylogenetic relationships of large groups of fish (Betancur-R *et al.*, 2017; Dai *et al.*, 2018; Rabosky *et al.*, 2018), species-level phylogenies including comprehensive taxa sampling are still scarce (Correa *et al.*, 2007). Given such paucity of knowledge about the evolutionary relationships between species, a reasonable approach is to investigate evolutionary patterns in well-studied groups.

Auchenipteridae (order Siluriformes) is a family of driftwood catfishes widespread through the Neotropical region (Ferraris Jr., 2003), for which a comprehensive and well-supported species-level phylogeny is available (Calegari *et al.*, 2019). This family presents a remarkable ecological variety; there are species that use trunks and rock cavities for diurnal shelter, and there are those epibenthic and active swimmers of open water-column or just below the water surface (Ferraris Jr., 2003; Freitas *et al.*, 2011). Further, although auchenipterids are commonly described as nocturnal or crepuscular fishes with omnivorous feeding habits, their diet can range from plankton filters (Rodriguez *et al.*, 1990; Pouilly *et al.*, 2004) to primarily piscivorous species (Royero, 1993; Barbarino Duque & Winemiller, 2003). Despite the remarkable spectrum of interspecific variation in feeding habits shown by auchenipterids, the evolution of their diet has not been studied yet.

Here we use an exhaustive compilation of data on trophic habits (Freitas et al. in prep.) and a comprehensive species-level phylogeny (Calegari *et al.*, 2019) to investigate both diet evolutionary patterns and the existence of phylogenetic trophic niche conservatism on Auchenipterid fishes. More precisely, we evaluated evolutionary variations in auchenipterid diet by addressing two main questions: (i) how did feeding habits evolve across the phylogeny? and (ii) are feeding habits conserved over evolutionary time? To answer these questions, we applied phylogenetic comparative methods to describe the evolutionary pattern of auchenipterid diet and reconstruct their ancestral state. Fishes in general, and auchenipterids in particular, are thought to be trophic opportunists (and therefore mainly generalists), so we expect that most, if not all, trophic habits evolve repeatedly across the phylogeny. However, some trophic habits such as eating fruits are rarely found, so we expect that several trophic habits present PNC. Indeed, we found that some food types such as aquatic insects, terrestrial insects, crustaceans, and fish showed balanced transition rates between being present or absent in the diet of auchenipterid fishes. Despite that, most feeding habits (i.e. terrestrial insects, fish, fruit, plankton, and detritus) tended towards a conservatism pattern. Although other feeding habits such as aquatic insects and vegetal matter showed no phylogenetic structure, the stronger diet-overlaps among closely related species shown by our data support the existence of PNC on auchenipterid diet.

4.3. MATERIALS AND METHODS

4.3.1. Feeding data

We searched the scientific literature for studies on feeding habits of auchenipterid fishes using the online databases of *ISI Web of Science* (www.webofknowledge.com), *Scielo* (www.scielo.org), and *Google Scholar* (www.scholar.google.com). This survey considered only scientific literature published before December 2017 (gray literature were not used). To do so, we used the keywords “diet”, “feeding”, “feeding ecology”, and “trophic ecology”, in combination with “Auchenipteridae” or the names of genera (e.g., "*Auchenipterus* AND feeding"). To cover the specialized scientific journals with international and local relevancy, keywords were searched in Portuguese, English, Spanish, and French. Because many of these scientific journals are not indexed at *ISI* or *Scielo*, no standardized protocol was established to search the literature. The searches were done exhaustively and ended when we judged we had covered the total information. We selected all publications from journals with editorial board that mentioned the feeding habit of auchenipterid species and distinguish eight diet categories as follows: (1) aquatic insects (larvae and adults), (2) terrestrial insects (larvae and adults), (3) crustacean, (4) fish, (5) fruit (including seeds), (6) vegetal matter (plant parts as barks, roots, and flowers), (7) plankton (icthyo-, phyto-, and zooplankton), and (8) detritus. We allowed a single species to be assigned to more than one feeding habit.

We found 67 published studies on auchenipterid gut contents, covering 54 species across 19 genera, which represent approximately 40% of the total species richness for the family (= 125 valid species). However, we based our analyses only on species that had any diet information available and that had information regarding its phylogenetic relationship (see below). Thus, our numbers were reduced to 62 publications and 47 species from 19 genera. Although data coverage was limited for the most speciose genera, its distribution along auchenipterids was relatively unbiased (see Freitas et al., in prep. for further information on the data and its completeness).

4.3.2. Phylogenetic data

We used a recent divergent time analysis for Auchenipteridae lineages dated in millions of years, provided by Calegari *et al.* (unpublished). Despite unpublished, this evolutionary hypothesis exactly reflects the consensus phylogenetic tree of Calegari *et al.* (2019) produced under Bayesian inference. This phylogenetic hypothesis is the most recent and comprehensive species-level phylogeny for Auchenipteridae, using 97 out of 124 valid species (updated to 125 species after Rocha *et al.,* 2019), and shows remarkable consistency with the root topology of other recent phylogenies (Freitas et al., in prep.). For our study, we pruned the phylogeny to include only species whose trophic data were available from the literature (see Supplementary Material S1; dating information is not provided). Based on this phylogeny and the compiled data on auchenipterid diet, we evaluated the existence of phylogenetic trophic niche conservatism (PNC) through ancestral state reconstruction, phylogenetic signal and the relationship between dietary overlap (dissimilarity) and phylogenetic distances.

4.3.3. Ancestral diet reconstruction

We performed ancestral diet reconstruction using the maximum-likelihood method (Cunningham *et al.*, 1998; Joy *et al.*, 2016), with the function *ace* of the R package *ape* (Paradis & Schliep, 2019). In this analysis, we estimated the transition rates of each feeding habit using simulations based on two models: (i) the symmetrical model (SYM), considering that the transitions between states 1 and 0 (presence and absence of the feeding habit, respectively), and *vice-versa*, occur at equal rates; and (ii) the all-rates-different model (ARD), which predicts different transition rates between states, forward (0 to 1) and backward (1 to 0). Those models were ranked by the weighted Akaike index (wAIC; Akaike, 1974; Burnham & Anderson, 2004) using the function *aicw* of the R package *geiger* (Harmon *et al.*, 2008). The marginal likelihood for the diet state at each ancestor was inferred using the best model (either SYM or ARD).

4.3.4. Phylogenetic signal analysis

For each feeding habit, we quantified the phylogenetic signal using Fritz & Purvis' (2010) D-statistic, a measurement applicable to binary traits (i.e. presence/absence of a trait). Values of the D-statistic equal or close to one imply a random distribution of the trait across the phylogeny, whereas values equal to or close to zero indicate Brownian motion evolution (Fritz & Purvis, 2010). Negative and positive extreme values of D-statistics can also be found, representing extremely clumped and overdispersed phylogenetic patterns, respectively (Fritz & Purvis, 2010; see also Olalla-Tárraga *et al.*, 2011). Significance (p-value) was estimated by comparing the D values of each binary trait against 1000 permutations of simulated distributions under (1) randomly reshuffled trait values across the tips of the tree, and (2) trait allocation based on Brownian motion. The calculation of the D-statistics was implemented using the function *phylo.d* of the R package *caper* (Orme *et al.*, 2012). Following Cooper et al. (2010), we adopted the niche-drift view, when niche evolution fits a BM model, as evidence for PNC.

4.3.5. Species dietary overlap across phylogeny

Under PNC, closely related species are expected to show higher dietary similarity than distantly related species. To test this expectation, we evaluated the relationship between trophic niche overlap and phylogenetic distance among species. To do so, we first estimated similarities in diets between species pairs by computing Pianka's symmetric index (Pianka, 1973). The index is calculated through the formula: $O_{jk} = \sum_i^n p_{ij} * p_{ik} \Big/ \sqrt{\sum_i^n {p_{ij}}^2 * \sum_i^n {p_{ik}}^2}$, where O_{jk} = value of the Pianka's niche overlap between species *j* and species *k*, p_{ij} = the proportion of the resource *i* among the total resources used by species *j*, p_{ik} = the proportion of the resource *i* among the total resources used by species *k*, and n = Total number of resources. Pianka's index varies between 0 (total separation) and 1 (total overlap). Then, we transformed this similarity index to a measure of dissimilarity by subtracting the observed value from unity (1 minus O_{jk}). This dissimilarity measure was then correlated to the phylogenetic distance between species pairs utilizing a Mantel test using the function *multi.mantel* of the R package *phytools* (Revell, 2012). In this case, a positive correlation means lower diet dissimilarity occurring at shorter phylogenetic distances, and consequently, supporting PNC. Additionally, we performed a Mantel correlogram (Gittleman & Kot, 1990; Diniz-Filho *et al.*, 2013b) using the function *mantel.correlog* of the R package *vegan* (Oksanen *et al.*, 2019). This technique consists of dividing the matrices into several sub-matrices, each one relating the diet-overlap within intervals of phylogenetic distances. Thus, correlations between these sub-matrices were obtained by performing Mantel tests (Diniz-Filho *et al.*, 2013b) (in our case, 10 subsets were used). Through this test, we can describe possible variations in the correlation between species diet and phylogenetic distances.

4.4. RESULTS

Published studies on gut contents of auchenipterid fishes revealed diets dominated by aquatic and terrestrial insects (cited for 28 and 27 species, respectively), followed by crustaceans (23 species), fish (16 species), fruit and vegetal matters (six species each), and plankton and detritus (five species each). For a complete list of species, feeding habits, and references, see Supplementary Material S2.

For all feeding habits, the reconstruction of the ancestral diet was better fit by the ARD than by the SYM model (mean wAIC equal to 0.89 and 0.11, respectively; Table 1). However, diet reconstruction at the root of the phylogeny was unpredictable, with all feeding habits having high probabilities to occur (Figure 1). Based on the transition rates between states (using the ARD model), terrestrial insects, fish, and detritus showed the lowest rate to be either lost or won across the evolutionary history of Auchenipteridae (Figure 2). We found similar rates of transition for aquatic insects and crustaceans. Fruits and plankton showed higher probabilities of being lost in the diet than to (re)appear as a trophic habit. Lastly, vegetal matters showed high transition rates, either to be lost or present in the diet of auchenipterid species (Figure 2).

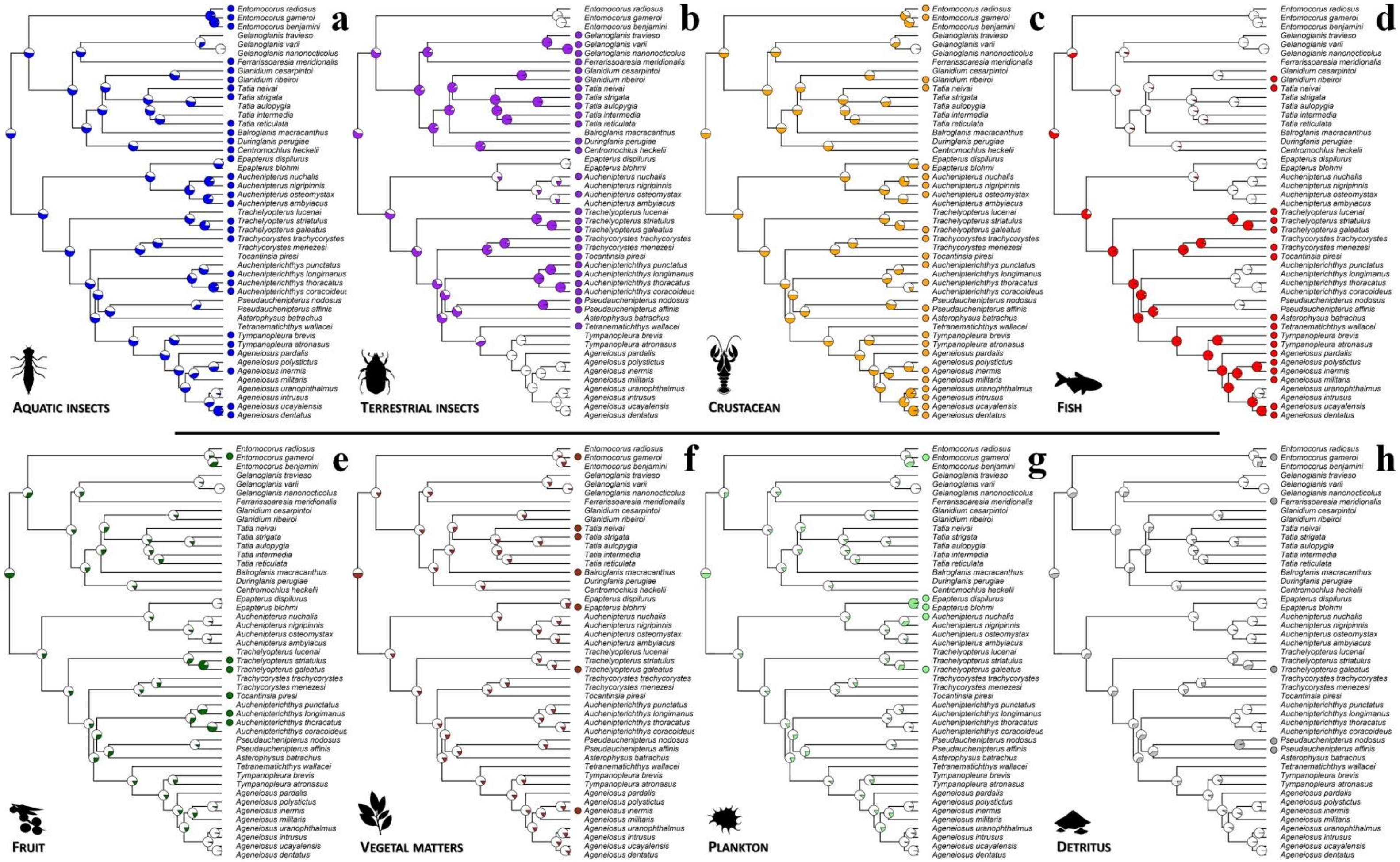


Figure 1. Reconstruction of the evolutionary history of feeding habits for auchenipterid fishes, based on the 47 valid species in the phylogeny proposed by Calegari *et al.* (2019) with data on feeding habits. Piecharts at nodes illustrate the ancestral probability of the diets. **a** – aquatic insects (blue), **b** – terrestrial insects (purple), **c** – crustacean (orange), **d** – fish (red), **e** – fruit (dark green), **f** – vegetal matter (brown), **g** – plankton (light green), and **h** – detritus (gray).

Table 1. Model fitting results for the comparison between symmetrical (SYM) and all-rates-different (ARD) models of diet evolution for auchenipterid fishes. LogL stands for Log-likelihood, ΔAIC for variations in the Akaike Information Coefficient, and wAIC for the Akaike weight.

Feeding habits	**SYM model**			**ARD model**		
	logL	**ΔAIC**	**wAIC**	**logL**	**ΔAIC**	**wAIC**
Aquatic insects	-30.9	3.2	0.168	-30.2	0.0	0.832
Terrestrial insects	-20.3	2.5	0.226	-20.0	0.0	0.774
Crustaceans	-31.4	1.8	0.287	-31.4	0.0	0.713
Fish	-19.5	4.7	0.089	-18.1	0.0	0.911
Fruit	-22.0	11.4	0.003	-17.2	0.0	0.997
Vegetal matters	-27.7	19.0	0.000	-19.1	0.0	1.000
Plankton	-18.6	11.4	0.003	-13.8	0.0	0.997
Detritus	-16.2	4.2	0.110	-15.0	0.0	0.890

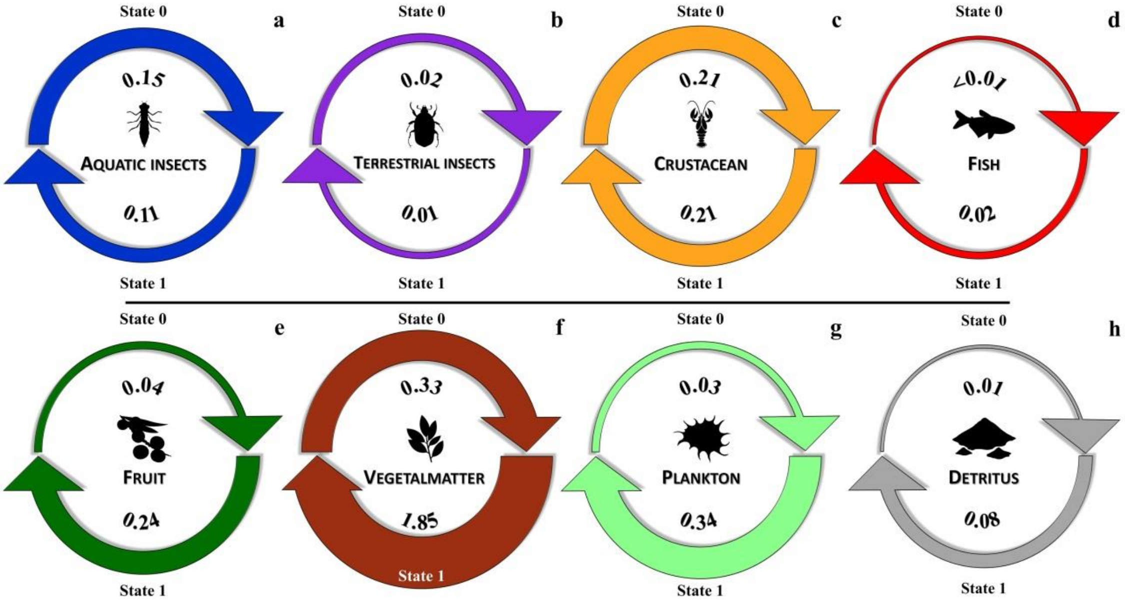


Figure 2. Forward and backward transition rates between the absence and presence of trophic habits (states 0 and 1, respectively) in the phylogeny of auchenipterid fishes, using the all-rates-different (ARD) model. Arrow thickness indicates the rate of character transition for each feeding habit. **a** – aquatic insects (blue), **b** – terrestrial insects (purple), **c** – crustacean (orange), **d** – fish (red), **e** – fruit (dark green), **f** – vegetal matter (brown), **g** – plankton (light green), and **h** – detritus (gray).

Based on the phylogenetic signal, we detected a trend towards conservatism under the Brownian expectation for five feeding habits: terrestrial insects, fish, fruit, plankton and detritus. Conversely, the feeding upon aquatic insects and vegetal matters showed weak phylogenetic signals, and were not clumped under the Brownian motion model, being therefore randomly distributed along the phylogeny of auchenipterids. The crustacean feeding habit showed an intermediate value for phylogenetic signal, but differed from both random pattern and Brownian model (Table 2).

Table 2. Evaluation of phylogenetic signal (D-statistic; Fritz & Purvis, 2010) for feeding habits of auchenipterid fishes. N spp. = the number of species assigned with the specific trait. Values in bold are statistically significant.

Traits	**N spp.**	**D-statistic**	$\mathbf{P_{random}}$	$\mathbf{P_{Brownian}}$
Aquatic insects	28	0.813	0.195	**0.005**
Terrestrial insects	27	-0.333	**<0.001**	0.828
Crustaceans	23	0.616	**0.044**	**0.029**
Fish	16	-0.272	**<0.001**	0.775
Fruit	6	0.317	**0.030**	0.288
Vegetal matters	6	1.143	0.644	**0.005**
Plankton	5	0.252	**0.029**	0.365
Detritus	5	0.490	0.089	0.222

A weak positive correlation between the dietary overlap and phylogenetic distance matrices was detected (Mantel test; $r = 0.253$; $P < 0.001$). However, when considering submatrices, the Mantel correlogram showed a significant and positive correlation in the first three classes of phylogenetic distance ranged between 6.3 and 27.4 (branch length). For intermediate phylogenetic distances (between 69.5 and 80.2), we also detected significant positive correlation values. These two results indicate that species tend to show greater diet overlap (i.e. lower diet dissimilarity) than expected by chance at both shorter and intermediate phylogenetic distances. Conversely, the last distance class (longer branch length) showed significant and negative values of correlation, which means that a higher diet-overlap is also occurring among species at longer phylogenetic distances. See Figure 3 for the detailed results of the Mantel correlograms.

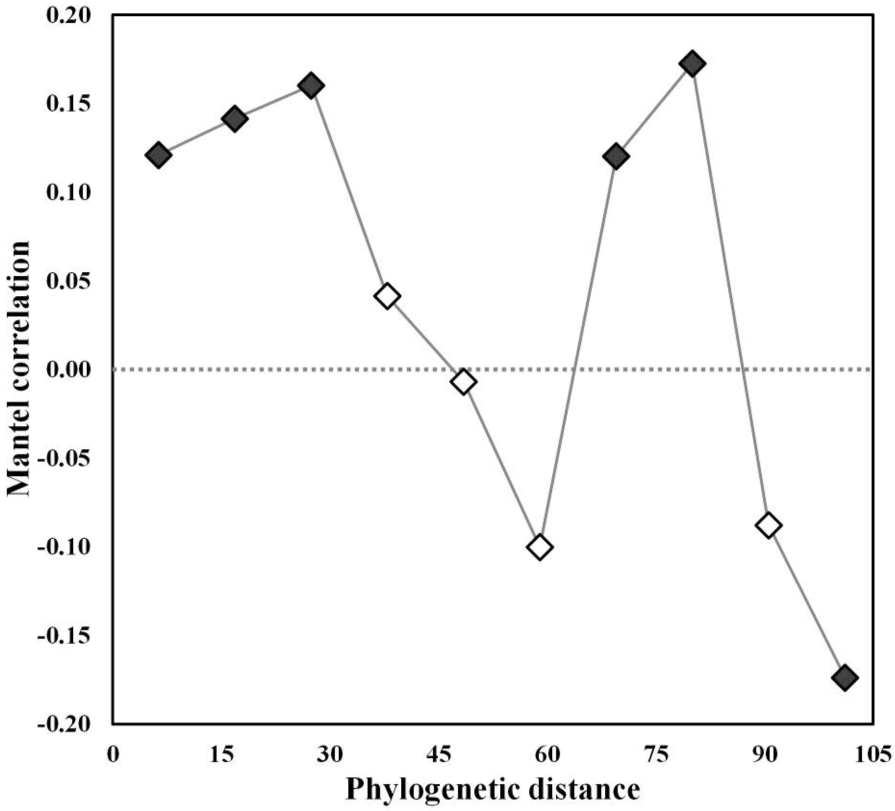


Figure 3. Mantel correlogram between diet-overlap and phylogenetic distances of auchenipterid fishes among classes of phylogenetic distances. Closed diamonds show significant correlation after correcting p-values for multiple testing.

4.5. DISCUSSION

Most of the feeding habits in the diet of auchenipterid fishes are conserved across evolutionary time, so diet is more similar among closely related species. However, the diversity of trophic habits and their high degree of conservatism prevents from identifying the ancestral trophic preferences at the root of the phylogeny based on the diet reconstruction. The extent to which animals retain their ancestral niche along evolutionary time has been subject to debate over the last decades (Losos, 2008; Wiens *et al.*, 2010; Pyron *et al.*, 2015). Our results of conservatism in terrestrial insect and fish feeding habits corroborate the broad expectation on conservatism of carnivorous habit in animals, which seems to be highly phylogenetically conserved across the animal tree of life (Román-Palacios *et al.*, 2019). However, the degree of trophic niche conservatism varies between habits; five out of the eight feeding habits we evaluated showed a trend towards phylogenetic conservatism, while two others showed a random clustering pattern across the phylogeny. Note that the results for feeding upon crustaceans must be interpreted with caution since we cannot assume it as conserved or randomly clustered (Fritz & Purvis, 2010), and additional efforts are needed to better elucidate the evolutionary pattern of this trait. In any case, such preeminence of trophic niche conservatism has been also observed for other Neotropical fish families such as Serrasalmidae (Correa *et al.*, 2007) and Loricariidae (Lujan *et al.*, 2012). Whether this is a general trend or corresponds to a handful of taxa deserves further study, as the evolution of trophic niches is unknown for most Neotropical fish taxa.

Conventionally, studies on phylogenetic niche conservatism and ancestral state reconstruction using Eltonian niche data (e.g. trophic ecology) designate a single trait per species (e.g. carnivore, omnivore, and herbivore) (Correa *et al.*, 2007; Baeckens *et al.*, 2017). However, the use of relatively coarse dietary characterizations can produce bias clustering species without overlapping in the food items they consume (Román-Palacios *et al.*, 2019). A more detailed feeding classification such as the one we used allows identifying distinct patterns that would be hidden in the coarser diet classes. In our study, each food category had its evolutionary path reconstructed for two main reasons. First, freshwater Neotropical fishes are widely known for their food opportunism (Abelha *et al.*, 2001), lacking trophic specializations (with rare exceptions). Therefore, investigating the presence of each diet type across the phylogeny may reveal how such trophic opportunism could have arisen along with the diversification of new diet sources. Second, classifying auchenipterids into broader diet categories could obscure important information about their trophic behavior. For instance, a single category for insectivorous would hide the diversity of foraging behaviors needed to feed on either aquatic or terrestrial insects – foraging in the bottom of the water body (Cabeceira *et al.*, 2015) or near the water surface (Freitas *et al.*, 2011), respectively.

4.5.1. Diet conservatism varies through phylogeny

The positive (although weak) overall relationship between dietary overlap and phylogenetic distances among species provides support for the existence of phylogenetic trophic niche conservatism among auchenipterids. Indeed, closely related species showed more similarity in their diets, and PNC was supported within taxa and/or clades with shorter phylogenetic distances (Figure 4). This corroborates the idea that ecological niches seem to be conserved mostly over shorter timescales or in specific lineages within a larger clade (Wiens *et al.*, 2010; Peterson, 2011). Likewise, although limited support for PNC was found in the trophic niche of mammals as a whole, stronger evidence was detectable at finer phylogenetic scales (i.e. taxonomic orders) and for different feeding habits (i.e. predation on mammals and birds; Olalla-Tárraga *et al.*, 2017). As highlighted by Debastiani & Duarte (2017), even when overall statistics (such as the Mantel test we used) cannot detect phylogenetic structure, complex patterns can still exist at small and restricted time-scales. Interestingly, we also detected high levels of dietary overlap among phylogenetically distant auchenipterids, which would not support PNC as a strong and pervasive driving force in trophic niche evolution. Despite no specific evaluation has been performed herein, the dietary overlap among distantly related taxa within Auchenipteridae could be resulted from repeated evolutionary convergence along unrelated linages (Figure 4).

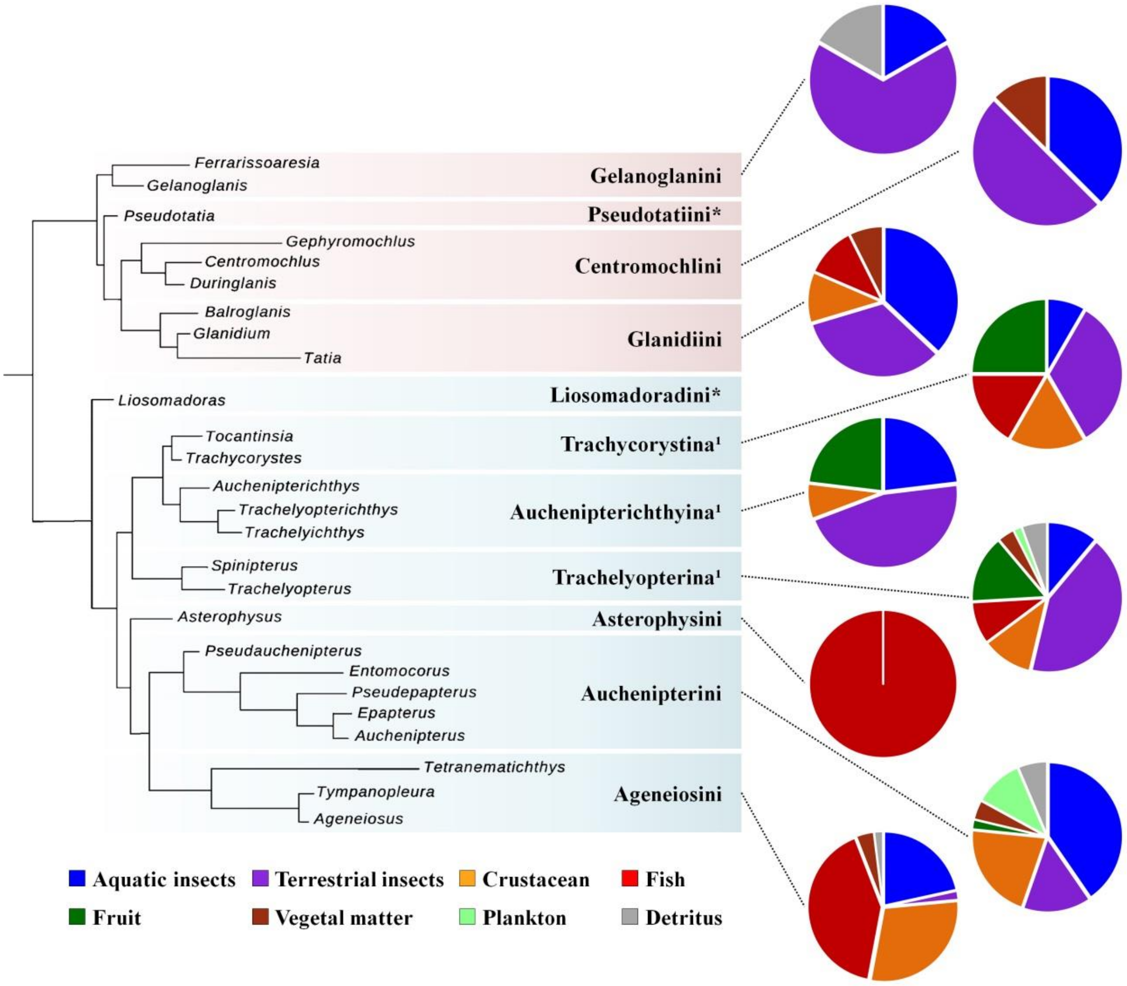


Figure 4. Feeding diversity among Auchenipteridae lineages (sensu Calegari *et al.*, 2019). Pie charts indicate the proportion of each food type. Clades with the number one superscript (¹) represent the tribe Trachelyopterini. *no information on feeding habit is available for these clades.

4.5.2. *Convergence in the evolutionary history of auchenipterid diet*

Our reconstruction of the ancestral diet produced a high degree of uncertainty at the root of the phylogeny for all feeding habits. The data distribution among species may not have been sufficient to estimate the value at the root. In this case, it would be informative to add fossil information given its usefulness in assessing the direction of character evolution (Donoghue *et al.*, 1989). In the absence of such information, based on our results, the available evidence suggests a generalist ancestral carnivorous habit for all auchenipterids, following also the general expectation for animals (Román-Palacios *et al.*, 2019).

According to our reconstruction the probability of retaining the diet depends on the feeding category. All carnivorous habits, including aquatic and terrestrial insects, crustaceans and fish, showed constant or even increased chances to appear in the evolutionary history of the clade, and also showed balanced transition rates between being present or absent on different lineages (Figure 2 and 4). Among them, terrestrial insects and fish feeding habits seem to be more stable across the phylogeny; once one of these two traits appears in the clade’s history, the shift to the absence of state is most unlikely, corroborating their conservatism over time. Taking into consideration the number of auchenipterid species that fed on aquatic and terrestrial insects, we can assume that these habits have been largely retained across the two main Auchenipteridae clades (Centromochlinae and Auchenipterinae subfamilies, *sensu* Calegari et al., 2019). Thus, the most probable scenario is the reversion of the condition for few species of those large groups, as some species had lost these insectivorous habits. Interestingly, carcinophagy and piscivory appeared to evolve mainly within the Auchenipterinae, and could perhaps represent the ancestral acquisition of these habits that remain as potential feeding conditions exclusively shared within the members of this subfamily.

Conversely, fruits, vegetable matter, plankton, and detritus feeding habits seem to have originated independently in different auchenipterid clades since they appeared and reappeared multiple times across the clades’ history (Figure 1). These categories showed lower probabilities to be retained across the tree because of their low probabilities to appear (except in the case of vegetable matter; see Figure 2) and high rates of being lost, which reflect an unstable evolutionary scenario. Note that these diet categories are composed at least by parts of autotrophic organisms. Herbivorous habits demand highly specialized requirements for animals, such as maintaining endosymbiont organisms in the digestive system to help extracting nutrients from plant cell walls (Ley *et al.*, 2008), and in the case of fish, modification of jaws and development of specialized gill rakers on the branchial arches (Lauder, 1982; Lazzaro, 1987; Sanderson *et al.*, 2016). Still, herbivory is thought to have evolved from carnivory (Sanchez & Trexler, 2016; Román-Palacios *et al.*, 2019), to take advantage from the lower costs of eating plants. Among auchenipterids, for instance, *Auchenipterichthys longimanus* uses fruits as energy source in the breeding season (Freitas *et al.*, 2011), although the higher availability of fruits in this season could be a parsimonious explanation for this diet choice. Further, the tribe Auchenipterini (*sensu* parsimonious hypothesis; Calegari et al., 2019) shows several modifications in the branchial apparatus (i.e. number and size of the gill rakers). These morphological structures allow the fish to filter small particles suspended in the water column, such as plankton. In our study, plankton was mostly registered in the diet of species which belong to this tribe (i.e. *Auchenipterus, Entomocorus,* and *Epapterus*), with the exception of *Trachelyopterus galeatus*. We believe that the relationship between the large feeding requirements of auchenipterids and the evolutionary advantages of being an herbivore in water bodies which receive constant inputs of vegetal matter can explain the high transition rates among these diet categories. Indeed, this points to a case of convergent evolution of the diet, and adaptation for which there is widely compelling evidence in animals (Zelditch *et al.*, 2017). Among fishes, the East African rift lakes cichlids are the most impressive (and astonishing) example of adaptive radiation of species. While cichlids display pronounced trophic variety, there is an outstanding morphological convergence among unrelated taxa that evolved to acquire similar diet specialization (Seehausen, 2006). Dietary convergence has been also noticed among Africa and South American electric fishes (Winemiller & Adite, 1997), planktivorous surgeonfishes (Acanthuridae; Friedman *et al.*, 2016), and other vertebrates such as snakes (Vincent *et al.*, 2009) or birds (Olsen, 2017).

4.5.3. *Caveats*

It is important to note that our results may be partly affected by the lack of comprehensive knowledge on the evolution and trophic preferences of auchenipterid fishes. On the one hand, a recent study showed some important limitations in the assessment of ecological characteristics of the species of this group due to taxonomic inaccuracies (Freitas *et al.*, 2020 *in press*). On the other, data on both diet and phylogenetic relationships are scarce and fragmentary in some level (Freitas *et al.*, in prep.). For instance, there are still some uncertainties on the phylogenetic position of some taxa, such as the genus *Entomocorus*. Through Bayesian inference, this genus was recovered within the Centromochlinae subfamily, contrary to the classification hypothesis based on parsimony that recovered this genus in Auchenipterinae (Calegari et al., 2019). Yet, according to the aforementioned authors, *Entomocorus* undoubtedly belongs to Auchenipterinae based on the sharing of many synapomorphies proposed to the subfamily. This change may affect the distribution of diets among groups in our study. Although no specific analysis was performed, phylogenetic signal and ancestral reconstruction could present slightly different results if *Entomocorus* was recovered as an Auchenipterinae member, since the feeding habits of the latter genus were quite similar to the closest related genera (*Epapterus*, *Auchenipterus*, and *Pseudepapterus*, sensu Calegari et al., 2019; no feeding habit was found for latter genus). For instance, if *Entomocorus* were placed in Auchenipterinae, the feeding habit crustaceans would be almost exclusive for this subfamily (except for *Glanidium ribeiroi* and *Tatia neivai*); fruit and plankton would occur only for Auchenipterinae; plankton would be almost an exclusive feeding category for the clade “*Entomocorus* + (*Epapterus* + *Auchenipterus*)”. All these results call for an even higher degree of conservatism, which provides further support for our inference of ancestral and conserved carnivorous habits, as well as for the exclusive development of plankton feeding by Auchenipterinae through morphological changes in their gill rakers. This, together with the relatively even distribution of data on dietary habits among genera (see Freitas *et al.*, in prep.) makes us believe that although the results of our study may change as new phylogenetic and trophic information is available, the general conclusions we raise about the evolution of auchenipterid diet will hold out.

4.5.4. *Concluding remarks*

In sum, our results point to a generalized trophic niche conservatism for Auchenipteridae, in particular for most carnivorous habits. Indeed, insects may be the ancestral food source of this family, although evidence is yet insufficient to corroborate this point. Feeding on fruits and, perhaps, plankton also show some degree of conservatism, although in general herbivorous habits seem to have evolved independently in different clades along the phylogeny, despite the difficulty in developing adaptations to take profit of vegetal matter. These findings improve our understanding of the dietary evolution of the Auchenipteridae, which may be a relevant lineage to understand the diversification of the Neotropical fish fauna as a whole (Hubert & Renno, 2006; Albert & Reis, 2011). Further investigations on ecological and functional constraints for the auchenipterids are required to improve our understanding of their evolution, and assess the eventual development of trophic generalism through repeated evolution of adaptations to feeding upon plant matter.

ACKNOWLEDGEMENTS

The authors are grateful to the financial support provided by several grants from the Brazilian *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES - Finance Code 001; TMSF - 23038.042984/2008-30), and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq; PDM - 308694/2015-5; LFAM - 305017/2016-0; JH – PVE 314523/2014-6). LJ receives grant from the Brazilian *Instituto Nacional de Ciência e Tecnologia* (INCT) in Ecology, Evolution and Biodiversity Conservation funded by CNPq (465610/2014-5) and *Fundação de Amparo à Pesquisa do Estado de Goiás* (FAPEG; process number 201810267000023). FV is supported by CONACyT, Mexico, through the Instituto de Ecología, A.C. LFAM also thanks to PROCAD (UFPA/UFG/UNESP: Ecologia, Genética e Conservação do Cerrado, Amazônia e Mata Atlântica).

4.6. REFERENCES

Abelha, M.C.F., Agostinho, A.A. & Goulart, F. 2001. Plasticidade trófica em peixes de água doce. *Acta Sci.* **23**: 425–434.

Abrams, P.A. 2000. The Evolution of Predator-Prey Interactions: Theory and Evidence. *Annu. Rev. Ecol. Syst.* **31**: 79–105.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**: 716–723.

Albert, J. & Reis, R.E. 2011. *Historical Biogeography of Neotropical Freshwater Fishes*. University of California Press, London.

Anderson, S.R. & Wiens, J.J. 2017. Out of the dark: 350 million years of conservatism and evolution in diel activity patterns in vertebrates. *Evolution (N. Y).* **71**: 1944–1959.

Assis, L.C.S. 2018. Revisiting the Darwinian shortfall in biodiversity conservation. *Biodivers. Conserv.* **27**: 2859–2875.

Baeckens, S., Van Damme, R. & Cooper, W.E. 2017. How phylogeny and foraging ecology drive the level of chemosensory exploration in lizards and snakes. *J. Evol. Biol.* **30**: 627–640.

Barbarino Duque, A. & Winemiller, K.O. 2003. Dietary segregation among large catfishes of the Apure and Arauca Rivers, Venezuela. *J. Fish Biol.* **63**: 410–427.

Bels, V. & Herrel, A. 2019. Feeding, a Tool to Understand Vertebrate Evolution Introduction to “Feeding in Vertebrates.” pp. 1–18.

Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., Miya, M., *et al.* 2017. Phylogenetic classification of bony fishes. *BMC Evol. Biol.* **17**: 162.

Blomberg, S.P., Garland, T. & Ives, A.R. 2003. Testing for phylogenetic signal in comparative data: behaviotal traits are more labile. *Evolution (N. Y).* **57**: 717–745.

Burin, G., Kissling, W.D., Guimarães, P.R., Şekercioğlu, Ç.H. & Quental, T.B. 2016. Omnivory in birds is a macroevolutionary sink. *Nat. Commun.* **7**: 11250.

Burnham, K.P. & Anderson, D.R. 2004. *Model Selection and Multimodel Inference* (K. P. Burnham & D. R. Anderson, eds). Springe, New York, NY.

Cabeceira, F.G., Parisotto, D.C., Zuanon, J. & Akama, A. 2015. The microhabitat, behavior and diet of Centromochlus meridionalis, a small catfish of Amazon streams (Teleostei: Auchenipteridae). *Ichthyol. Explor. Freshwaters* **26**: 221–228.

Calegari, B.B., Vari, R.P. & Reis, R.E. 2019. Phylogenetic systematics of the driftwood catfishes (Siluriformes: Auchenipteridae): a combined morphological and molecular analysis. *Zool. J. Linn. Soc.* **187**: 661–773.

Correa, S.B., Winemiller, K.O., LóPez-Fernández, H. & Galetti, M. 2007. Evolutionary Perspectives on Seed Consumption and Dispersal by Fishes. *Bioscience* **57**: 748–756.

Crisp, M.D. & Cook, L.G. 2012. Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? *New Phytol.* **196**: 681–694.

Culumber, Z.W. & Tobler, M. 2016. Ecological divergence and conservatism: spatiotemporal patterns of niche evolution in a genus of livebearing fishes (Poeciliidae: Xiphophorus). *BMC Evol. Biol.* **16**: 1–16.

Cunningham, C.W., Omland, K.E. & Oakley, T.H. 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* **13**: 361–366.

Dai, W., Zou, M., Yang, L., Du, K., Chen, W., Shen, Y., *et al.* 2018. Phylogenomic Perspective on the Relationships and Evolutionary History of the Major Otocephalan Lineages. *Sci. Rep.* **8**: 1–12.

Davis, A.M., Unmack, P.J., Pusey, B.J., Johnson, J.B. & Pearson, R.G. 2012. Marine-freshwater transitions are associated with the evolution of dietary diversification in terapontid grunters (Teleostei: Terapontidae). *J. Evol. Biol.* **25**: 1163–1179.

Debastiani, V.J. & Duarte, L. da S. 2017. Evolutionary Models and Phylogenetic Signal Assessment via Mantel Test. *Evol. Biol.* **44**: 135–143.

Dehling, D.M. & Stouffer, D.B. 2018. Bringing the Eltonian niche into functional diversity. *Oikos* **127**: 1711–1723.

Diniz-Filho, J.A.F., Loyola, R.D., Raia, P., Mooers, A.O. & Bini, L.M. 2013a. Darwinian shortfalls in biodiversity conservation. *Trends Ecol. Evol.* **28**: 689–695.

Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski, R., Landeiro, V.L., Telles, M.P. de C., *et al.* 2013b. Mantel test in population genetics. *Genet. Mol. Biol.* **36**: 475–485.

Donoghue, M.J., Doyle, J.A., Gauthier, J., Kluge, A.G. & Rowe, T. 1989. The Importance of Fossils in Phylogeny Reconstruction. *Annu. Rev. Ecol. Syst.* **20**: 431–460.

Ferraris Jr., C.J. 2003. Family Auchenipteridae. In: Reis, R.E., Kullander, S. & Ferraris Jr., C.J. *Check List of the Freshwater Fishes of South and Central America*, pp. 473–485.

Floeter, S.R., Bender, M.G., Siqueira, A.C. & Cowman, P.F. 2018. Phylogenetic perspectives on reef fish functional traits. *Biol. Rev.* **93**: 131–151.

Freitas, T.M.S., Almeida, V.H.C., Valente, R.M. & Montag, L.F.A. 2011. Feeding ecology of Auchenipterichthys longimanus (Siluriformes: Auchenipteridae) in a riparian flooded forest of Eastern Amazonia, Brazil. *Neotrop. Ichthyol.* **9**: 629–636.

Friedman, S.T., Price, S.A., Hoey, A.S. & Wainwright, P.C. 2016. Ecomorphological convergence in planktivorous surgeonfishes. *J. Evol. Biol.* **29**: 965–978.

Fritz, S.A. & Purvis, A. 2010. Selectivity in Mammalian Extinction Risk and Threat Types: a New Measure of Phylogenetic Signal Strength in Binary Traits. *Conserv. Biol.* **24**: 1042–1051.

Garvey, J.E. & Whiles, M. 2016. *Trophic Ecology*. CRC Press, Boca Raton: Taylor & Francis, 2016.

Gerking, S.D. 1994. *Feeding Ecology of Fish*. Academic Press, San Diego.

Gittleman, J.L. & Kot, M. 1990. Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects. *Syst. Zool.* **39**: 227.

Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. 2008. GEIGER:

investigating evolutionary radiations. *Bioinformatics* **24**: 129–131.

Hawkins, B.A., Diniz-Filho, J.A.F., Jaramillo, C.A. & Soeller, S.A. 2006. Post-Eocene climate change, niche conservatism, and the latitudinal diversity gradient of New World birds. *J. Biogeogr.* **33**: 770–780.

Helfman, G.S., Collette, B.B., Facey, D.E. & Bowen, B.W. 2009. *The diversity of fishes: Biology, Evolution, and Ecology*. Wiley-Blackwell, West Sussex.

Hickman, C., Roberts, L., Keen, S., Larson, A. & Elsenhour, D. 2012. *Animal diversity*. McGraw-Hill, New York.

Hortal, J., de Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M. & Ladle, R.J. 2015. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**: 523–549.

Hubert, N. & Renno, J.-F. 2006. Historical biogeography of South American freshwater fishes. *J. Biogeogr.* **33**: 1414–1436.

Joy, J.B., Liang, R.H., McCloskey, R.M., Nguyen, T. & Poon, A.F.Y. 2016. Ancestral Reconstruction. *PLOS Comput. Biol.* **12**: e1004763.

Lauder, G. V. 1982. Patterns of Evolution in the Feeding Mechanism of Actinopterygian Fishes. *Am. Zool.* **22**: 275–285.

Lazzaro, X. 1987. A review of planktivorous fishes: Their evolution, feeding behaviours, selectivities, and impacts. *Hydrobiologia* **146**: 97–167.

Ley, R.E., Lozupone, C.A., Hamady, M., Knight, R. & Gordon, J.I. 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6**: 776–788.

Litsios, G., Pellissier, L., Forest, F., Lexer, C., Pearman, P.B., Zimmermann, N.E., *et al.* 2012. Trophic specialization influences the rate of environmental niche evolution in damselfishes (Pomacentridae). *Proc. R. Soc. B Biol. Sci.* **279**: 3662–3669.

Losos, J.B. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* **11**: 995–1003.

Lujan, N.K., Winemiller, K.O. & Armbruster, J.W. 2012. Trophic diversity in the evolution and community assembly of loricariid catfishes. *BMC Evol. Biol.* **12**: 1–12.

Manlick, P.J., Petersen, S.M., Moriarty, K.M. & Pauli, J.N. 2019. Stable isotopes reveal limited Eltonian niche conservatism across carnivore populations. *Funct. Ecol.* **33**: 335–345.

Moyle, P.B. & Cech-Jr., J.J. 2004. *Fishes: an introduction to ichthyology*. Prentice Hall, New Jersey.

Nelson, J.S. 2016. *Fishes of the World*. John Wiley & Sons, Nova Jersey.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al.* 2019. Vegan: Community Ecology Package.

Olalla-Tárraga, M.Á., González-Suárez, M., Bernardo-Madrid, R., Revilla, E. & Villalobos, F. 2017. Contrasting evidence of phylogenetic trophic niche conservatism in mammals worldwide. *J. Biogeogr.* **44**: 99–110.

Olsen, A.M. 2017. Feeding ecology is the primary driver of beak shape diversification in waterfowl. *Funct. Ecol.* **31**: 1985–1995.

Orme, C.D.L., Freckleton, R.P., Thomas, G.H., Petzoldt, T., Fritz, S.A., Isaac, N.J.B., *et al.* 2012. caper: Comparative Analyses of Phylogenetics and Evolution in R. *Methods Ecol. Evol.* **3**: 145–151.

Paradis, E. & Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Peterson, A.T. 2011. Ecological niche conservatism: a time-structured review of evidence. *J. Biogeogr.* **38**: 817–827.

Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. 1999. Conservatism of Ecological Niches in Evolutionary Time. *Science (80-.).* **285**: 1265–1267.

Pianka, E.R. 1973. The Structure of Lizard Communities. *Annu. Rev. Ecol. Syst.* **4**: 53–74.

Pouilly, M., Yunoki, T., Rosales, C. & Torres, L. 2004. Trophic structure of fish assemblages from Mamoré River floodplain lakes (Bolivia). *Ecol. Freshw. Fish* **13**: 245–257.

Price, S.A., Hopkins, S.S.B., Smith, K.K. & Roth, V.L. 2012. Tempo of trophic evolution and its impact on mammalian diversification. *Proc. Natl. Acad. Sci.* **109**: 7008–7012.

Pyron, R.A., Costa, G.C., Patten, M.A. & Burbrink, F.T. 2015. Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biol. Rev.* **90**: 1248–1262.

Rabosky, D.L., Chang, J., Title, P.O., Cowman, P.F., Sallan, L., Friedman, M., *et al.* 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**: 392–395.

Revell, L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**: 217–223.

Rocha, M., Rossoni, F., Akama, A. & Zuanon, J. 2019. A new species of spiny driftwood catfish Spinipterus (Siluriformes: Auchenipteridae) from the Amazon basin. *J. Fish Biol.* jfb.14211.

Rodriguez, M.A., Richardson, S.E. & Lewis, W.M. 1990. Nocturnal Behavior and Aspects of the Ecology of a Driftwood Catfish, Entomocorus gameroi (Auchenipteridae). *Biotropica* **22**: 435–438.

Román-Palacios, C., Scholl, J.P. & Wiens, J.J. 2019. Evolution of diet across the animal tree of life. *Evol. Lett.* **3**: 339–347.

Rosado, B.H.P., Figueiredo, M.S.L., de Mattos, E.A. & Grelle, C.E. V. 2016. Eltonian shortfall due to the Grinnellian view: functional ecology between the mismatch of niche concepts. *Ecography (Cop.).* **39**: 1034–1041.

Royero, R. 1993. Notas sobre la capacidad de alimentacion del bagre sapo (Asterophysus batrachus) , (Siluriformes - Auchenipteridae) y algunos datos sobre su distribucion en Venezuela. *Acta Biol. Venez.* **14**: 69–73.

Sanchez, J.L. & Trexler, J.C. 2016. The adaptive evolution of herbivory in freshwater systems. *Ecosphere* **7**: e01414.

Sanderson, S.L., Roberts, E., Lineburg, J. & Brooks, H. 2016. Fish mouths as engineering structures for vortical cross-step filtration. *Nat. Commun.* **7**: 11092.

Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proc. R. Soc. B Biol. Sci.* **273**: 1987–1998.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* **10**: 1115–1123.

Svanbänk, R. & Persson, L. 2004. Individual diet specialization, niche width and population dynamics: implications for trophic polymorphisms. *J. Anim. Ecol.* **73**: 973–982.

Vincent, S.E., Brandley, M.C., Herrel, A. & Alfaro, M.E. 2009. Convergence in trophic morphology and feeding performance among piscivorous natricine snakes. *J. Evol. Biol.* **22**: 1203–1211.

Webb, C.O., Ackerly, D.D., McPeek, M.A. & Donoghue, M.J. 2002. Phylogenies and Community Ecology. *Annu. Rev. Ecol. Syst.* **33**: 475–505.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H. V., *et al.* 2010. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* **13**: 1310–1324.

Wiens, J.J. & Graham, C.H. 2005. Niche conservatism: Integrating Evolution, Ecology, and Conservation Biolog. *Annu. Rev. ofEcology, Evol. Syst.* **36**: 519–533.

Wiens, J.J., Lapoint, R.T. & Whiteman, N.K. 2015. Herbivory increases diversification across insect clades. *Nat. Commun.* **6**: 8370.

Winemiller, K.O. & Adite, A. 1997. Convergent evolution of weakly electric fishes from floodplain habitats in Africa and South America. *Environ. Biol. Fishes* **49**: 175–186.

Wootton, R.J. 1998. *The ecology of teleost fishes*, 2nd ed. Springer Netherlands, Dordrecht.

Zelditch, M.L., Ye, J., Mitchell, J.S. & Swiderski, D.L. 2017. Rare ecomorphological convergence on a complex adaptive landscape: Body size and diet mediate evolution of jaw shape in squirrels (Sciuridae). *Evolution (N. Y).* **71**: 633–649.

5. CONCLUSÃO GERAL

Os resultados obtidos nesta Tese apresentam contribuições sobre aspectos ecológicos e evolutivos da família Auchenipteridae. Apesar dos recentes avanços nas descrições e revisão dos táxons, a atual riqueza de espécies da família representa apenas uma fração do que ainda pode ser descrito. Além do déficit sobre as espécies que ainda não ganharam descrições formais, é possível observar um viés no uso de uma identificação taxonômica atualizada nas coleções científicas (Seção 1). Essa obliquidade nomenclatural alcança os repositórios online, dados esses globalmente utilizados em estudos macroecológicos. O uso de dados potencialmente enviesados pode levar os pesquisadores a fornecer uma perspectiva incompleta ou até equivocada das variações do mundo natural. Desta forma, é imprescindível um maior aporte em investimentos em taxonomia, seja financiando projetos de revisões de táxons, seja na melhor capacitação de pessoal.

Além da lacuna acerca da riqueza de espécies, quantificamos algumas carências de dados sobre esse importante grupo de peixes Neotropicais (Seção 2). Observamos um notável nível de sub-amostragem geográfica para essa família, indicando que maiores esforços de coleta são necessários para melhor compreensão da distribuição das espécies. Já o conhecimento sobre a história evolutiva da família Auchenipteridae apresentou lacunas pouco pronunciadas. O bom nível de conhecimento acumulado sobre essa temática tem produzido conclusões cada vez mais robustas sobre as relações de parentesco entre as linhagens. Por outro lado, observamos um déficit na quantidade de informações sobre a dieta das espécies, visto que inéditas combinações de hábitos alimentares ainda estão surgindo em publicações científicas.

Por fim, com os atuais dados sobre a dieta dos auchenipterídeos, é possível apontar que a maioria dos hábitos alimentares das espécies é conservada ao longo do tempo evolutivo (Seção 3). Isso sugere que ao longo da evolução dos clados, os hábitos alimentares presente no ancestral tendem a ser mantidos ao longo da diversificação. Ou seja, as linhagens mais proximamente relacionadas tentem a ter dietas mais similares entre si do que com táxons mais distantes. Esse conhecimento é essencial para entender o processo de evolução dos caracteres e diversificação ecológica, lançando uma nova luz sobre a evolução dos peixes auchenipterídeos.

Apesar de certa independência entre as seções aqui desenvolvidas, ponderamos de maneira geral: ainda há muito trabalho a ser feito para esgotar (se possível for) as lacunas no conhecimento sobre esses peixes. Para tal nível de conhecimento ser alcançado precisaremos da conjunta colaboração de ecólogos, zoólogos taxonomistas, biogeógrafos, geneticistas, entre outros.

6. ARTIGOS PUBLICADOS DURANTE O PERÍODO DO DOUTORADO

1. **FREITAS, TIAGO M. S.**; DE PAULA, ANA THAIS T.; LEÃO, HÍNGARA; BENONE, NARAIANA L.; MONTAG, LUCIANO F. A. Length-weight relationships of 11 fish species from streams of Anapu River Basin, State of Pará, eastern Amazon, Brazil. *Journal of Applied Ichthyology*, v. 1, p. 1-3, 2019.

2. **FREITAS, T. M. S.**; MONTAG, L. F. A.Population and reproductive parameters of the red-tailed catfish, *Phractocephalus hemioliopterus* (Pimelodidae: Siluriformes), from the Xingu River, Brazil. *Neotropical Ichthyology*, v. 17, p. e190015, 2019.

3. SALVADOR, G. N.; MAIA, C.; DUTRA, G. M.; PESSALI, T. C.; BENONE, N. L.; MONTAG, L. F. A.; **FREITAS, T. M. S.** Length-weight relationships of eight fish species from Guyana coastal drainages, Amapá, Brazil. *Journal of Applied Ichthyology*, v. 35, p. 1, 2019.

4. SALVADOR, G. N.; FREDERICO, R. G.; PESSALI, T. C.; VIEIRA, F.; **FREITAS, T. M. S.** Length-weight relationship of 21 fish species from Rio Doce River basin, Minas Gerais, Brazil. *Journal of Applied Ichthyology*, v. 1, p. 1, 2018.

5. AYRES-SANTOS, B.; **FREITAS, T. M. S.**; MONTAG, L. F. A. Influence of flood pulse on ecological aspects of an iliophagous fish from Eastern Amazonia, Brazil. *Boletim do Museu Paraense Emilio Goeldi. Ciências Naturais*, v. 13, p. 205, 2018.

6. **FREITAS, T. M. S.**; ALMEIDA, V. H. C.; MONTAG, L. F. A. Evidence of seed dispersal of *Virola surinamensis* (Myristicaceae) by a catfish in Eastern Amazonia, Brazil. *Boletim do Museu Paraense Emilio Goeldi. Ciências Naturais*, v. 13, p. 213, 2018.

7. **FREITAS, T. M. S.**; PRUDENTE, B. S.; FREITAS, D. T. H.; BENONE, N. L.; LEAO, H.; DUTRA, G. M.; MONTAG, L. F. A. Fishes of Caxiuanã National Forest: 20 years (1993 to 2012) of sampling in a protected area in the Eastern Amazon. *Boletim do Museu Paraense Emilio Goeldi. Ciências Naturais*, v. 13, p. 185, 2018.

8. **FREITAS, T. M. S.**; SOUZA, J. B. S.; PRUDENTE, B. S.; MONTAG, L. F. A. Length-weight relationship in ten fish species from the Nhamundá River, the Amazon Basin, Brazil. *Acta Amazonica*, v. 47, p. 75-78, 2017.

9. **FREITAS, T. M. S.**; MONTAG, L. F. A.; BARTHEM, R. B. Distribution, feeding and ecomorphology of four species of Auchenipteridae (Teleostei: Siluriformes) in Eastern Amazonia, Brazil. *Iheringia. Série Zoologia*, v. 107, p. 1-6, 2017.

10. **FREITAS, T. M. S.**; DUTRA, G. M.; SALVADOR, G. N. Length-weight relationships of 18 fish species from Paraíba do Sul basin, Minas Gerais, Brazil. *Journal of Applied Ichthyology*, v. 33, p. 1-3, 2017.

11. HASHIGUTI, D. T. F.; BEGOT, T. O.; PRUDENTE, B. S.; **FREITAS, T. M. S.**; MONTAG, L. F. A. Length-weight relationships of six fish species from Tapajós River, Eastern Amazon, Brazil. *Journal of Applied Ichthyology*, v. 33, p. e13453, 2017.

12. SALVADOR, G. N.; PESSALI, T. C.; ANDRADE, A. F. A.; **FREITAS, T. M. S.** Length-weight relationships of seven fish species from Jequitinhonha River basin, southeastern Brazil. *Journal of Applied Ichthyology*, v. 33, p. 1-3, 2017.

13. **FREITAS, T.M.S.**; ALMEIDA, V. H. C.; MONTAG, L. F. A.; FONTOURA, N.F. Predicting size at first sexual maturity from length/weight relationship: a case study with an Amazonian catfish. *Neotropical Ichthyology*, v. 14, p. 00-00, 2016.

ARTIGOS ACEITO PARA PUBLICAÇÃO

1. **FREITAS, T.M.S.**; MONTAG, L. F. A.; DE MARCO, P.; HORTAL, J. How reliable are species identifications in biodiversity big data? Evaluating the records of a Neotropical fish family in online repositories. *Systematics and Biodiversity*. *in press* 2020.

7. ANEXOS

SESSÃO 2 – R SCRIPT COMMANDS TO EVALUATE THE LINNEAN SHORTFALL

```
##################################
##### Loading packages and data #####
##################################

library(vegan)
data <- read.table("Matrix_auchenipterid.txt",header=T) # input of binary data

######################################
##### Calculating Jackknife2 ('jack2') #####
######################################

jackk<-specpool(data)

n0 <- 123
#current number of valid species

K1 <- 152.1 + 70
#'jack2' value + number of species already described before the dataset

#######################################################
##### Function to calculate the 'population' growth model #####
#######################################################

cresc.log=function(N0=n0, r=0.01, K=K1, tseq=1:500)
{
  resulta=K/(1+((K-N0)/N0)*exp(-r*tseq))
  return(resulta)
}

#r = intrinsic rate of increase;  r will vary manually for simulation of growth curves
#K = carrying capacity; or 'jack2'+ N before the data (=3166.6 in our data)
#tseq = time limit for population growth

##################################
##### Running the logistic model #####
##################################

cresc.log(N0=n0, r=0.01, K=K1, tseq=1:500) # intevalo de 500 anos

#manual simulation for 10 populational curves by increasing r
pop1J<-cresc.log(N0=n0, r=0.05, K=K1, tseq=1:500)
pop2J<-cresc.log(N0=n0, r=0.1, K=K1, tseq=1:500)
pop3J<-cresc.log(N0=n0, r=0.15, K=K1, tseq=1:500)
pop4J<-cresc.log(N0=n0, r=0.2, K=K1, tseq=1:500)
pop5J<-cresc.log(N0=n0, r=0.25, K=K1, tseq=1:500)
pop6J<-cresc.log(N0=n0, r=0.3, K=K1, tseq=1:500)
pop7J<-cresc.log(N0=n0, r=0.35, K=K1, tseq=1:500)
pop8J<-cresc.log(N0=n0, r=0.4, K=K1, tseq=1:500)
pop9J<-cresc.log(N0=n0, r=0.45, K=K1, tseq=1:500)
pop10J<-cresc.log(N0=n0, r=0.5, K=K1, tseq=1:500)

estimation_results<-data.frame(pop1J,pop2J,pop3J,pop4J,pop5J,pop6J,pop7J,pop8J,pop9J,pop10J)

write.table(simulation_Jack2, file = 'estimation_results.txt', row.names = F, col.names = T, sep="\t")
```

SESSÃO 2 – R SCRIPT COMMANDS TO EVALUATE THE WALLACEAN SHORTFALL

```
###############################
##### Inventory completeness #####
########## Ecoregions ###########
###############################

# Loading packages and the dataset of species occurrence
library(extrafont)
library(ggplot2)
library(grid)
library(raster)
library(rgdal)
library(rlang)
library(spaa)
library(sqldf)
library(vegan)
fonts()

rm(list = ls())
setwd("xxxx")

# Loading file with occurance data
df_occ<-read.table("Over_Ecoregions.txt", sep=";",h=T)
head(df_occ)

# Loading shapefile
ECO_shp <- readOGR(dsn="xxxx")

proj4string(ECO_shp)
plot(ECO_shp)

#Change the projection into Albers equal area
ECO_shp_proj <- spTransform(ECO_shp , CRS("+proj=aea +lat_1=-5 +lat_2=-42 +lat_0=-32 +lon_0=-60 +x_0=0 +y_0=0 +ellps=aust_SA +units=m +no_defs"))
proj4string(ECO_shp_proj)
plot(ECO_shp_proj)

# Overlay points with polygons
lon<-df_occ[,c("Longitude")]
lat<-df_occ[,c("Latitude")]
coords<-SpatialPoints(cbind(lon,lat))
df1_SPpoints <- SpatialPointsDataFrame(coords, df_occ)
class(df1_SPpoints)
proj4string(df1_SPpoints) = CRS("+proj=longlat +datum=WGS84 +no_defs")
df1_SPpoints_proj <- spTransform(df1_SPpoints , CRS("+proj=aea +lat_1=-5 +lat_2=-42 +lat_0=-32 +lon_0=-60 +x_0=0 +y_0=0 +ellps=aust_SA +units=m +no_defs"))

proj4string(df1_SPpoints_proj) <- proj4string(ECO_shp_proj) #define that SpatialPointsDataFrame and shapefile have the same projection

names(ECO_shp@data) #check fields in the attribute table of the shape file.

df1_SPpoints_proj$ECO_Name<- over(df1_SPpoints_proj, ECO_shp_proj)$Name #add the column "Name" from the shape file to the SpatiaPointsDataFrame
names(df1_SPpoints_proj@data) #check the attribute table of the SpatialPointsDataFrame.

head(df1_SPpoints_proj@data)
```

```
#######################################################
##### Calculate inventory completeness for each ecoregion #####
#######################################################

df_occ2<-as.data.frame(df1_SPpoints_proj) #convert the SpatialPointsDataFrame to data.frame
class(df_occ2)

nrow(df_occ2) #check whether each records as a unique ID
length(unique(df_occ2$IND))

df_occ2$id<-row.names(df_occ2) #create an unique ID for each record
df_occ2$abundance<-1 #create a column called "abundance" filled with "1"
df_occ2$abundance<-as.integer(df_occ2$abundance)

sum(is.na(df_occ2$Ecoregion)) #check whether there is a record with value NA in the field ECO_name

eco = unique(df_occ2$Ecoregion) #a vector with unique names of ecoregions
res1 = list() #a list that we will store our Site x Species matrix (as requered by vegan)
spaccum = list() #a list to store the results of the species accumulation curve
slope = list() #a list to store the slopes
names(df_occ2)

for (i in 1:length(eco)) {
  dat2 = df_occ2[df_occ2$Ecoregion == eco[i], c("id", "Species", "abundance")]
  res1[[i]] = data2mat(dat2)
  spaccum[[i]] = specaccum (res1[[i]], method = "exact")
  slope[[i]]<-specslope(spaccum[[i]], length(spaccum[[i]][[4]])-1)
}

ss<-as.data.frame(unlist(slope)) #convert list of slope values to data.frame
head(ss)
cc<-as.data.frame(eco) #convert vector of unique names of Ecoregions to data.frame
head(cc)

slope_eco<-cbind(cc,ss) #bind these two objects
colnames(slope_eco)[2]<-"Slope" #rename columns
colnames(slope_eco)[1]<-"Ecoregion" #rename columns
head(slope_eco)
min(slope_eco$Slope)
max(slope_eco$Slope)
hist(slope_eco$Slope)
newdata <- slope_eco[order(-slope_eco$Slope),]
head(newdata)

# Format the field "slope"
slope_eco$Slope_round<-as.numeric(format(round(slope_eco$Slope, 3)))
head(slope_eco)
str(slope_eco)

slope_eco$Completeness<-1-slope_eco$Slope_round # Calculate completeness
#values closer to 1 indicate well-sampled areas
hist(slope_eco$Completeness)

# Ecoregion “Fluminense” showed slope higher than 1. This is probably an error. So, we decided to give
to this ecoregion the slope value equal to 0
slope_eco$Completeness<-ifelse(slope_eco$Slope_round > 1, 0, slope_eco$Completeness)
hist(slope_eco$Completeness)
head(slope_eco)
slope_eco$Completeness_round<-as.numeric(format(round(slope_eco$Completeness, 3)))
```

```
###########################
##### Building the maps ######
###########################

# Calculate the number of records and the number of species of each ecoregion
N_rec_sp<-sqldf("SELECT Ecoregion, COUNT(Species) AS N_rec,
              COUNT(DISTINCT(Species)) AS N_sp
              FROM df_occ2
              GROUP BY Ecoregion")
head(N_rec_sp)

# Merged the object obtained in item above with the Slopes_Ecoregions
nrow(N_rec_sp)
nrow(slope_eco)
head(N_rec_sp)
head(slope_eco)

Ecoregions_comp<-merge(N_rec_sp, slope_eco, by = "Ecoregion")
head(Ecoregions_comp)
class(Ecoregions_comp)

# Defining well-sampled ecoregions
Ecoregions_comp$well_sampled<-ifelse(Ecoregions_comp$Slope <=0.01 & Ecoregions_comp$N_rec
>=50, 1, 0)
well_sampled<-subset(Ecoregions_comp, Ecoregions_comp$well_sampled ==1)
head(Ecoregions_comp)
sum(Ecoregions_comp$well_sampled) # eight ecoregions with slope value lower or equal to 0.01

Ecoregions_comp$R_S_ratio<-Ecoregions_comp$N_rec/Ecoregions_comp$N_sp
head(Ecoregions_comp)
plot(Ecoregions_comp$Slope_round, Ecoregions_comp$R_S_ratio)
hist(Ecoregions_comp$N_rec)

write.csv(Ecoregions_comp, "Ecoregions_comp.txt") # Arquivo com valores de completeness

Ecoregions_comp<-read.table("Ecoregions_comp.txt", header = TRUE, sep = ",")
head(Ecoregions_comp)

#########################
##### Building graphs #####
#########################

names(Ecoregions_comp)

# Select cells with reliable values of completeness (slope)

names(Ecoregions_comp)
Ecoregions_comp$slope_ok<-ifelse(Ecoregions_comp$N_rec <=15 &
!is.na(Ecoregions_comp$Slope_round), 1, 0)

Ecoregions_comp$well_sampled<-ifelse(Ecoregions_comp$Slope_round <=0.01 &
                                  Ecoregions_comp$N_rec >= 20 &
                                   !is.na(Ecoregions_comp$Slope_round), 1, 0)

sum(Ecoregions_comp$well_sampled)
Ecoregions_comp$slope_category<-as.factor(ifelse(Ecoregions_comp$slope_ok == 1, 1,
                                         ifelse(Ecoregions_comp$well_sampled == 1 , 2, 3)))

max(Ecoregions_comp$N_rec)
```

```
p <- ggplot(Ecoregions_comp, aes(N_rec, Slope_round, color=slope_category))+
  #geom_point(alpha = 1, size = 1.5)+
  geom_point(aes(colour = factor(slope_category)), alpha = 1 , size = 1.8, shape = 19)+
  geom_vline(xintercept = 15, colour = "darkred", size = 0.5, linetype="F1")+
  #geom_rect(aes(xmin = 0, xmax = 50, ymin = -Inf, ymax = Inf), fill = "red", alpha = 0.5)+
  #scale_color_manual(values=cols)+
  scale_colour_manual(values = c("grey70", "red", "black"))+
  scale_x_continuous(name="Number of records", limits=c(0, 1200))+
  scale_y_continuous(name="Slope of SACs", limits=c(0, 1))+
  theme(legend.title=element_blank())+
  theme(panel.grid.minor = element_blank())+
  theme(panel.grid.major = element_blank())+
  #theme(axis.title.y=element_blank())+
  #theme_bw()+
  #theme(text=element_text(size=16,  family="Times New Roman", color = "black"))+
  #theme(panel.background=element_rect("white"))+
  theme(panel.background=element_rect(fill = "white", colour = "black", size = 0.5))+
  theme(axis.text.x = element_text(color="black",size=12))+
  theme(axis.text.y = element_text(color="black",size=12))+
  theme(axis.title.x = element_text(color="black", size=12))+
  theme(axis.title.y = element_text(color="black", size=12))+
  #theme(panel.grid.major = element_line(colour = "grey90"))+
  #theme(panel.grid.major.y = element_blank())+
  #theme(panel.grid.minor.y = element_blank())+
  theme(legend.position="none")+
  theme(plot.margin=unit(c(1,1,1,1),"cm"))
p

p1<-p + annotate("rect", xmin=c(-Inf), xmax=c(50), ymin=c(-Inf) ,
          ymax=c(Inf), alpha=0.2, color="grey60", fill="grey50")

p1
p2<-p1 + annotate("rect", xmin=c(-Inf), xmax=c(10), ymin=c(-Inf) ,
            ymax=c(Inf), alpha=0.2, fill="grey10")
p2

hist(Ecoregions_comp$Completeness)
hist(Ecoregions_comp$Slope)
newdata <- Ecoregions_comp[order(-Ecoregions_comp$Slope),]
head(newdata)

# Merged the object above with the shapefiles of the ecoregions
shp <- readOGR(file.choose()) # loading the shapefiles of the ecoregions
head(shp@data)

Ecoregions_comp<-merge(x= shp, y = Ecoregions_comp, by.x ="Name", by.y = "Ecoregion")
head(Ecoregions_comp@data)

colnames(Ecoregions_comp@data)[5]<-"Slope_1" #rename columns
colnames(Ecoregions_comp@data)[8]<-"Slope_2" #rename columns

Ecoregions_comp@data$N_rec[is.na(Ecoregions_comp@data$N_rec)] <- 0
Ecoregions_comp@data$N_sp[is.na(Ecoregions_comp@data$N_sp)] <- 0
Ecoregions_comp@data$well_sampled[is.na(Ecoregions_comp@data$well_sampled)] <- 0

writeOGR(obj=Ecoregions_comp, dsn = getwd(), layer="Ecoregions_comp", driver="ESRI Shapefile",
overwrite_layer = TRUE)
```

```
# The shapefile Ecoregions_comp shows the number of records and number os species for all ecoregions

getwd()
class(Ecoregions_comp)

Ecoregions_Nrec_50 <- Ecoregions_comp[Ecoregions_comp$N_rec >=50,]

writeOGR(obj=Ecoregions_Nrec_50, dsn = getwd(), layer="Ecoregions_Nrec_50", driver="ESRI
Shapefile", overwrite_layer = TRUE)
# The shapefile Ecoregions_Nrec_50 shows only the ecoregions with more than 50 records

# Get stats for legend
df<-as.data.frame(Ecoregions_comp)
head(df)
nrow(df)

shp <- readOGR(file.choose()) # loading the shapefiles of the ecoregions
head(shp@data)
df<-as.data.frame(shp)
dat <- transform(df, bin = ifelse(Records == 0,"Rec 0",
                              ifelse((Records >=1 &  Records <=10),"1Rec 1 - 10",
                              ifelse((Records >=11 & Records <= 50), "2Rec 1 - 40",
                              ifelse((Records >=51 & Records <=70), "3Rec 41 - 70",
                              ifelse((Records >=71 & Records <=150), "4Rec 71 - 150",
                              ifelse((Records >=151 & Records <=250), "5Rec 151 - 250",
                              ifelse((Records >=251 & Records <=350), "6Rec 251 - 350",
                              ifelse((Records >= 351 & Records <=700), "7Rec 351 - 700",
                              ifelse((Records >=701 & Records <=1170), "8Rec 701 - 1170", NA))))))))))

head(dat)
dat2<-as.data.frame(table(dat$bin))
dat2
dat <- transform(df, bin2 = ifelse(Records == 0,9,
                        ifelse((Records >=1 &  Records <=10),8,
                        ifelse((Records >=11 & Records <= 40), 7,
                        ifelse((Records >=41 & Records <=70), 6,
                        ifelse((Records >=71 & Records <=150), 5,
                        ifelse((Records >=151 & Records <=250), 4,
                        ifelse((Records >=251 & Records <=350), 3,
                        ifelse((Records >= 351 & Records <=700), 2,
                        ifelse((Records >=701 & Records <=1170), 1, NA))))))))))

dat3<-table(dat$bin2)
barplot(dat3, space=c(0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1), horiz = TRUE)

head(df)
dat <- transform(df, bin3 = ifelse(Species == 0,9,
                        ifelse((Species >=1 &  Species <=5),8,
                        ifelse((Species >=6 & Species <= 10), 7,
                        ifelse((Species >=11 & Species <=15), 6,
                        ifelse((Species >=16 & Species <=20), 5,
                        ifelse((Species >=21 & Species <=25), 4,
                        ifelse((Species >=26 & Species <=30), 3,
                        ifelse((Species >= 31 & Species <=45), 2,
                        ifelse((Species >=46 & Species <=55), 1, NA))))))))))

dat4<-table(dat$bin3)
hist(df$Species)
df<-as.data.frame(shp)
barplot(dat4, space=c(0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1, 0.1), horiz = TRUE)
```

```
head(df)
df_50rec<-subset(df, df$Records >=50)
dat <- transform(df_50rec, bin4 = ifelse((Slp_rnd >=0 &  Slp_rnd <=0.01),1,
            ifelse((Slp_rnd >=0.011 & Slp_rnd <= 0.02), 2,
            ifelse((Slp_rnd >=0.021 & Slp_rnd <=0.03), 3,
            ifelse((Slp_rnd >=0.031 & Slp_rnd <=0.04), 4,
            ifelse((Slp_rnd >=0.041 & Slp_rnd <=0.05), 5,
            ifelse((Slp_rnd >=0.051 & Slp_rnd <=0.06), 6,
            ifelse((Slp_rnd >= 0.061 & Slp_rnd <=0.07), 7,
            ifelse((Slp_rnd >=0.071 & Slp_rnd <=0.16), 8, NA))))))))

head(dat)
dat5<-table(dat$bin4)
dat5
hist(df$Species)
barplot(dat5, space=c(0.1,0.1,0.1,0.1,0.1,0.1,0.1,0.1), horiz = TRUE, xlim=c(0,12))

head(df)

################################
##### Inventory completeness #####
############ Grids #############
################################

setwd("xxxxx")

library(extrafont)
library(ggplot2)
library(KnowBR)
library(rgeos)
library(rlang)
require(maptools)
require(PBSmapping)
require(rgdal)
fonts()

# Loading the file with occurance data
df_occ<-read.table("Over_Ecoregions.txt", sep=";",h=T)
head(df_occ)

# Loaging the shapefile
GRID_shp <- readOGR(dsn="C:/Users/juliana.stropp/Dropbox (UFAL-
ECO)/R_scripts/Tiago_R/Dados_July_2018/Shapes South America/Grids",layer="Grids
Auchenipteridae_JS")
proj4string(GRID_shp)
proj4string(GRID_shp) = CRS("+proj=aea +lat_1=-5 +lat_2=-42 +lat_0=-32 +lon_0=-60 +x_0=0
+y_0=0 +ellps=aust_SA +units=m +no_defs")
GRID_shp_proj <- spTransform(GRID_shp , CRS("+proj=longlat +datum=WGS84 +no_defs"))

plot(GRID_shp_proj)

# Getting the centroid for each cell
grid_centr<-gCentroid(GRID_shp_proj,byid=TRUE)
class(grid_centr)
plot(grid_centr)

GRID_with_ctr<-cbind(GRID_shp_proj, grid_centr@coords)
class(GRID_with_ctr)
head(GRID_with_ctr@data)
```

```
plot(GRID_with_ctr)

# Overlay occurrence data with GRID shapefile

# Convert the occurrence data into "SpatialPointsDataFrame" ###
lon<-df_occ[,c("Longitude")]
lat<-df_occ[,c("Latitude")]
coords<-SpatialPoints(cbind(lon,lat))
df1_SPpoints <- SpatialPointsDataFrame(coords, df_occ)
class(df1_SPpoints)
proj4string(df1_SPpoints) = CRS("+proj=longlat +datum=WGS84 +no_defs")

# Check if each cell has a unique ID
names(GRID_with_ctr@data)
head(GRID_with_ctr@data)
nrow(GRID_with_ctr@data)
length(unique((GRID_shp@data$id)))

# Defining the same projection for SpatialPointsDataFrame and shapefile
proj4string(df1_SPpoints) <- proj4string(GRID_with_ctr)

# Add the column id (cells) into the occurrence data
df1_SPpoints$cell_id<- over(df1_SPpoints, GRID_with_ctr)$id

# Add the column latitude and longitude of the centroid into the occurrence data
df1_SPpoints$x<- over(df1_SPpoints, GRID_with_ctr)$x
df1_SPpoints$y<- over(df1_SPpoints, GRID_with_ctr)$y
proj4string(df1_SPpoints)
head(df1_SPpoints@data)

###############################################
##### Calculating completeness with KnowBR #####
###############################################

head(df1_SPpoints@data)

# Converte df1_SPpoints_proj to data frame
df_occ2<-as.data.frame(df1_SPpoints)

records<-data.frame(df_occ2$Species,df_occ2$x, df_occ2$y)
names(records)<-c("Species","Longitude","Latitude")
head(records)
records$Species<-as.character(records$Species)
records$Counts<-1
str(records)
head(records)
data(adworld)

plot(records$Longitude, records$Latitude)

KnowB(data=records, cell=45, estimator = 0)

# Importing the results of KnowBR to the R
estimators_KnowB<-read.csv('Estimators.csv', header = TRUE, sep=';', dec = ",")
head(estimators_KnowB)
hist(estimators_KnowB$Slope.exact)
hist(estimators_KnowB$Records)
min(estimators_KnowB$Records)

max(estimators_KnowB$Slope.exact, na.rm = TRUE)
```

```
estimators_KnowB[1:23,]
str(estimators_KnowB)
write.csv(estimators_KnowB, "estimators_KnowB_cell_final.csv", dec = ".", sep=';') # ok

# Merged the results of KnowBR with the grid-shapefile
# Convert the results of KnowBR to SpatialPointsDataframe
head(estimators_KnowB)
head(GRID_with_ctr@data)

estimators_KnowB$cell_coords<-paste(estimators_KnowB$Longitude,estimators_KnowB$Latitude, sep
= "_")
GRID_with_ctr$cell_coords<-paste(GRID_with_ctr$x,GRID_with_ctr$y, sep = "_")

estimators_poly<-merge(GRID_with_ctr, estimators_KnowB, by = 'cell_coords')
class(estimators_poly)

estimators_poly@data$Records[is.na(estimators_poly@data$Records)] <- 0
estimators_poly@data$Observed.richness[is.na(estimators_poly@data$Observed.richness)] <- 0

writeOGR(obj=estimators_poly, dsn = getwd(), layer="Cell_comp3", driver="ESRI Shapefile",
overwrite_layer = TRUE)

###########################
##### Plotting the results #####
###########################

p <- ggplot(estimators_KnowB, aes(Records, Slope.exact))+
  geom_point(alpha = 1, size = 1.5)+
  #geom_point(aes(colour = factor(well_sampled)), alpha = 1, size = 1.5)+
  geom_vline(xintercept = 15, colour = "darkred", size = 0.5, linetype="F1")+
  #geom_rect(aes(xmin = 0, xmax = 50, ymin = -Inf, ymax = Inf), fill = "red", alpha = 0.5)+
  #scale_color_manual(values=cols)+
  scale_colour_manual(values = c("grey70", "black"))+
  scale_x_continuous(name="Number of records", limits=c(0, 250))+
  scale_y_continuous(name="Slope of SACs", limits=c(0, 1))+
  theme(legend.title=element_blank())+
  theme(panel.grid.minor = element_blank())+
  theme(panel.grid.major = element_blank())+
  #theme(axis.title.y=element_blank())+
  theme(panel.background=element_rect(fill = "white", colour = "grey20"))+
  theme(axis.text.x = element_text(color="grey20",size=12))+
  theme(axis.text.y = element_text(color="grey20",size=12))+
  theme(axis.title.x = element_text(color="grey20", size=14))+
  theme(axis.title.y = element_text(color="grey20", size=14))+
  #theme(panel.grid.major = element_line(colour = "grey90"))+
  #theme(panel.grid.major.y = element_blank())+
  #theme(panel.grid.minor.y = element_blank())+
  theme(legend.position="none")
p

# Select cells with reliable completeness (slope) values
estimators_KnowB$slope_ok<-ifelse(estimators_KnowB$Records <=15 &
!is.na(estimators_KnowB$Slope.exact), 1, 0)


estimators_KnowB$well_sampled<-ifelse(estimators_KnowB$Slope.exact <=0.01 &
                                       estimators_KnowB$Records >= 20 &
                                       !is.na(estimators_KnowB$Slope.exact), 1, 0)
```

```
sum(estimators_KnowB$well_sampled)
estimators_KnowB$slope_category<-as.factor(ifelse(estimators_KnowB$slope_ok == 1, 1,
                                   ifelse(estimators_KnowB$well_sampled == 1 , 2, 3)))

unique(estimators_KnowB$slope_category)

write.csv(estimators_KnowB, "estimators_KnowB_cell_final.csv", dec = ".", sep=';')
estimators_KnowB<-read.csv("estimators_KnowB_cell_final.csv", header = TRUE)
head(estimators_KnowB)

#############################
##### Building the graphs #####
#############################

p <- ggplot(estimators_KnowB, aes(Records, Slope.exact, color=slope_category))+
  #geom_point(alpha = 1, size = 1.5)+
  geom_point(aes(colour = factor(slope_category)), alpha = 1 , size = 1.8, shape = 19)+
  geom_vline(xintercept = 15, colour = "darkred", size = 0.5, linetype="F1")+
  #geom_rect(aes(xmin = 0, xmax = 50, ymin = -Inf, ymax = Inf), fill = "red", alpha = 0.5)+
  #scale_color_manual(values=cols)+
  scale_colour_manual(values = c("grey70", "red", "black"))+
  scale_x_continuous(name="Number of records", limits=c(0, 250))+
  scale_y_continuous(name="Slope of SACs", limits=c(0, 1))+
  theme(legend.title=element_blank())+
  theme(panel.grid.minor = element_blank())+
  theme(panel.grid.major = element_blank())+
  #theme(axis.title.y=element_blank())+
  #theme_bw()+
  theme(text=element_text(size=16,  family="Times New Roman", color = "black"))+
  #theme(panel.background=element_rect("white"))+
  theme(panel.background=element_rect(fill = "white", colour = "black", size = 0.5))+
  theme(axis.text.x = element_text(color="black",size=12, family="Times New Roman"))+
  theme(axis.text.y = element_text(color="black",size=12, family="Times New Roman"))+
  #theme(axis.title.x = element_text(color="grey20", size=14, family="Times News Roman"))+
  #theme(axis.title.y = element_text(color="grey20", size=14, family="Times News Roman"))+
  #theme(panel.grid.major = element_line(colour = "grey90"))+
  #theme(panel.grid.major.y = element_blank())+
  #theme(panel.grid.minor.y = element_blank())+
  theme(legend.position="none")
p

getwd()
# exportar arquivo:
head(estimators_KnowB)

write.csv(estimators_KnowB, "estimators_KnowB_cells.csv", sep = ",")
df<-read.csv("estimators_KnowB_cells.csv", sep = ",")
head(df)

#########################################
##### Building histograms for the maps #####
#########################################

shp <- readOGR(file.choose())
df<-as.data.frame(shp)
```

```
dat <- transform(df, bin = ifelse(Records == 0,"9",
                     ifelse((Records >=1 &  Records <=15),"8",
                     ifelse((Records >=16 & Records <= 35), "7",
                     ifelse((Records >=36 & Records <=55), "6",
                     ifelse((Records >=56 & Records <=75), "5",
                     ifelse((Records >=76 & Records <=95), "4",
                     ifelse((Records >=96 & Records <=115), "3",
                     ifelse((Records >= 116 & Records <=155), "2",
                     ifelse((Records >=156 & Records <=197), "1", NA))))))))))

head(dat)
dat2<-table(dat$bin)
dat2
barplot(dat2, horiz = TRUE, xlim = c(0, 1400))
names(df)

dat2 <- transform(df, bin = ifelse(Obsrvd_ == 0,"9",
                     ifelse((Obsrvd_ >=1 &  Obsrvd_ <=6),"8",
                     ifelse((Obsrvd_ >=7 & Obsrvd_ <= 9), "7",
                     ifelse((Obsrvd_ >=10 & Obsrvd_ <=12), "6",
                     ifelse((Obsrvd_ >=13 & Obsrvd_ <=15), "5",
                     ifelse((Obsrvd_ >=16 & Obsrvd_ <=18), "4",
                     ifelse((Obsrvd_ >=18 & Obsrvd_ <=21), "3",
                     ifelse((Obsrvd_ >=21 & Obsrvd_ <=24), "2",
                     ifelse((Obsrvd_ >=25 & Obsrvd_ <=28), "1", NA))))))))))

head(dat)
dat3<-table(dat2$bin)
dat3
barplot(dat3, horiz = TRUE, xlim = c(0, 1400))

head(df)
dat4 <- transform(df, bin = ifelse((Slp_xct >=0 &  Slp_xct <=0.01), "8",
                      ifelse((Slp_xct >0.01 & Slp_xct <= 0.05), "7",
                      ifelse((Slp_xct >0.05 & Slp_xct <=0.09), "6",
                      ifelse((Slp_xct >0.09 & Slp_xct <=0.14), "5",
                      ifelse((Slp_xct >0.14 & Slp_xct <=0.2), "4",
                      ifelse((Slp_xct >0.2 & Slp_xct <=0.29), "3",
                      ifelse((Slp_xct >0.29 & Slp_xct <=0.37), "2",
                      ifelse((Slp_xct >0.37 & Slp_xct <=42), "1", NA)))))))))

head(dat4)
dat5<-table(dat4$bin)
dat5
barplot(dat5, horiz = TRUE, xlim = c(0, 100))
```

SESSÃO 2 – R SCRIPT COMMANDS TO EVALUATE THE DARWINIAN SHORTFALL

```
##################################
##### Loading packages and data #####
##################################

library(ape)
filos = list.files()[1:8]

#Read and save phylogenies in a list
filo_lista = list()
for(i in 1:length(filos)){
  filo_lista[[i]]= read.tree(filos[i])
  print(i)
}
names(filo_lista) = filos

###########################
##### Data management #####
###########################

#Standardize branch lengths (= 1)
for(i in 1:length(filos)){
  filo_lista[[i]] <- compute.brlen(filo_lista[[i]], 1)
  plot(filo_lista[[i]])
}

#Change the tip names
#Some genera had more than one species
#Parts and removes the "_" from the genera names
for(i in 1:length(filos)){
  nam=strsplit(filo_lista[[i]]$tip.label,"_")
  nam=lapply(nam, function(x)x[1])
  nam=unlist(nam)
  filo_lista[[i]]$tip.label=nam
  plot(filo_lista[[i]])
}
#Calculate minimum number of taxa in a pairwise comparison between phylogenies
num_tax<-data.frame(matrix(ncol=8,nrow=8))
rownames(num_tax)=colnames(num_tax)=filos
for(i in 1:length(filos)){
  for(j in 1:length(filos)){
    num_tax[i,j]=sum(unique(filo_lista[[i]]$tip.label)%in%unique(filo_lista[[j]]$tip.label))
  }
}
min(num_tax)

################################################
##### Calculating similarity between phylogenies #####
################################################

simi_filo<-data.frame(matrix(ncol=8,nrow=8))
rownames(simi_filo)=colnames(simi_filo)=filos
rho_list=list()

for(i in 1:length(filos)){
  for(j in 1:length(filos)){
    #Selecting common genera among phylogenies
    nom_fi=unique(filo_lista[[i]]$tip.label)
    nom_fj=unique(filo_lista[[j]]$tip.label)
```

```
  nom_com=nom_fi[nom_fi%in%nom_fj]
  rho_val=c()

  for(n in 1:100){
   nom_sel=sample(nom_com,10)
   pns1=drop.tip(filo_lista[[i]],nom_fi[!nom_fi%in% nom_sel])
   pns2=drop.tip(filo_lista[[j]],nom_fj[!nom_fj%in% nom_sel])
      #Checking for repeated genera and remove
   if(length(pns1$tip.label)>10){
    sps_eli = names(table(pns1$tip.label))[table(pns1$tip.label)>1]
    for(o in 1:length(sps_eli)){
     pns1=drop.tip(pns1,sample(which(pns1$tip.label==sps_eli[o]),(sum(pns1$tip.label==sps_eli[o])-1)))
    }
   }
   if(length(pns2$tip.label)>10){
    sps_eli = names(table(pns2$tip.label))[table(pns2$tip.label)>1]
    for(o in 1:length(sps_eli)){
     pns2=drop.tip(pns2,sample(which(pns2$tip.label==sps_eli[o]),(sum(pns2$tip.label==sps_eli[o])-1)))
    }
   }
      #Repeated genera had branch lengths of 2;
   #Defaults to 1
   pns1=compute.brlen(pns1, 1)
   pns2=compute.brlen(pns2, 1)

   #Nodes distances between species
   dis_p1=cophenetic.phylo(pns1)
   dis_p2=cophenetic.phylo(pns2)
dis_p2=dis_p2[match(rownames(dis_p1),rownames(dis_p2)),match(colnames(dis_p1),colnames(dis_p2))]
   rho_val[n]=cor(as.dist(dis_p1),as.dist(dis_p2),method="spearman")
  }

#Save all correlation values for each pair of phylogenies
  rho_list[[length(rho_list)+1]] = rho_val
  names(rho_list)[length(rho_list)]=paste(names(filo_lista)[i],names(filo_lista)[j])
  simi_filo[i,j]=mean(rho_val)
  print(c(i,j))
 }
}
#Salving results
save(rho_list,file="rho_list_filo_pairs.RData")
resF=data.frame(names=names(rho_list),mean=unlist(lapply(rho_list,mean)),sd=unlist(lapply(rho_list,sd)))
write.csv(resF,"media_sd_filo_pairs.csv")

################################
##### Phylogenies dendrogram #####
################################

dis_filo=1-as.dist(as.matrix(simi_filo))
dendro=hclust(dis_filo,method = "average")
plot(as.dendrogram(dendro))

#Save dendrogram as tiff
tiff("dendrograma.tiff",res = 600,width = 30, height = 20, units = "cm", compression = "lzw")
plot(as.dendrogram(dendro))
dev.off()
```

SESSÃO 3 – R SCRIPT COMMANDS FOR PHYLOGENETIC NICHE CONSERVATISM

Three scripts were built

1. **functions**
2. **data management**
3. **run**

```
# WRITING FUNCTIONS - script “functions”

###########################
##### Phylogenetic signal #####
###########################

merge_sp <- function(i){

  sp <- i[, 1]
  t <- as.matrix(i[, -1])
  if(nrow(t) > 1){
    return(colSums(t))
  }else{
      return(t)
    }
}

run_d <-function(d_data, i, phy, permut){
  dat <- d_data[, c(1, i)]
  colnames(dat) <- c("names.col", "binvar")
  teste <- phylo.d(data = dat,
                   phy = phy,
                   names.col = names.col,
                   binvar = binvar,
                   permut = permut)
}

#########################
##### Pinkas's function #####
#########################

pianka_index <- function(i){
  return(sum(i[1, ] * i[2 ,])/sqrt(sum(i[1,]^2) * sum(i[2, ]^2)))
  }

pairwise_pianka <- function(x){
  scaled_x <- t(apply(x, 1, function(j) j/sum(j)))
  size <- nrow(x)
  pianka_matrix <- matrix(as.double(size^2), nrow = size, ncol = size)

  for(lines in 1:size){
    for(cols in (1:size)[-lines]){

      pianka_matrix[lines, cols] <- pianka_index(scaled_x[c(lines, cols), ])

    }
  }
  diag(pianka_matrix) <- 1
  return(pianka_matrix)
}

run_analyses <- function(trait_discrete, phylo, perm){
```

```
###################################
##### Discrete phylogenetic signal #####
###################################

d_data <- data.frame(row.names(trait_discrete), trait_discrete, stringsAsFactors = FALSE)
colnames(d_data)[1] <- "species"

d_discrete <- lapply(2:ncol(d_data), function(c){

  return(try(run_d(d_data, c, phylo, permut = perm)))
}
)

#################
##### Mantel #####
#################

pianka <- pairwise_pianka(trait_discrete)
pianka_invert <- 1-pianka
phylo_dist_parci <- cophenetic(phylo)
mantel_parci <- phytools::multi.mantel(pianka_invert, phylo_dist_parci, nperm = perm)

return(list(discrete = d_discrete, pianka = mantel_parci))
}

################################
##### Ancestral reconstruction #####
################################

anc_diet <- function(diet, phy){

  equal <- ace(x = diet, phy = phy, type = "discrete", model = "SYM", marginal = FALSE)
  ard <- ace(x = diet, phy = phy, type = "discrete", model = "ARD", marginal = FALSE)
  aic <- c("equal" = AICc(equal, n = length(diet)), "ard" = AICc(ard, n = length(diet)))
  w <- aicw(aic)
  return(list(equal = equal, ard = ard, w = w))
}

AICc <- function(i, n){
  lik <- i$loglik
  k <- length(i$rates)
  return(-2*(lik + k) + ((2*k^2 + 2*k)/(n-k-1)))
}

average_states <- function(i){
  average <- i$ard$lik.anc * i$w["ard", "w"] +
    i$equal$lik.anc * i$w["equal", "w"]
  return(average)
}

average_rates <- function(i){
  average <- i$equal$rates * i$w["equal", "w"] +
    i$ard$rates * i$w["ard", "w"]
  names(average) <- c("1-0", "0-1")
  return(average)
}
```

DATA MANAGEMENT - script “data management”

```
#########################
##### Loading packages #####
#########################

library("ape")
library("phytools")
library("geiger")
library("caper")
library("phylocurve")
library("geiger")
library("MuMIn")
library("vegan")
source("functions.R")

####################################################
##### Reading the valid names for Auchenipteridae fishes #####
####################################################

valid_names <- read.csv("Valid_names_auchenipteridae2019.csv", stringsAsFactors = FALSE)
valid_names[, 2] <- gsub("^ | $", "", valid_names[, 2])
valid_names[, 1] <- gsub("^ | $", "", valid_names[, 1])
discrete <- read.csv("traits_discrete.csv", header = TRUE, stringsAsFactors = FALSE)
discrete[, 1] <- gsub("_" ," ", discrete[, 1])

##################################
##### Reading and editing the tree #####
##################################

bayesiana <- read.nexus("consenso2_Datada.tree")
mrca <- getMRCA(bayesiana, grep("E_|Ag_", bayesiana$tip.label,value = TRUE))
desc <- phytools::getDescendants(bayesiana, mrca)
clade <- bayesiana$tip.label[desc[desc<=length(bayesiana$tip.label)]]
tree <- keep.tip(bayesiana, clade)
ex <- grep("^.*_.*_|Trachelyi_Roraima|Trachelyi_Santarem|Glanidium_RS2|Glanidium_RS1|Tatia_ater|Spinipterus_oncinha|Trachelyi_Tefe|Trachelyi_xingu", tree$tip.label, value = TRUE)
tree$tip.label <- gsub("Auche_thoracathus", "Auche_thoracatus", tree$tip.label)
tree <- drop.tip(tree, ex)

for(i in 1:length(tree$tip.label)){
  tree$tip.label[i] <- valid_names[grep(tree$tip.label[i], valid_names[, 1]), 2]

}

####################################
##### Discrete matrix (feeding habits) #####
####################################

by_sp <- split(discrete, as.factor(discrete[, 1]))
sp_sum <- t(sapply(by_sp, merge_sp))
colnames(sp_sum) <- colnames(discrete[, -1])
discrete_merged <- ifelse(sp_sum >= 1, 1, 0)
discrete_merged <- discrete_merged[row.names(discrete_merged) %in% tree$tip.label, ]
```

```
############################################################
##### Dropping the phylogeny according to the discrete matrix #####
############################################################

tree_drop <- drop.tip(tree, tree$tip.label[!(tree$tip.label %in% row.names(discrete_merged))])

######################
##### Final dataset #####
######################

tree_discrete <- discrete_merged[sapply(tree_drop$tip.label, function(i) grep(i,
row.names(discrete_merged))), ]
```

RUNNING ANALYZES - script “run”

```
source("data_management.R")

#######################################
##### Phylogenetic signal (Statistic D) #####
############# and diet-overlap ##########
#######################################

tree_results <- run_analyses(tree_discrete, tree_drop, perm = 10000)

Result_StatisticD<-data.frame(traits=colnames(tree_discrete),D=NA,P_random=NA, P_Brownian=NA)
for(i in 1:length(tree_results$discrete)){
  Result_StatisticD[i,2] <- tree_results$discrete[[i]]$DEstimate
  Result_StatisticD[i,3] <- tree_results$discrete[[i]]$Pval1
  Result_StatisticD[i,4] <- tree_results$discrete[[i]]$Pval0
}
Result_StatisticD
write.table(Result_StatisticD, file = "Result_StatisticD.txt", sep = "\t")

#Plotting the pruned tree
tiff("Phylogenetic tree.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize=1.1)
dev.off()

############################
##### Mantel correlogram #####
############################

pianka <- pairwise_pianka(tree_discrete)
pianka_invert <- 1-pianka
phy_dist <- cophenetic(tree_drop)
correlogram <- mantel.correlog(D.eco = pianka_invert, D.geo = phy_dist, n.class = 10, cutoff = FALSE)

plot(correlogram)
write.table(correlogram$mantel.res, file = "Result_correlogram.txt", sep = "\t")

###################################
##### Ancestor diet reconstruction #####
###################################

anc_res <- apply(tree_discrete, 2, anc_diet, phy = tree_drop)

#Extracting Likehood
equallik <- sapply(anc_res, function(i) logLik(i$equal))
```

```
ardlik <- sapply(anc_res, function(i) logLik(i$ard))
likehood <- data.frame(equallik,ardlik)

?aicw
?ace

#Extracting delta
delta1 <- anc_res[[1]]$w$delta
delta2 <- anc_res[[2]]$w$delta
delta3 <- anc_res[[3]]$w$delta
delta4 <- anc_res[[4]]$w$delta
delta5 <- anc_res[[5]]$w$delta
delta6 <- anc_res[[6]]$w$delta
delta7 <- anc_res[[7]]$w$delta
delta8 <- anc_res[[8]]$w$delta

delta<-t(cbind(delta1,delta2,delta3,delta4,delta5,delta6,delta7,delta8))
colnames(delta) <- c("delta_equal", "delta_ard")
rownames(delta) <- row.names(likehood)

#ARD rates

rate1 <- anc_res[[1]]$ard$rates
rate2 <- anc_res[[2]]$ard$rates
rate3 <- anc_res[[3]]$ard$rates
rate4 <- anc_res[[4]]$ard$rates
rate5 <- anc_res[[5]]$ard$rates
rate6 <- anc_res[[6]]$ard$rates
rate7 <- anc_res[[7]]$ard$rates
rate8 <- anc_res[[8]]$ard$rates

rates<-t(cbind(rate1,rate2,rate3,rate4,rate5,rate6,rate7,rate8))
colnames(rates) <- c("1-0", "0-1")
rownames(rates) <- row.names(likehood)

#average rates
#node_states <- lapply(anc_res, average_states)
#rate_average <- t(sapply(anc_res, average_rates))


#################
##### Models #####
#################

models <- t(sapply(anc_res, function(i) i$w[, "w"]))
colnames(models) <- c("equal", "ard")

ancestor_results <- cbind(models,likehood,delta)
write.table(ancestor_results, file = "Result_ancestor.txt", sep = "\t")

###############
##### Plot ######
###############

  #Aquatic insects
  tiff("1aquaticinsects.tiff", compression = "lzw",
      width = 2500, height = 3090, res=300)
  plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
  nodelabels(pie = node_states[[1]], piecol=c("white","blue"), cex=0.7)
  dev.off()
```

```
#Terrestrial insects
tiff("2terrestrialinsects.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize=1.1)
nodelabels(pie = node_states[[2]], piecol=c("white","purple"), cex=0.7)
dev.off()

#Crustacean
tiff("3crustacean.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
nodelabels(pie = node_states[[3]], piecol=c("white","orange"), cex=0.7)
dev.off()

#Fish
tiff("4fish.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
nodelabels(pie = node_states[[4]], piecol=c("white","red"), cex=0.7)
dev.off()

#Fruit
tiff("5fruit.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
nodelabels(pie = node_states[[5]], piecol=c("white","darkgreen"), cex=0.7)
dev.off()

#Vegetal matters
tiff("6vegetalmatters.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
nodelabels(pie = node_states[[6]], piecol=c("white","brown"), cex=0.7)
dev.off()

#Plankton
tiff("7plankton.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
nodelabels(pie = node_states[[7]], piecol=c("white","lightgreen"), cex=0.7)
dev.off()

#Detritus
tiff("8detritus.tiff", compression = "lzw",
     width = 2500, height = 3090, res=300)
plotTree(tree_drop, show.tip.label = TRUE, ftype="i", fsize = 1.1)
nodelabels(pie = node_states[[8]], piecol=c("white","gray"), cex=0.7)
dev.off()
```