Universidade de Brasília
Instituto de Ciências Biológicas
Programa de Pós-graduação em Ecologia

# Estimando o tamanho mínimo amostral em estudos para peixes em uma zona de transição Cerrado-Amazônia

Isabela Carolina Silva Vieira Braga

Brasília, DF
2021

Universidade de Brasília
Instituto de Ciências Biológicas
Programa de Pós-graduação em Ecologia

# Estimando o tamanho mínimo amostral em estudos para peixes em uma zona de transição Cerrado-Amazônia

Isabela Carolina Silva Vieira Braga

Dissertação apresentada ao Programa de Pós-graduação em Ecologia do Instituto de Ciências Biológicas da Universidade de Brasília, como requisito parcial para a obtenção do título de Mestre em Ecologia

Orientador: Dr. Murilo Sversut Dias

Brasília, DF

2021

“Cada um sabe a dor e a delícia de ser o que é.”

*Caetano Veloso*

# Agradecimentos

Agradeço ao CNPq pela bolsa de mestrado, ao programa de Pós-graduação em Ecolgia e à UnB pela minha formação. Agradeço também aos membros do Laboratório de Conservação e Ecologia, especificamente ao Dr. Luciano Montag e à Msc. Pâmela Virgolino, pela disponibilização de dados que foram imprescindíveis para que este trabalho fosse executado. Agradeço ao professor Dr. Murilo Sversut Dias, pela paciência e disponibilidade, fundamentais para a construção desta dissertação.

Agradeço imensamente à minha família por ter me encaminhado neste caminho. Agradecimentos especiais à minha mãe e à minha vó, que abdicaram de muitas coisas para que eu pudesse viver as minhas. Vocês sempre acreditaram em mim e eu nunca conseguirei retribuir totalmente o que fizeram por mim, a não ser esperar que eu esteja lhes proporcionando orgulho. Agradeço também aos meus tios, Adriana e Sandro, e aos meus primos, Kamyla e Bruno, pelo companheirismo de anos. Ao meu amor, Guilherme Sena, por ser meu suporte, caminhar ao meu lado e me por pra cima quando tudo parecia cair. À minha amiga Gardenia, que sempre foi solícita e paciente ao escutar meus lamúrios. Aos meus amigos e companheiros de trabalho Aline, Artur, Raiane e Fernanda, que sempre tiveram uma palavra de conforto.

Este trabalho não teria sido realizado sem o envolvimento de várias pessoas. Agradeço a cada um que foi a campo para coleta e que se dispuseram a compartilhar seus dados para que eu pudesse executar o projeto. Fazer ciência no Brasil é um ato de resistência, por isso agradeço a todos os cientistas que persistiram e ajudaram a construir este trabalho.

# Sumário

# Lista de figuras

**Figure 1.** Species accumulation curve for all nine sample points. Species richness is the average of simulated resampling for each number of plots.

**Figure 2.** Boxplot of the values of difference of richness of the simulations for each N of plots and the total richness observed (Richness difference with total), in function of the number of 15 m transects, for the nine streams. The red lines represent the fit of the segmented regression model. Figures a) to c) represents streams with 100% of primary forest cover; d), e) and f) streams with primary and secondary forest cover and; g), h) and i) streams with primary, secondary and agriculture cover.

**Figure 3.** Result of the linear regression analysis performed between the minimum sampling effort (Number of 15 m plots) as a function of the coordinate values for each sampled point. The red line represents the fit of the linear regression model.

**Figure 4.** The axes of graphs correspond to axes of Principal Coordinate Analysis (PCoA) applied with data on the relative abundance of the species composition of the simulated transects. The numbers correspond to the number of plots resampled and the lines to the confidence intervals on both axes of the scores resulting from the multivariate analysis (PCoA). Figures a) to c) represents streams with 100% of primary forest cover; d), e) and f) streams with primary and secondary forest cover and; g), h) and i) streams with primary, secondary and agriculture cover.

**Figure 5.** Similarity in the composition of species simulated for each number of transercts (N) relative to the total composition in 150 m, for the nine streams. The simulated values for each N were represented using the boxplot. The red line represents

the fit of the segmented regression model. Figures a) to c) represents streams with 100% of primary forest cover; d), e) and f) streams with primary and secondary forest cover and; g), h) and i) streams with primary, secondary and agriculture cover.

**Figure 6.** Result of the linear regression analysis performed between the minimum sampling effort (Number of 15 m plots) as a function of the coordinate values for each sampled point. The red line represents the fit of the linear regression model.

# Resumo

Enquanto uma amostragem excessiva consome tempo e recursos financeiros que poderiam ser investidos em outras áreas, uma subamostragem pode gerar métricas enviesadas e levar a conclusões errôneas sobre a comunidade biológica estudada. Estudos sobre esforço amostral com peixes de riachos, no Brasil, são escassos. Definir um mínimo de esforço amostral é importante, principalmente para se obter estimativas precisas em levantamentos taxonômicos, fundamentais para conservação de grupos ameaçados, como assembleias de peixes de áreas afetadas pela supressão vegetal causada pelo avanço agropecuário. Nosso objetivo foi estabelecer um esforço mínimo de amostragem adequado para estimar com precisão a riqueza e composição de espécies, e testar se esse mínimo varia em função da cobertura florestal. Nosso protocolo consistiu em simular valores de riqueza e composição de espécies em trechos diferentes tamanhos. Nós executamos reamostragens, randomizadas e repetidas 50x, com dados previamente coletados em trechos de diferentes tamanhos. Para cada repetição de cada tamanho de trecho, calculamos a diferença de riqueza e composição de espécies com o maior trecho coletado (i.e., 150 m) e analisamos a variação dessa diferença em função do tamanho amostral. Com a regressão segmentada, determinamos para riqueza e composição um mínimo amostral a partir do qual a relação diferença de riqueza/composição em função do tamanho amostral foi atenuada. Os resultados sugerem que o mínimo amostral para riqueza e composição é de ~45 m. Nenhuma evidência foi encontrada de que este mínimo varia dentro dos níveis de desmatamento das áreas avaliadas. Os resultados obtidos aqui devem ser utilizados com precaução, visto que este resultado foi obtido para uma área bem específica.

**Palavras-chave:** tamanho amostral, esforço amostral, comprimento de trecho, peixes de água doce, ictiofauna.

# Abstract

While oversampling consumes time and financial resources that can be invested in other areas, undersampling can generate biased metrics and mislead about the studied biological community. Studies on sampling effort with fish from streams in Brazil are scarce. Defining a minimum sampling effort is important, mainly to obtain precise sources in taxonomic surveys, essential for the conservation of threatened groups, such as fish assemblages from areas affected by vegetation suppression caused by agricultural advances. Our objective was to establish an adequate minimum sampling effort to accurately estimate species richness and composition, and test if this minimum varies as a function of forest cover. Our protocol consisted of simulated values of species richness and composition in different sized stretches. We performed re-sampling, randomized and repeated 50 times, with data previously collected in different sized stretches. For each repetition of each stretch size, we calculated the difference in species richness and composition with the longest stretch collected (i.e., 150 m), and analyzed the variation of this difference as a function of sample size. With segmented regression, we determined a sample minimum, for richness and composition, from which the difference in richness/composition as a function of sample size was attenuated. The results obtained that the sample minimum for richness and composition is ~ 45 m. No evidence was found that this minimum varies within the deforestation levels of the assessed areas. The results selected here should be used with caution, as this result has been published for a very specific area.

**Key words:** sample size, sampling effort, reach length, freshwater fish, ichthyofauna.

# Introdução Geral

As métricas de biodiversidade fornecem uma maneira de entender os sistemas ecológicos. A maioria dos descritores ecológicos, entretanto, são sensíveis ao tamanho da amostra, de forma que um número crescente de repetições ou áreas pesquisadas pode alterar os valores das métricas, muitas vezes levando a inferências tendenciosas (Gotelli & Colwell, 2001). Por outro lado, a aplicação de um tamanho de amostra excessivo consome recursos que limitam o tamanho da amostra ou podem ser reinvestidos em outros objetivos. Um esforço mínimo de amostragem geralmente difere entre pesquisas e grupos taxonômicos (Walther et al., 1995; Taylor 2002, Bergallo et al. 2003; Delabie et al., 2000). Para peixes, os pesquisadores frequentemente definem um esforço mínimo de amostragem com base em como as mudanças no tamanho da amostragem afetam as métricas relacionadas às assembleias, geralmente riqueza, abundância e diversidade (Paller, 1995; Angermeier & Smogor, 1995; Fischer & Paukert, 2009). Avaliar como a similaridade na composição das espécies aumenta em função do tamanho da amostra poderia complementar as estimativas do esforço mínimo de amostragem fornecido apenas pela riqueza. No Brasil, estudos conduzidos em riachos da Amazônia Central e da Mata Atlântica obtiveram diferentes estimativas de mínimo amostral. Para a Amazônica, utilizando curvas de acumulação de espécies, observou-se que a riqueza de espécies atingia um platô apenas com 180 m de trecho amostrado (Anjos & Zuanon, 2007) Para a Mata Atlântica, observou-se que com 100 m a riqueza de espécies não atingia um platô (Terra et al., 2013).

Os pequenos riachos que compõem as grandes bacias hidrográficas brasileiras têm forte associação com a mata ripária, pois ela regula uma série de fatores no ambiente aquático (Lowrance *et al*., 1997). Ela impede o superaquecimento da água, fornece materiais alóctones (e.g., folhas, sementes, galhos, etc) que sustentam uma cadeia alimentar no curso d’água, promovem heterogeneidade ambiental, e torna possível a presença de organismos com diferentes hábitos de vida (Macedo *et al*., 2013; Tank *et al*., 2010; Bojsen & Barriga, 2002). A retirada da mata associada aos riachos aumenta a entrada de sedimentos nos cursos d’água, promove aumento da temperatura da água e diminuição no aporte de nutrientes (Macedo et al., 2013; Miserendino, 2011). Estas mudanças ambientais promovem, consequentemente, uma mudança na composição e riqueza de espécies presentes localmente, considerando que apenas uma parte delas está realmente apta a viver em um novo ambiente simplificado. Casatti *et al* (2012) mostrou, por exemplo, que em um gradiente de degradação da mata ripária há a dominância de espécies de peixe tolerantes a ambientes assoreados e de hipóxia (e.g. *Phalloceros harpagos* e *Poecilia reticulata*) em detrimento de especialistas que são dependentes de diferentes substratos e maiores níveis oxigênio dissolvido (e.g. *Characidium zebra*). Como a supressão da vegetação ripária (i.e., vegetação circundante aos riachos) traz efeitos na riqueza e composição de espécies, isso pode ser refletido também no esforço amostral mínimo, que pode ser diferente entre áreas mais ou menos impactadas.

Os objetivos deste estudo foram: I) determinar o esforço amostral mínimo para assembleias de peixes de riachos e; II) analisar se este mínimo amostral varia em função do nível de supressão vegetal. Nós hipotetizamos que I) o esforço mínimo amostral seria de 100 m, visto que estimativas anteriores foram altas e II) e o esforço amostral mínimo para riachos com maior supressão vegetal seria menor, já que áreas mais impactadas mostraram menor riqueza de espécies e dominância de algumas espécies em detrimento de outras.

Para analisar estas questões acima, nós criamos uma simulação para estimar o mínimo amostral necessário para representar a composição de espécies de um dado local. Como a composição de espécies se trata de um dado multivariado, nós utilizamos a mesma lógica da construção das curvas de acumulação de espécies para re-amostrar aleatoriamente sub-parcelas de riachos de 15 m de comprimento e registrar a abundância relativa de cada espécies para diferentes comprimentos de trechos de riachos. Em seguida, nós avaliamos diferenças de riqueza e composição de espécies para cada um dos comprimentos de riachos relativos ao tamanho total de 150 m de riachos e utilizamos uma regressão segmentada para definir o ponto de inflexão a partir do qual a assíntota era atingida.

## Referências

Angermeier, P. L., & Smogor, R. A. (1995). Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(5), 936-949. https://doi.org/10.1139/f95-093

Anjos, M. B. D., & Zuanon, J. (2007). Sampling effort and fish species richness in small terra firme forest streams of central Amazonia, Brazil. *Neotropical Ichthyology*, *5*(1), 45-52. https://doi.org/10.1590/S1679-62252007000100006

Bergallo, H. G., Esbérard, C. E., Mello, M. A. R., Lins, V., Mangolin, R., Melo, G. G., & Baptista, M. (2003). Bat species richness in Atlantic Forest: what is the minimum sampling effort?. *Biotropica*, *35*(2), 278-288. https://doi.org/10.1111/j.1744-7429.2003.tb00286.x

Bojsen, B. H., & Barriga, R. (2002). Effects of deforestation on fish community structure in Ecuadorian Amazon streams. *Freshwater Biology*, *47*(11), 2246-2260. https://doi.org/10.1046/j.1365-2427.2002.00956.x

Casatti, L., Teresa, F. B., Gonçalves-Souza, T., Bessa, E., Manzotti, A. R., Gonçalves, C. D. S., & Zeni, J. D. O. (2012). From forests to cattail: how does the riparian zone influence stream fish?. *Neotropical Ichthyology*, *10*(1), 205-214. https://doi.org/10.1590/S1679-62252012000100020

Delabie, J. H., Fisher, B. L., Majer, J. D., & Wright, I. W. (2000). Sampling effort and choice of methods. *Ants: standard methods for measuring and monitoring biodiversity*, 145-154.

Fischer, J. R., & Paukert, C. P. (2009). Effects of sampling effort, assemblage similarity, and habitat heterogeneity on estimates of species richness and relative abundance of stream fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(2), 277-290. https://doi.org/10.1139/F08-209

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology letters, 4(4), 379-391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Lowrance, R., Altier, L. S., Newbold, J. D., Schnabel, R. R., Groffman, P. M., Denver, J. M., ... & Staver, K. W. (1997). Water quality functions of riparian forest buffers in Chesapeake Bay watersheds. *Environmental Management*, *21*(5), 687-712.

Macedo, M. N., Coe, M. T., DeFries, R., Uriarte, M., Brando, P. M., Neill, C., & Walker, W. S. (2013). Land-use-driven stream warming in southeastern Amazonia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1619), 20120153.doi:10.1098/rstb.2012.0153.

Miserendino, M. L., Casaux, R., Archangelsky, M., Di Prinzio, C. Y., Brand, C., & Kutschker, A. M. (2011). Assessing land-use effects on water quality, in-stream habitat, riparian ecosystems and biodiversity in Patagonian northwest streams. *Science of the Total Environment*, *409*(3), 612-624

Paller, M. H. (1995). Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. *North American Journal of Fisheries Management*, *15*(1), 110-120. https://doi.org/10.1577/1548-8675(1995)015%3C0110:RANOFS%3E2.3.CO;2

Taylor, A. F. (2002). Fungal diversity in ectomycorrhizal communities: sampling effort and species detection. In *Diversity and integration in Mycorrhizas* (pp. 19-28). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1284-2_3

Tank, J. L., Rosi-Marshall, E. J., Griffiths, N. A., Entrekin, S. A., & Stephen, M. L. (2010). A review of allochthonous organic matter dynamics and metabolism in streams. *Journal of the North American Benthological Society*, *29*(1), 118-146. https://doi.org/10.1899/08-170.1

Terra, F.B., Hughes, R. M., & Araújo, F. G. (2013). Sampling sufficiency for fish assemblage surveys of tropical Atlantic forest streams, Southeastern Brazil. Fisheries, 38(4), 150-158. https://doi.org/10.1080/03632415.2013.775572

Walther, B. A., Cotgreave, P., Price, R. D., Gregory, R. D., & Clayton, D. H. (1995). Sampling effort and parasite species richness. Parasitology today, 11(8), 306-310. https://doi.org/10.1016/0169-4758(95)80047-6

**Submetido para Ecology of freshwater fish. Fator de impacto: 1.950.**

# ESTIMATING THE MINIMUM SAMPLE SIZE FOR STREAM FISH IN A CERRADO-AMAZON TRANSITION ZONE

Isabela Carolina Silva Vieira Braga[1,2], Pâmela Virgulino Freitas[3,4], Luciano F. A. Montag[4], Murilo Sversut Dias[2]

[1]Programa de Pós-graduação em Ecologia – Universidade de Brasília (UnB)

[2]Departamento de Ecologia, Instituto de Ciências Biológicas, Brasília, Distrito Federal, Brasil

[3]Programa de Pós-graduação em Ecologia – Universidade Federal do Pará/Embrapa

[4] Laboratório de Ecologia e Conservação, Instituto de Ciências Biológicas, Belém, Pará, Brasil

**Corresponding author**

Isabela Carolina S. V. Braga, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Brasil.

E-mail: isabelabragav@gmail.com & msdias@unb.br

Telphone number: +55 61 983551785

## ABSTRACT

Applying an excessive sampling size consumes resources that either limits sampling size or could be reinvested into other goals. Acquiring accurate estimates on some assemblage metrics (i.e., abundance, richness, species composition) therefore remains a constant challenge, as well as assessing whether the applied effort suffices the survey's goals. S tudies on sampling effort in streams provide variable results. Yet, sites subjected to different impact regimes could also require different sampling efforts to properly record species composition. The main question here is what is the minimum reach length required to reliable ecological estimates of stream fish assemblages? Our goal was to set a minimum sampling effort suitable for accurately estimate species richness and composition. We also detailed how a minimum sampling effort changes under a gradient of disturbance related to changes in forest cover from riparian forests to monocultures. We used a data set of nine streams of the Upper Xingu river basin. Thus, we executed permutations with presence/abcense and abundance data, in wich we constructed acumulation curves, difference of richness and performed Bray-Curtis Distance. We executed Segmented Regression to determine the minimum sampling effort from similarity and richness difference as a function of the number of plots em each stream. We concluded that a minimum of ~45 m stretch is necessary to better describe assemblage structure of stream fish, on terms of species richness and composition. No evidence was found that this minimum varies within levels of deforestation.

**Key words:** sample size, sampling effort, reach length, freshwater fish, ichthyofauna.

## INTRODUCTION

Biodiversity metrics provide a way to understand ecological systems. Precise estimations enable to test predictions on state-of-the-art theories (Angermeier & Schlosser, 1989; Triants *et al.*, 2003; Mathews *et al.*, 2019), monitor conservation efforts (Conroy & Noon, 1996; Fleishman et al., 2006; Freemark et al., 2006) and evaluate impacts on biota at distinct scales (Ganasan & Hughes, 1998; Bradford *et al.*, 1998). Most ecological descriptors, however, are sensitive to sampling size, such that an increasing number of replicates or surveyed areas may change the metrics values, often leading to biased inferences (Gotelli & Colwell, 2001). On the other hand, applying an excessive sampling size consumes resources that either limits sampling size or could be reinvested into other goals.

Inferences in community ecology commonly rely on metrics to survey and describe features of local species pools (e.g., richness, diversity and evenness). Acquiring accurate estimates on these metrics therefore remains a constant challenge, as well as assessing whether the applied effort suffices the survey’s goals. When sampling species richness, for instance, researchers usually test the quality of their estimation of total species richness in a given area by plotting accumulation curves (Soberón & Llorente, 1993; Mao et al., 2000; Ugland *et al.*, 2003). Those curves inform how the recorded species richness grow as the sampling effort increase (Gotelli & Colwell, 2001), also indicating whether this relationship reached an asymptote, *i.e.* a moment in which increasing sampling effort does not result in more detected species. Detecting this moment of no increase in richness (the minimum sample size or effort) is fundamental to setting best experimental designs and optimal cost-effective surveys.

A minimum sampling effort usually differs among surveys and taxonomic groups (Walther *et al.*, 1995; Taylor 2002, Bergallo *et al.* 2003; Delabie *et al.*, 2000). Researchers frequently set a minimum sampling effort based on how changes in sampling size affect metrics related to their subjects, usually richness, abundance and diversity in community ecology (Paller, 1995; Angermeier & Smogor, 1995; Fischer & Paukert, 2009). Species richness is the most frequent metric used to define minimum sampling efforts (maybe because more ways to evaluate it are available), but evaluating other metrics and indexes are highly recommented (Walther & Martin, 2001). For instance, assessing how similarity in species composition increases as a function of sample size could complement estimates on minimum sampling effort provided by richness alone.

Freshwater ecosystems harbor a large proportion of animal biodiversity worldwide despite occupying a small surface area (Jenkins, 2003). The Amazon basin alone includes around 15% of described fish species (Tedesco *et al.*, 2017), with many species present in small streams (Castro, 1999; Dias et al., 2021). In these environments, riparian forests usually surround streams, affecting how they connect to the nearby ecosystems (Lowrance et al., 1997). Shading provided by the canopy reduces sunlight inputs into streams, the forest itself provides a source of allochthonous resources (as leafs, seeds and branches) capable of sustaining an underwater food chain and promote habitat heterogeneity suitable for sheltering organisms with diverse life histories (Macedo *et al.*, 2013; Tank *et al.*, 2010; Bojsen & Barriga, 2002).

To establish local mono-cultures, land owners frequently suppress native riparian vegetation, essentially changing the stream ecosystemic dynamics by allowing a high sediment and nutrient income, followed by an increase in daily temperature (Macedo *et al.*, 2013; Miserendino, 2011). These changes usually favor generalist species, capable of enduring a wider range of environmental conditions (Ilha et al., 2019). The urgency in recording biodiversity highlights the importance of accurately sampling natural communities and those subject to intense land-use conversion.

Studies in streams usually employ stream longitudinal sections (i.e., reaches) as the sampling unit, and its size is highly variable among studies (e.g., 50 - 150 m in Amazonia) (Mendonça et al. 2005; Montag *et al.*, 2018), 30 - 500 m in Atlantic Forest (Pessanha et al., 2003; Ferreira & Petrere, 2009; Terra et al., 2013) 30 – 150 m in Cerrado (Schneider et al., 2011; Macedo et al., 2014; de Carvalho et al., 2017). There is no consensus on the minimum size needed for reliable richness and composition estimates, and studies on sampling effort in streams provide variable results. The minimum length of stream reach to useful estimates of species richness increases with stream order, first, second and third order streams requiring samples of at least 180 m, 213 m and 253 m, respectively (Anjos e Zuanon, 2007). Surveys set on human-impacted streams (suffering from urbanization, waste discharge and deforesting) in Atlantic forest indicate the sampled stream reach length must extend between 80 to 100 times the stream width in order to correctly sample species richness, though reach length of about 40 times the stream width have proven reliable estimates of biological integrity index (Terra et al., 2013). A study conducted in forth order rivers on other continents, however, show a smaller sampling effort (~ 65 m) could miss rare species and lead to high variability on richness estimates, compromising the suitability of assemblages assessments (Angemeier & Smogor, 1995).

Yet, sites subjected to different impact regimes could also require different sampling efforts to properly record species composition. For instance, deforestation usually leads to a homogenization of stream habitats, often reducing the abundance of sensitive species (Casatti et al., 2012; Brejão *et al.*, 2018). Under an increasing gradient of riparian forest degradation, fish species tolerant to silting and hypoxia (as *Phalloceros harpagos* e *Poecilia reticulata*) prevail over environment specialists that rely upon a diverse set of substrates and high oxygen content to thrive (as *Characidium zebra*) (Casatti et al 2012). If fish richness and composition diverge between pristine and disturbed systems, the minimum sampling effort (as stream reach length) required to sample those assemblages might also differ. Heavy disturbed sites, for instance, should require short stream reaches in comparison to undisturbed areas, in such a way that pristine streams shall require a much greater effort. The same reasoning may apply conversely, such that sampling disturbed sites using the same effort suitable for pristine areas consumes unnecessary field time and resources. Therefore, discerning sampling biases from other environmental constrains, as human related disturbances, becomes essential to draw reasonable ecological parameter estimates. So, defining a minimum sampling effort suitable to compare disturbed and undisturbed streams becomes paramount to freshwater ecologists.

The main question here is what is the minimum reach length required to reliable ecological estimates of stream fish assemblages? Our goal was to set a minimum sampling effort (in terms of stream reach length) suitable for accurately estimate species richness and composition. We also detailed how a minimum sampling effort changes under a gradient of disturbance related to changes in forest cover from riparian forests to monocultures. We hypothesize that disturbed assemblages require less sampling effort than undisturbed ones as impacted streams shelter more homogeneous assemblages with smaller species richness.

## MATERIAL AND METHODS

### *Study area*

Sampling was conducted in first and second order streams of Tanguro and Darro drainage basins, in Tanguro farm area, located in Querência/MT (Attachment 1). The region is a Cerrado/Amazon transition area and has marked seasonality, with the rainy season occurring from October to April and the dry season from May to September. Streams are located between longitudes 52 ° 23'30” W and 142 52 ° 18’50” W, and latitudes 13 ° 9’12” S and 12 ° 41’40” S. The farm covers 82,000 hectares, 60% of which is covered by original vegetation. The rest of the area is used for cultivation, mainly of soy (Ilha, 2015).

### *Data set*

The data used here were collected in a study by Freitas et al. (2019), who evaluated the effect of vegetation cover loss on the functional and taxonomic diversity of stream fishes. Nine streams were sampled (six first-order and three second-order streams) in August 2017. Figure of the points on the map and the analysis can be seen in the study by Freitas et al. (2019). The Tanguro farm is composed of polygons of primary and secondary forests, and areas with monoculture farming. The sampled points form a gradient of degradation, as each one is composed of primary and secondary forests, and agricultural cover. The points ua1, ua2 and ua3 correspond to more conserved areas, being composed of 100% of primary forest; the points ua4, ua5 and ua6 correspond to areas in secondary forests; and the points ua7, ua8 and ua9 correspond to areas with primary and secondary forest, but with a higher percentage of agricultural areas.

The sampling protocol consisted of 150 m-long stretches delimited in nine streams. Each stretch was cross-sectioned 11 times, resulting in 10 plots of 15 m-long each, without spacing between them. The sampling effort consisted of three people collecting for 12 min in each 15 m plot, totaling 2 hours per stretch (Attachment 2). The collection was carried out with aquatic and dip nets (40 and 55 cm in diameter, respectively), both with 1 mm metallic mesh. Fishes collected were separated for each of the ten plots, euthanized in clove oil, fixed in 10% formalin and then transferred to 70% alcohol. Individuals were identified with specialists and with help of identification keys. In total, 2,943 individuals were collected, from 29 species, 27 genera, 17 families and 6 orders (Freitas, 2019).

### *Defining minimum sampling effort*

In each 15 m plot, species and number of individuals in each were recorded, generating a database of 10 rows and 29 columns (total species in the whole data set) for each stream sampled. This initial data matrix was used for the procedures described below.

We executed random resamplings of N sample size with the 10 plots in each of the nine streams, repeating the operation 50 times for each N of plots. For example, for N = 1, we randomly sampled a single plot that contained species composition data (abundance of each species present in the plot); for N = 2, we simulated the sampling of two plot chosen at random. We did this procedure up to N = 10, each N sampling have been repeated 50 times.

*Species richness*: For each of the nine collected streams, we analyzed the number of species recorded for each number of selected plots (i.e. number of 15 m plots; hereafter, N of plots), resulting in nine species accumulation curves. For the curves construction, we calculated the average species richness of the 50 randomizations for each N. With the randomization of resamplings, which is identical to classical species accumulation curves (Gotelli & Colwell 2001), it is possible to observe an increase in the number of species as a function of N plots (sample size).

*Species composition*: To analyze the difference in species composition as a function of the size of the sampled stretch (i.e., N of 15 m plots), we grouped and add the abundance data for each species separately to create a species composition matrix, in which the lines correspond to the 50 random samplings for each N ranging from 1 to 10, and the columns correspond to the species of fish and the matrix cells correspond to the abundances of each species added in each of the resamples. The resulting final matrix has 500 rows (50 resamples for each N ranging from 1 to 10) and 29 columns (species).

### *Statistical analysis*

We extracted the richness values for each N value of plots and calculate the difference between these richness and the total richness collected in each stream (i.e., the 150 m stretch). This difference reflects how close the expected richness for each N value is to the observed total value. With the species composition data resulting from the simulations, we applied the Bray-Curtis pair-wise distance (based on relative species abundance data, i.e., standardized by the sum of individuals sampled in the assembly) to create a matrix of dissimilarity based on the composition of the community. We followed the same protocol as in the richness difference and used the distance matrix to estimate how similar each simulation was to the total species composition observed for all 10 plots of each stream. Thus, we obtained an error measure (distance) in relation to the observed total species composition of each point for each N. Then, we plotted the two difference measures (i.e., difference in richness and difference in composition) as a variable response (y-axis) and the number of plots (N) as a predictor variable (x-axis).

To improve the visualization of composition variation as a function of the sample N, we applied a Principal Coordinate Analysis (PCoA) in the Bray-Curtis distance matrix to summarize the composition in two orthogonal axes. Thus, we calculated the mean and standard error of the 50 values of scores for each axis for each N of plots, totaling 10 values of mean and standard error of species composition in the multivariate space. Thereby, it was possible to visualize the average variability in the species composition for each selected N and to visualize the similarity of the composition with the total composition observed in the 10 stretches together.

To extract the minimum sample necessary to detect the local composition, we performed two Segmented Regressions (Piecewise Regression) using the values of difference in richness and composition as a response variable and number of plots sampled as an explanatory variable. This regression model is used when the analyzed data have break points and/or non-linear relations, with a break point from which the response variable will assume another pattern as a function of the explanatory variable (Portz et al., 2000; Muggeo, 2003). The regression returns a break value (Toms & Lesperance, 2003), which in this analysis represents the value of the minimum sample for each sample site. We repeat the same procedure for each of the nine streams.

Finally, we tested whether there is a relationship between the minimum sample size estimated as a function of the degree of impact by running a linear model (Linear Regression) with the minimum break values detected for each stream as a response variable and with the coordinates of the PCA axes of each site as an explanatory variable base on vegetation cover in 100 m buffers. We extracted the coordinate values for each point from the result of the PCA performed in the work of Freitas (2019; unpublished data), from Axis I, which summarized 84% of the data variation. Negative values of the PCA axis represent points with less forest coverage and positive points with greater forest coverage. All analyzes were performed using the R program (R core team, 2019), using the *‘segmented’* (Muggeo, 2003) and *‘vegan’* (Oksanen et al., 2019) packages.

## RESULTS

### *Species richness*

For all nine sample sites, the accumulation curves tend to stabilize with the increase in the number of plots sampled (Figure 1).

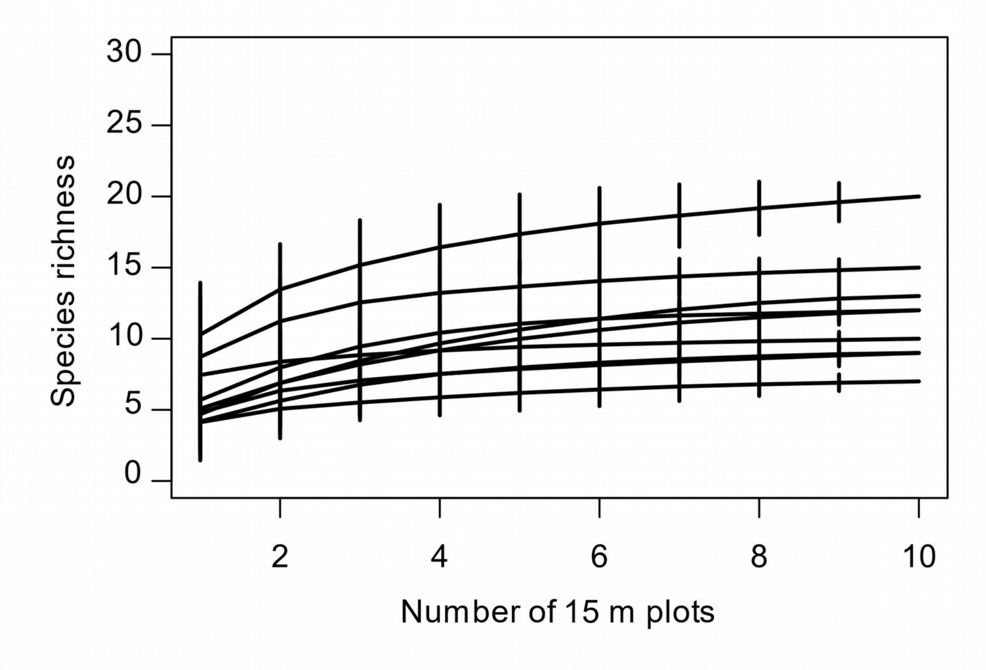


**Figure 1.** Species accumulation curve for all nine sample points. Species richness is the average of simulated resampling for each number of plots.

Using the difference in species richness between N plots and the total number of species found in the plots, the average break point (i.e., minimum sample in relation to species richness) resulting from the segmented regression for all nine streams analyzed was 3.07 (Standard Error: ± 0.79, Confidence Interval: 1,52 - 4,62), indicating that, on average, 3 plots of 15 m (i.e., a minimum length of 45 m) are needed to obtain a representative sample of the community richness (Fig. 2). The segmented regression was significant for all nine points analyzed, however, with this threshold on average, four species would not be detected in relation to the total richness of each stream (Richness difference = 4, Fig. 2).

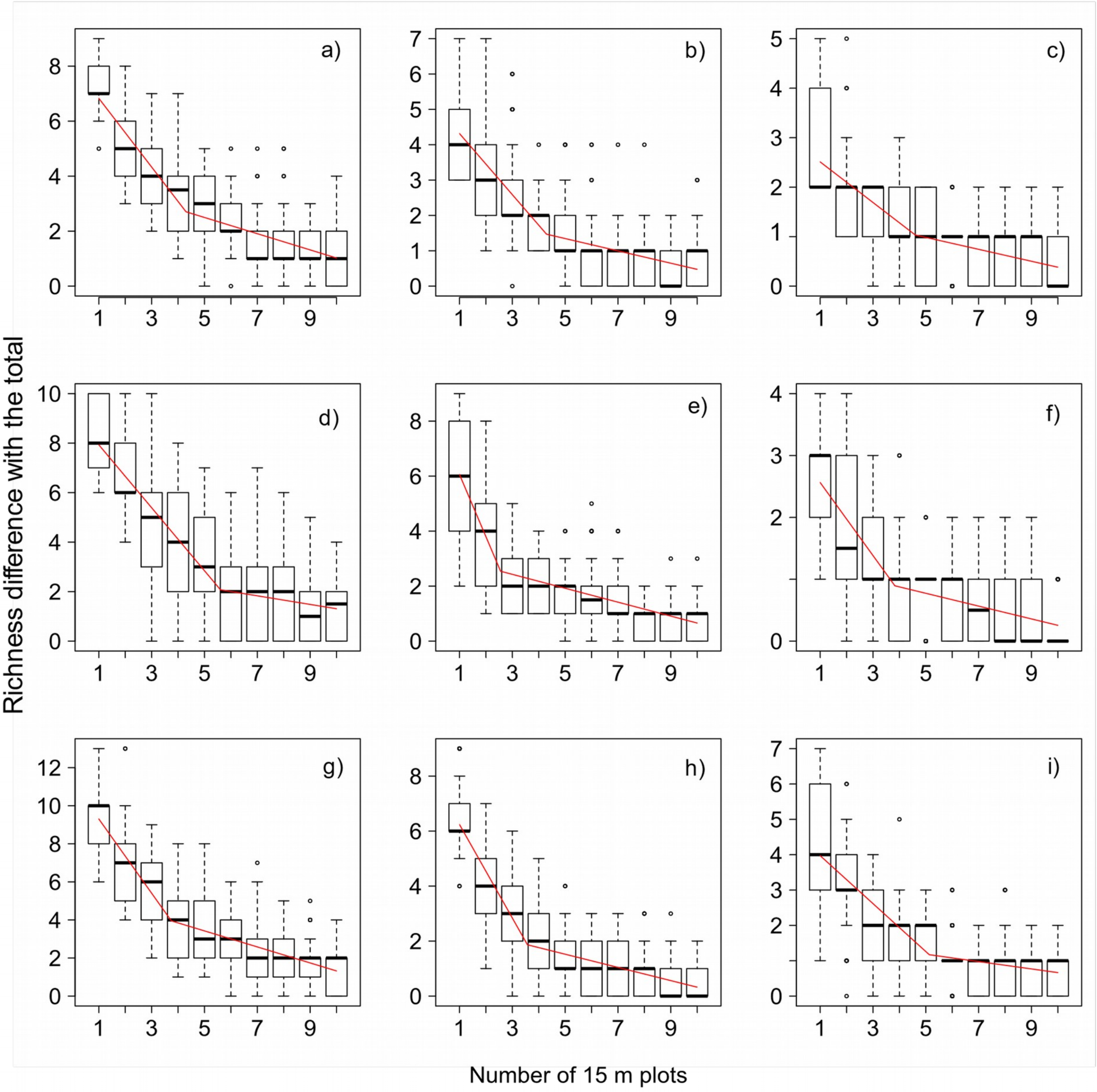


**Figure 2.** Boxplot of the values of difference of richness of the simulations for each N of plots and the total richness observed (Richness difference with total), in function of the number of 15 m plots, for the nine streams. The red lines represent the fit of the segmented regression model. Figures a) to c) represents streams with 100% of primary

forest cover; d), e) and f) streams with primary and secondary forest cover and; g), h) and i) streams with primary, secondary and agriculture cover.

The linear regression model performed with the minimum sampling effort, extracted from the segmented regression analysis, for each point as a function of coordinates of Axis I (i.e., canopy cover) was not significant ($t_{1,7} = 0.594$; $p = 0.571$), showing that the minimum sampling effort, to be adopted so that it has the smallest difference in richness with the total richness on the 150 m stretch, is the same for areas with more or less vegetation cover (Fig.3).

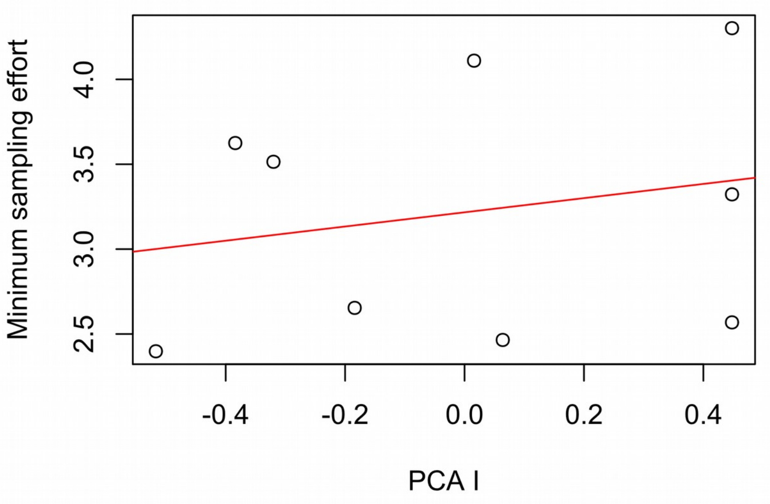


**Figure 3.** Result of the linear regression analysis performed between the minimum sampling effort (Number of 15 m plots), for species richness, as a function of the coordinate values for each sampled point. The red line represents the fit of the linear regression model.

*Species composition*

The scores resulting from the PCoA analysis show that the greater the number of sampled plots, the more similar and, hence, the closer the points are to the species composition of the largest sampled stretch (i.e., point with 10 sampled plots) (Figure 4). In addition, large confidence intervals show in the first N values and the decrease in the interval with the increase in N number shows less variability in species composition as N increases. This demonstrates that the larger the sampled section, more accurate and precise are the composition estimates (Figure 4).

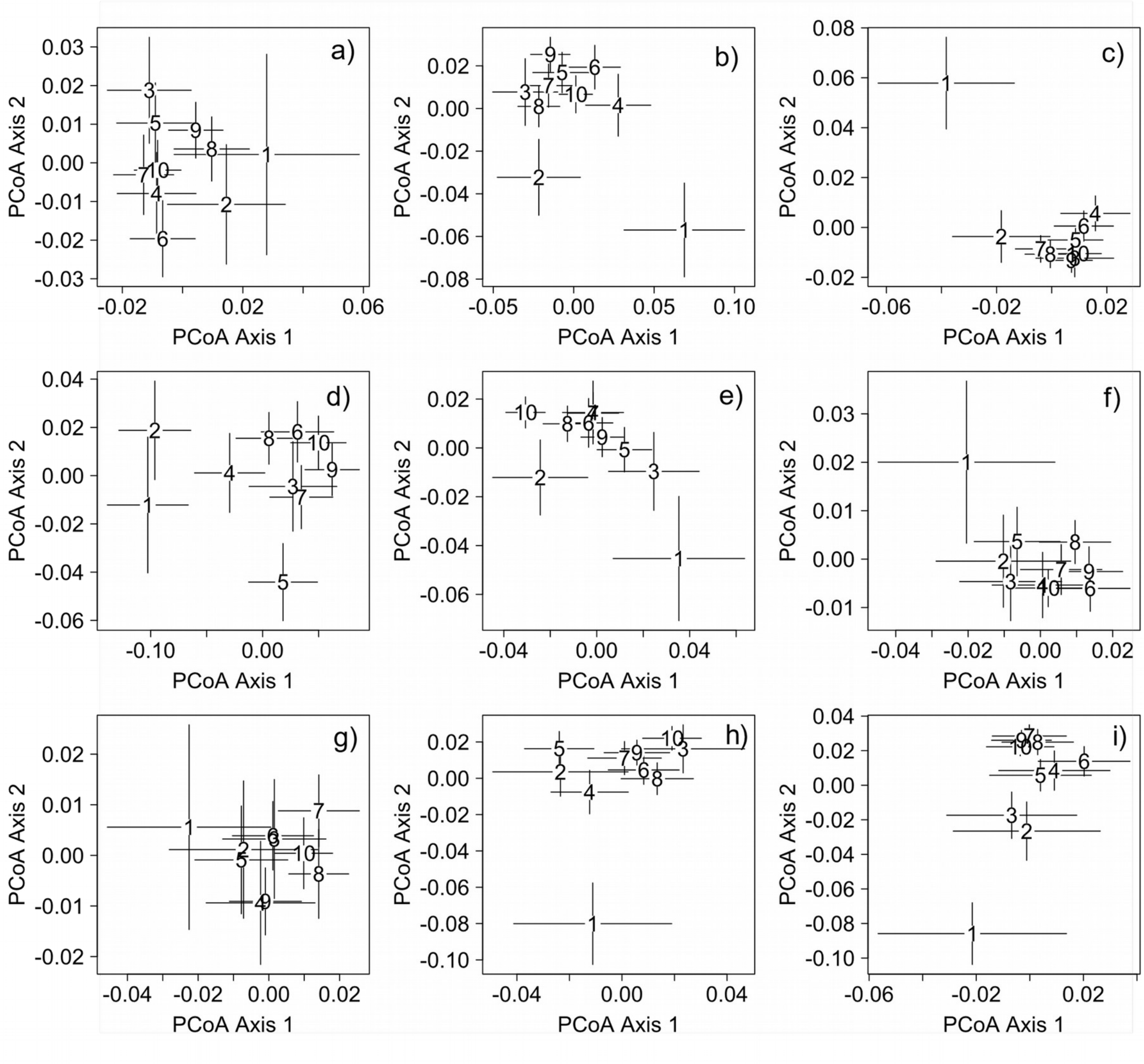


**Figure 4.** The axes of graphs correspond to axes of Principal Coordinate Analysis (PCoA) applied with data on the relative abundance of the species composition of the simulated transects. The numbers correspond to the number of plots resampled and the lines to the confidence intervals on both axes of the scores resulting from the multivariate analysis (PCoA). Figures a) to c) represents streams with 100% of primary

forest cover; d), e) and f) streams with primary and secondary forest cover and; g), h) and i) streams with primary, secondary and agriculture cover.

The higher the N value, the shorter the distance from the total species composition, and the segmented regression was significant for all nine points analyzed (Figure 5). For the App2 point, the segmented regression returned a break value of 4,271 ($t_{1,496}$ = 11.80, p = <2e-16), visually shown by the slight accentuation of the curve (Fig. 5a). For the App4 point, the break value was 3,855 ($t_{1,496}$ = 11.68, p = 1.5e-11) (Fig. 5d). The Appp point had a break value of 2,635 ($t_{1,496}$ = 13.52, p = <2e-16) (Fig. 5g). These points were described here because of their percentage of vegetation cover, where the first one has 100% of primary forest cover, and the second and third has 60% of primary forest, 38% of secondary forest and 2% of agriculture cover, relying the gradient of vegetation cover.

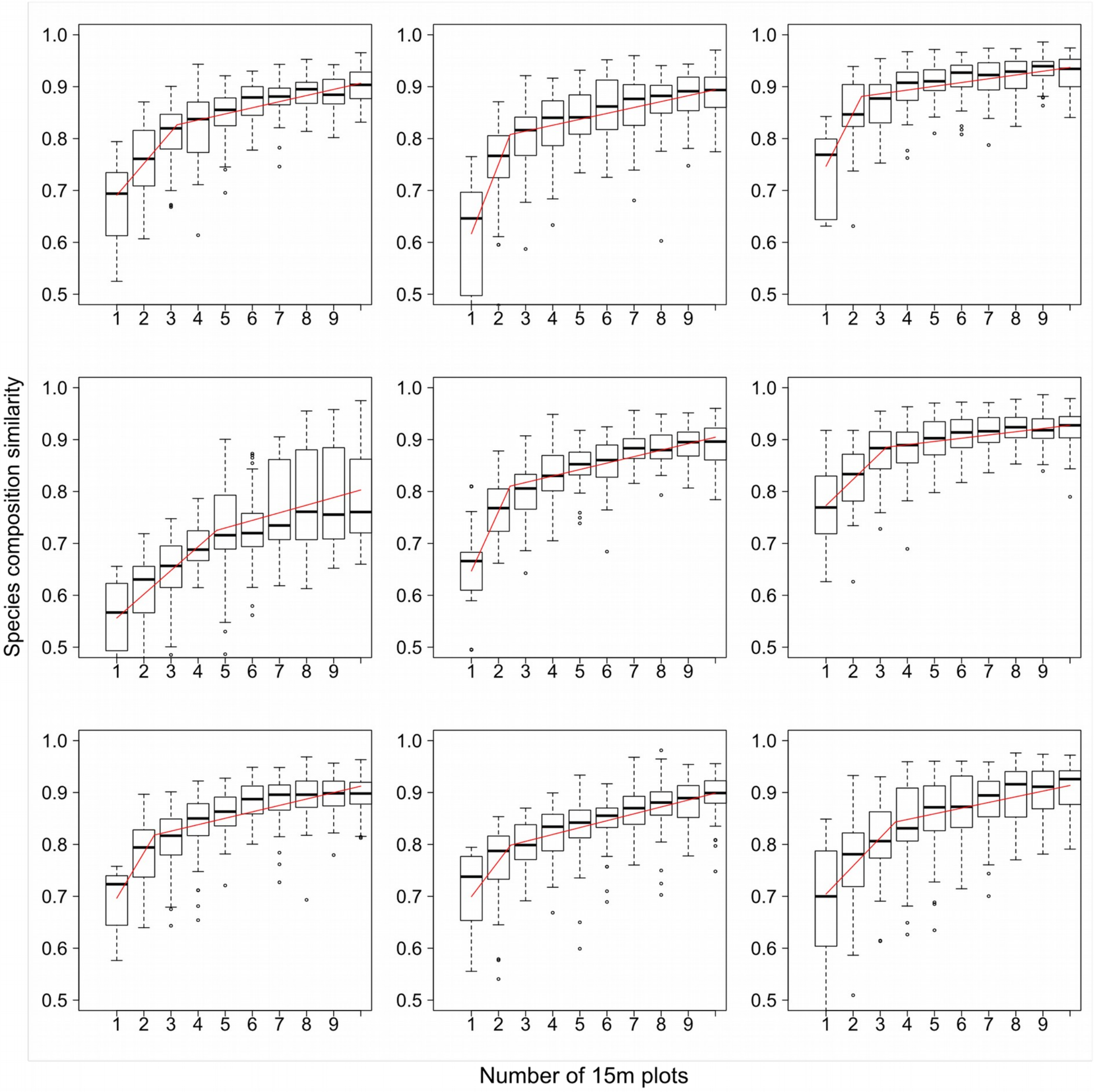


**Figure 5.** Similarity in the composition of species simulated for each number of transects (N) relative to the total composition in 150 m, for the nine streams. The simulated values for each N were represented using the boxplot. The red line represents the fit of the segmented regression model. Figures a) to c) represents streams with 100% of primary forest cover; d), e) and f) streams with primary and secondary forest cover and; g), h) and i) streams with primary, secondary and agriculture cover.

In general, the average break point for all nine streams analyzed was 2.98 (± 0.39), indicating that an average of three 15 m plots (i.e., a minimum length of 45 m) is required to obtain a representative sample of the community. The linear regression model was not significant ($t_{1.7} = 0.594$; $p = 0.571$), showing that the minimum sampling effort to be adopted, so that it has the smallest difference in composition with the 150 m stretch, is the same for areas with more or less vegetation cover (Fig. 6).

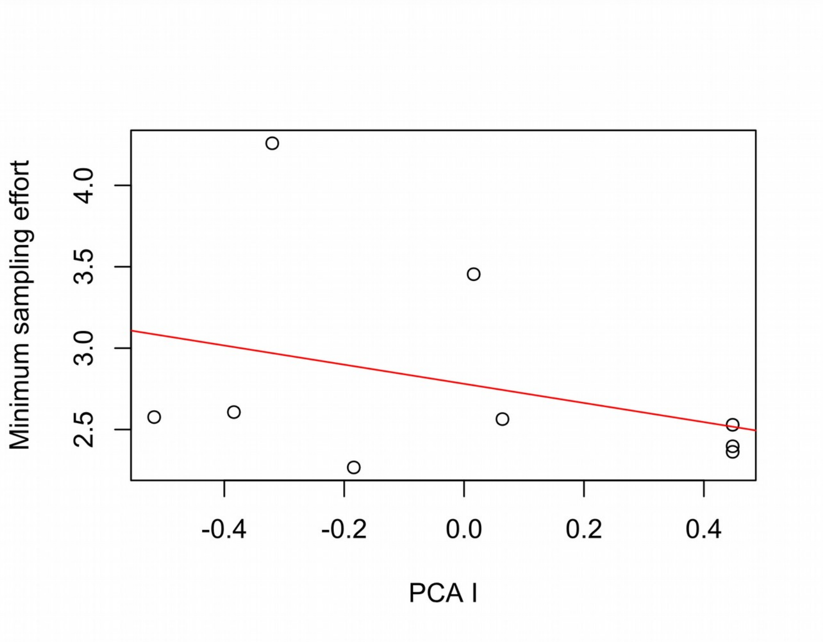


**Figure 6.** Result of the linear regression analysis performed between the minimum sampling effort (Number of 15 m plots), for species composition, as a function of the coordinate values for each sampled point. The red line represents the fit of the linear regression model.

## DISCUSSION

To understand biodiversity and the impacts of multiple land uses, it is necessary to have an excellent representation of the local fauna in ecological studies. Moreover, determining a sample size smaller than necessary can lead to errors in taxonomic surveys, while a larger sample size can lead to a waste of time and financial resources. Based on a well-established technique used to estimate whether sampling was sufficient to achieve a plausible estimate of the number of species (i.e., species accumulation curves), we expanded the accumulation curve technique for composition data of species and we estimated how many meters in a stream stretch sampled are needed in studies with stream assemblages. Our results show that at least three plots of 15m (i.e., a 45m stretch) are needed for a good representation of the total richness and species composition observed in each stretch. However, no evidence was found that these results change as a function of the degree of forest coverage at each sampled point, showing that the minimum sample effort must be the same regardless of the degree of forest coverage, contradicting our initial hypothesis that environments with degraded riparian forest would require less sampling effort.

Previous studies on sampling effort for stream fish in Brazil have yielded different results from those found here. In Central Amazonia, through analysis of species accumulation curves, the minimum sample defined for first and second order streams was 180 m and 213 m, respectively (Anjos and Zuanon, 2007). This result is about 4-5 times higher than the minimum found here, with a difference of 135 meters for first order streams. In the Atlantic Forest, it was observed that a stretch of 40x the width of the channel is not sufficient to sample all the richness of species in streams of this biome (Terra et al, 2013), which would result in lengths of 60 m (40x the average width of streams analyzed here) based on the width of watercourses analyzed here. Even compared to first to third order streams in North America (South Carolina, the minimum sample found of 370-435m; Paller, 1995), our estimates are much lower. Differences in the minimum sample may be due to the greater species richness found in the sampled plots in the Amazon (Mendonça et al., 2005, Espirito-Santo et al. 2009) and in the Atlantic Forest (Terra et al. 2013), or even to the methodology employed by other authors to be based on visual deductions from accumulation curves rather than on analytical criteria.

The minimum sample used is not related to the degree of land use (i.e., forest cover) used here. The vegetation cover of the streams does not vary considerably in the sampling points analyzed (Freitas, 2019; unpublished data), with five points covered by at least 60%, only one with 30% and four of them with up to 10% of primary or secondary forest. Although negative effects of the removal of riparian vegetation on stream fish assemblages have already been documented (Ilha et al., 2019; Brejão et al., 2018; Casatti et al., 2012; Leitão et al., 2018), Freitas (2019) did not find difference in richness, abundance and species composition related to vegetation cover. This may reflect an absence of vegetation cover-fish relationship, which seems improbable given the widespread knowlegde on this subject, or our control group do not fully represent pristine stream sites. Indeed, logging seem to had been performed in Tanguro farm during the 1970s and 1980s (*MSD, personal observation*), and this could increase the chance of type II errors in our analysis. Further studies should tackle this source of uncertainty.

Studies addressing the minimum sampling effort in fish focus on species richness as the main metric, in which species accumulation curves are intensively used. Using only species accumulation curves and visually estimating minimum efforts based curves limit conclusions about the structure of assemblages and could lead to wrong conclusions. Besides, richness estimates showed more variability between sites than composition estimates (i.e., the standard error for the minimum sampling effort for richness was almost twice the standard error for the minimum for composition). Here, we used not only the species richness, but also the species composition, in order to determine a minimum sampling effort capable of detecting the structure of the fish community in a more accurately way. Species composition has been used less frequently on sampling effort studies. As an example, currently there are methodologies where pseudo standard errors (i.e. *MultiSE*) are used as error measures, calculated from the dissimilarity between samples in multivariate analyzes (e.g. Sorensen, Jaccard, Bray-Curtis; Anderson & Santana-Garcon, 2015; Guerra-Castro et al., 2020). The error calculation is made from similarity/dissimilarity analyzes between simulated samples, taken from parameters of pilot data (i.e., probability of occurrence, density, etc), and after *MultiSE* is tested according to the number of samples. This analysis is similar to that performed here, in which we also used species composition analysis (i.e. Bray-Curtis distance) to calculate error measures (i.e. standard error extracted from the PCOA analysis), thus based on the similarity between the samples, defined a minimum sample. However, our work used regression models to define a minimum sample, whereas in the *MultiSE* analysis finite derivatives are used between the error values and sampling effort, represented by the percentage of precision improvement with the addition of each sample unit (used 10 %, 5% and 2.5%). This last analysis allowed to define three “cutoff points”, here called break values, thus defining a minimally necessary improvement (10%), sub-optimal improvement (5%) and optimal improvement (2.5%), indicating the minimum values to obtain greater precision in multivariate analyzes at communities. Here, to improve estimates of species richness and composition, we recommend that besides the minimum sampling effort be added the standard error.

Although our results showed that a minimum of 45m is required for a good representation of assemblage’ structure, this conclusion must be taken with caution. As shown by Figures 2 and 5, a 45 m stretch would still lead to non-detection of a few species (1-6 species, depending on the stream site), an under-representation of species composition, and, hence, loss of fish fauna information. These losses represent a huge impact on fish estimates as local stream sites usually support a few number of species (the lost two out of 10 species would represent 20% of information loss). To minimize this bias, we recommend the use of our standard error estimates in order to define an upper limit for sampling effort. The minimum sampling effort obtained, for instance, for species composition and richness were 45m, but including the upper bound of the confidence interval would result in 60m (4 sections) and 75m (5 section), respectively. These large river sections would produce more conservative assemblage estimates, and a potential better trade-off between under- and over-sampling.

The present work brought important results regarding the minimum sample to be adopted in studies with fish assemblages and the difference between more and less impacted locations. However, our estimates are based only on a technique for detecting the minimum samples (i.e., segmented regression) and only one break point. Associate other break points (e.g., three), other techniques to this definition (e.g., rate of change in error; Guerra-Castro et al, 2020) or even visual inspection (although subjective), can contribute to a better definition of the size minimum to be sampled. Finally, as our analyzes were restricted to a relatively small sample size (i.e., nine streams), future analyzes should take into account other aquatic organisms, a greater number of points, in other Brazilian biomes, with a more variable impact gradient and mainly in highly diverse areas, for greater clarity in the definition of the minimum sampling estimate.

## CONCLUSION

We therefore conclude that the minimum sample size to obtain more accurate richness and species composition estimates of stream fish assemblages is approximately 45 meters. We emphasize, however, that this result was obtained for a specific area, and caution should be taken when implementing this protocol in areas other than those analyzed here. Furthermore, we suggest that a sampling “maximum” should be also implemented so that richness and composition estimates are even more precise. We also concluded that the minimum sampling effort does not seem to vary depending on the vegetation cover. However, it is important to evaluate further if deforestation or other impacts drive the minimum sampling effort to delineate even better surveys of biological assemblages.

## ACKNOWLEDGEMENTS

We would like to thank the Ecology and Conservation laboratory at the Federal University of Pará (UFPA), for granting access to the data used for this research. We thank the Amaggi group for granting access to their property for researchers to collect the data. We would like to thank the National Council for Scientific and Technological Development - CNPq - for the scholarship granted, which made it possible for this research to be carried out, the National Institutes for Science and Technology (INCT) programme in Ecology, Evolution and Biodiversity Conservation (MCTIC/CNPq: #465610/2014-5), the Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) and the Fundação de Amparo à Pesquisa do Distrito Federal (FAPDF: #00193-00000002/2019-61).

## REFERENCES

Angermeier, P. L., & Schlosser, I. J. (1989). Species area relationship for stream fishes. *Ecology*, *70*(5), 1450-1462. https://doi.org/10.2307/1938204

Angermeier, P. L., & Smogor, R. A. (1995). Estimating number of species and relative abundances in stream-fish communities: effects of sampling effort and discontinuous spatial distributions. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(5), 936-949. https://doi.org/10.1139/f95-093

Anjos, M. B. D., & Zuanon, J. (2007). Sampling effort and fish species richness in small terra firme forest streams of central Amazonia, Brazil. *Neotropical Ichthyology*, *5*(1), 45-52. https://doi.org/10.1590/S1679-62252007000100006

Bergallo, H. G., Esbérard, C. E., Mello, M. A. R., Lins, V., Mangolin, R., Melo, G. G., & Baptista, M. (2003). Bat species richness in Atlantic Forest: what is the minimum sampling effort?. *Biotropica*, *35*(2), 278-288. https://doi.org/10.1111/j.1744-7429.2003.tb00286.x

Bojsen, B. H., & Barriga, R. (2002). Effects of deforestation on fish community structure in Ecuadorian Amazon streams. *Freshwater Biology*, *47*(11), 2246-2260. https://doi.org/10.1046/j.1365-2427.2002.00956.x

Bradford, D. F., Franson, S. E., Neale, A. C., Heggem, D. T., Miller, G. R., & Canterbury, G. E. (1998). Bird species assemblages as indicators of biological integrity in Great Basin rangeland. *Environmental Monitoring and Assessment*, *49*(1), 1-22. https://doi.org/10.1023/A:1005712405487

Brejão, G. L., Hoeinghaus, D. J., Pérez Mayorga, M. A., Ferraz, S. F., & Casatti, L. (2018). Threshold responses of Amazonian stream fishes to timing and extent of deforestation. Conservation biology, 32(4), 860-871. https://doi.org/10.1111/cobi.13061

Callisto, M., Moretti, M., & Goulart, M. (2001). Macroinvertebrados bentônicos como ferramenta para avaliar a saúde de riachos. Revista Brasileira de Recursos Hídricos, 6(1), 71-82.

Casatti, L., Teresa, F. B., Gonçalves-Souza, T., Bessa, E., Manzotti, A. R., Gonçalves, C. D. S., & Zeni, J. D. O. (2012). From forests to cattail: how does the riparian zone influence stream fish?. *Neotropical Ichthyology*, *10*(1), 205-214. https://doi.org/10.1590/S1679-62252012000100020

Castro, R. M. C. (1999). “Evolução da ictiofauna de riachos sul-americanos: padrões gerais e possíveis processos gerais e possíveis processos casuais,” in *Ecologia de Peixes de Riachos*, eds E. P. Caramashi and R. Mazzoni e Peres-Neto (Rio de Janeiro: PPGE Instituto de Biologia), 139–155.

Conroy, M. J., & Noon, B. R. (1996). Mapping of species richness for conservation of biological diversity: conceptual and methodological issues. *Ecological Applications*, *6*(3), 763-773. https://doi.org/10.2307/2269481

de Carvalho, D. R., de Castro, D. M. P., Callisto, M., Moreira, M. Z., & Pompeu, P. S. (2017). The trophic structure of fish communities from streams in the Brazilian Cerrado under different land uses: an approach using stable isotopes. *Hydrobiologia*, *795*(1), 199-217. DOI 10.1007/s10750-017-3130-6

Delabie, J. H., Fisher, B. L., Majer, J. D., & Wright, I. W. (2000). Sampling effort and choice of methods. *Ants: standard methods for measuring and monitoring biodiversity*, 145-154.

Dias, M. S., da Silva Cassemiro, F. A., Bailly, D., Ribeiro, L. S., de Freitas Cossignani, E., da Graça, W. J., ... & Villara, G. T. Macroecologia de peixes de riachos brasileiros. *Oecologia Australis*, *25*(2), 530. https://doi.org/10.4257/oeco.2021.2502.18

Espírito Santo, H. M. V., Magnusson, W. E., Zuanon, J., Mendonca, F. P., & Landeiro, V. L. (2009). Seasonal variation in the composition of fish assemblages in small Amazonian forest streams: evidence for predictable changes. *Freshwater Biology*, *54*(3), 536-548. https://doi.org/10.1111/j.1365-2427.2008.02129.x

Ferreira, F. C., & Petrere, M. (2009). The fish zonation of the Itanhaém river basin in the Atlantic Forest of southeast Brazil. *Hydrobiologia*, *636*(1), 11-34. DOI 10.1007/s10750-009-9932-4

Fleishman, E., Noss, R. F., & Noon, B. R. (2006). Utility and limitations of species richness metrics for conservation planning. *Ecological indicators*, *6*(3), 543-553. https://doi.org/10.1016/j.ecolind.2005.07.005

Fischer, J. R., & Paukert, C. P. (2009). Effects of sampling effort, assemblage similarity, and habitat heterogeneity on estimates of species richness and relative abundance of stream fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(2), 277-290. https://doi.org/10.1139/F08-209

Freemark, K. E., Meyers, M., White, D., Warman, L. D., Kiester, A. R., & Lumban-Tobing, P. (2006). Species richness and biodiversity conservation priorities in British Columbia, Canada. *Canadian Journal of Zoology*, *84*(1), 20-31. https://doi.org/10.1139/z05-172

FREITAS, P. V. Efeito da perda de cobertura florestal sobre a diversidade de peixes de riachos em uma zona de transição Cerrado-Amazônia. Dissertação (Mestrado em Ecologia) – Programa de pós-graduação em Ecologia, Universidade Federal do Pará. Pará, p. 58. 2019.

Ganasan, V., & Hughes, R. M. (1998). Application of an index of biological integrity (IBI) to fish assemblages of the rivers Khan and Kshipra (Madhya Pradesh), India. *Freshwater Biology*, *40*(2), 367-383. https://doi.org/10.1046/j.1365-2427.1998.00347.x

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters*, *4*(4), 379-391. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Goulart, M. D., & Callisto, M. (2003). Bioindicadores de qualidade de água como ferramenta em estudos de impacto ambiental. Revista da FAPAM, 2(1), 156-164.

Ilha, P., Rosso, S., & Schiesari, L. (2019). Effects of deforestation on headwater stream fish assemblages in the Upper Xingu River Basin, Southeastern Amazonia. *Neotropical Ichthyology*, *17*(1). http://dx.doi.org/10.1590/1982-0224-20180099

Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Jenkins, M. (2003). Prospects for biodiversity. Science, 302(5648), 1175-1177. DOI: 10.1126/science.1088666

Jézéquel, C., Tedesco, P. A., Darwall, W., Dias, M. S., Frederico, R. G., Hidalgo, M., ... & Oberdorff, T. (2020). Freshwater fish diversity hotspots for conservation priorities in the Amazon Basin. Conservation biology, 34(4), 956-965. https://doi.org/10.1111/cobi.13466

Lowrance, R., Altier, L. S., Newbold, J. D., Schnabel, R. R., Groffman, P. M., Denver, J. M., ... & Staver, K. W. (1997). Water quality functions of riparian forest buffers in Chesapeake Bay watersheds. *Environmental Management*, *21*(5), 687-712. https://doi.org/10.1007/s002679900060

Macedo, D. R., Hughes, R. M., Ligeiro, R., Ferreira, W. R., Castro, M. A., Junqueira, N. T., ... & Callisto, M. (2014). The relative influence of catchment and site variables on fish and macroinvertebrate richness in cerrado biome streams. *Landscape Ecology*, *29*(6), 1001-1016. https://doi.org/10.1007/s10980-014-0036-9

Macedo, M. N., Coe, M. T., DeFries, R., Uriarte, M., Brando, P. M., Neill, C., & Walker, W. S. (2013). Land-use-driven stream warming in southeastern Amazonia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1619), 20120153.doi:10.1098/rstb.2012.0153.

Matthews, T. J., Rigal, F., Triantis, K. A., & Whittaker, R. J. (2019). A global model of island species–area relationships. *Proceedings of the National Academy of Sciences*, *116*(25), 12337-12342. https://doi.org/10.1073/pnas.1818190116

Miserendino, M. L., Casaux, R., Archangelsky, M., Di Prinzio, C. Y., Brand, C., & Kutschker, A. M. (2011). Assessing land-use effects on water quality, in-stream habitat, riparian ecosystems and biodiversity in Patagonian northwest streams. *Science of the Total Environment*, *409*(3), 612-624. https://doi.org/10.1016/j.scitotenv.2010.10.034

Macedo, M. N., Coe, M. T., DeFries, R., Uriarte, M., Brando, P. M., Neill, C., & Walker, W. S. (2013). Land-use-driven stream warming in southeastern Amazonia. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 368(1619), 20120153. doi:10.1098/rstb.2012.0153

Montag, L. F., Winemiller, K. O., Keppeler, F. W., Leão, H., Benone, N. L., Torres, N. R., ... & Juen, L. (2019). Land cover, riparian zones and instream habitat influence stream fish assemblages in the eastern Amazon. *Ecology of Freshwater Fish*, *28*(2), 317-329. https://doi.org/10.1111/eff.12455

Muggeo, V. M. (2003). Estimating regression models with unknown break points. *Statistics in medicine*, *22*(19), 3055-3071. https://doi.org/10.1002/sim.1545

Nichols, J. D., Boulinier, T., Hines, J. E., Pollock, K. H., & Sauer, J. R. (1998). Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation biology, 12(6), 1390-1398. https://doi.org/10.1111/j.1523-1739.1998.97331.x

Oberdorff, T., Dias, M. S., Jézéquel, C., Albert, J. S., Arantes, C. C., Bigorne, R., ... & Zuanon, J. (2019). Unexpected fish diversity gradients in the Amazon basin. Science advances, 5(9), eaav8681. DOI: 10.1126/sciadv.aav8681

Paller, M. H. (1995). Relationships among number of fish species sampled, reach length surveyed, and sampling effort in South Carolina coastal plain streams. *North American*

*Journal of Fisheries Management*, *15*(1), 110-120. https://doi.org/10.1577/1548-8675(1995)015%3C0110:RANOFS%3E2.3.CO;2

Pessanha, A. L. M., Araujo, F. G., De Azevedo, M. C. C., & Gomes, I. D. (2003). Diel and seasonal changes in the distribution of fish on a southeast Brazil sandy beach. *Marine Biology*, *143*(6), 1047-1055. https://doi.org/10.1007/s00227-003-1138-0

Peterson, J. T., & Rabeni, C. F. (1995). Optimizing sampling effort for sampling warmwater stream fish communities. North American Journal of Fisheries Management, 15(3), 528-541. https://doi.org/10.1577/1548-8675(1995)015%3C0528:OSEFSW%3E2.3.CO;2

Portz, L., Dias, C. T. D. S., & Cyrino, J. E. P. (2000). Regressão segmentada como modelo na determinação de exigências nutricionais de peixes. Scientia Agricola, 57(4), 601-607. http://dx.doi.org/10.1590/S0103-90162000000400002

Ricketts, T. H., Dinerstein, E., Olson, D. M., & Loucks, C. (1999). Who's where in North America? Patterns of species richness and the utility of indicator taxa for conservation. BioScience, 49(5), 369-381. https://doi.org/10.2307/1313630

Schneider, M., Aquino, P. D. P. U. D., Silva, M. J. M., & Fonseca, C. P. (2011). Trophic structure of a fish community in Bananal stream subbasin in Brasília National Park, Cerrado biome (Brazilian Savanna), DF. *Neotropical Ichthyology*, *9*(3), 579-592. http://dx.doi.org/10.1590/S1679-62252011005000030

Silva, J. C., Gubiani, É. A., Piana, P. A., & Delariva, R. L. (2016). Effects of a small natural barrier on the spatial distribution of the fish assemblage in the Verde River, Upper Paraná River Basin, Brazil. Brazilian Journal of Biology, 76, 851-863. https://doi.org/10.1590/1519-6984.01215

Smith, K. L., & Jones, M. L. (2005). Watershed-level sampling effort requirements for determining riverine fish species composition. Canadian Journal of Fisheries and Aquatic Sciences, 62(7), 1580-1588. https://doi.org/10.1139/f05-098

SoberónM, J., & LlorenteB, J. (1993). The use of species accumulation functions for the prediction of species richness. *Conservation biology*, *7*(3), 480-488. https://doi.org/10.1046/j.1523-1739.1993.07030480.x

Tank, J. L., Rosi-Marshall, E. J., Griffiths, N. A., Entrekin, S. A., & Stephen, M. L. (2010). A review of allochthonous organic matter dynamics and metabolism in streams. *Journal of the North American Benthological Society*, *29*(1), 118-146. https://doi.org/10.1899/08-170.1

Taylor, A. F. (2002). Fungal diversity in ectomycorrhizal communities: sampling effort and species detection. In *Diversity and integration in Mycorrhizas* (pp. 19-28). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-1284-2_3

Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., ... & Oberdorff, T. (2017). A global database on freshwater fish species occurrence in drainage basins. Scientific data, 4(1), 1-6. https://doi.org/10.1038/sdata.2017.141

Terra, F.B., Hughes, R. M., & Araújo, F. G. (2013). Sampling sufficiency for fish assemblage surveys of tropical Atlantic forest streams, Southeastern Brazil. Fisheries, 38(4), 150-158. https://doi.org/10.1080/03632415.2013.775572

Toms, J. D., & Lesperance, M. L. (2003). Piecewise regression: a tool for identifying ecological thresholds. *Ecology*, *84*(8), 2034-2041. https://doi.org/10.1890/02-0472

Torrente Vilara, G., Zuanon, J., Leprieur, F., Oberdorff, T., & Tedesco, P. A. (2011). Effects of natural rapids and waterfalls on fish assemblage structure in the Madeira River (Amazon Basin). Ecology of Freshwater Fish, 20(4), 588-597. https://doi.org/10.1111/j.1600-0633.2011.00508.x

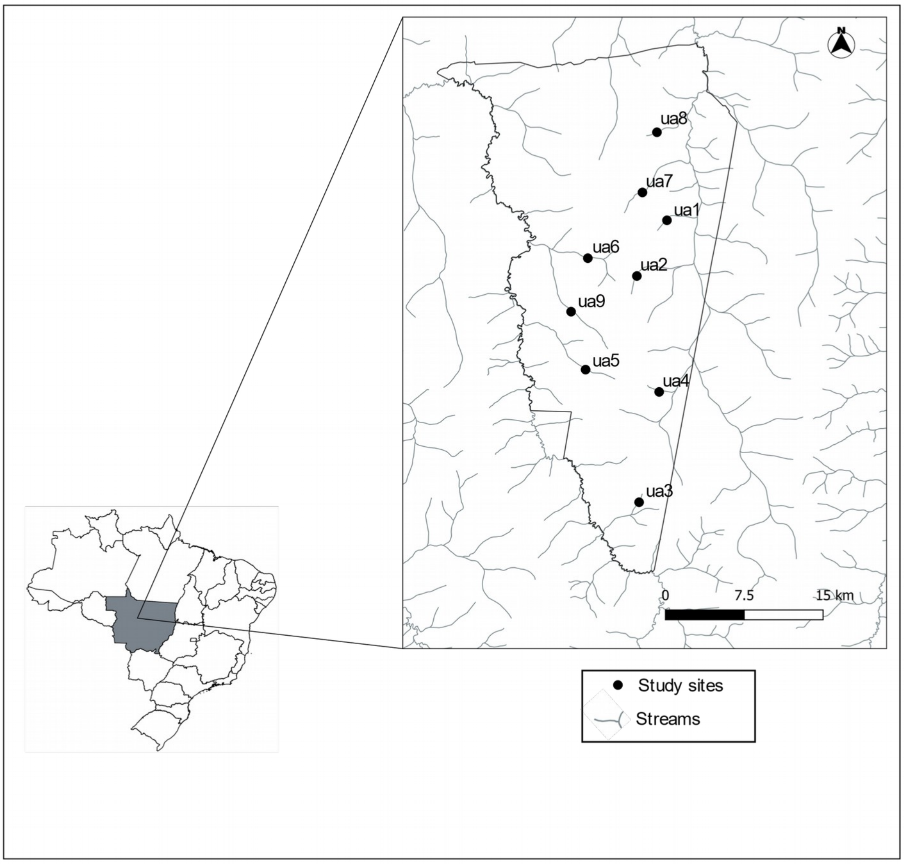
Triantis, K. A., Mylonas, M., Lika, K., & Vardinoyannis, K. (2003). A model for the species–area–habitat relationship. *Journal of Biogeography*, *30*(1), 19-27. https://doi.org/10.1046/j.1365-2699.2003.00805.x

Ugland, K. I., Gray, J. S., & Ellingsen, K. E. (2003). The species–accumulation curve and estimation of species richness. *Journal of Animal Ecology*, *72*(5), 888-897. https://doi.org/10.1046/j.1365-2656.2003.00748.xWalther, B. A., & MARTIN, J. L. (2001). Species richness estimation of bird communities: how to control for sampling effort?. *Ibis*, *143*(4), 413-419. https://doi.org/10.1111/j.1474-919X.2001.tb04942.x

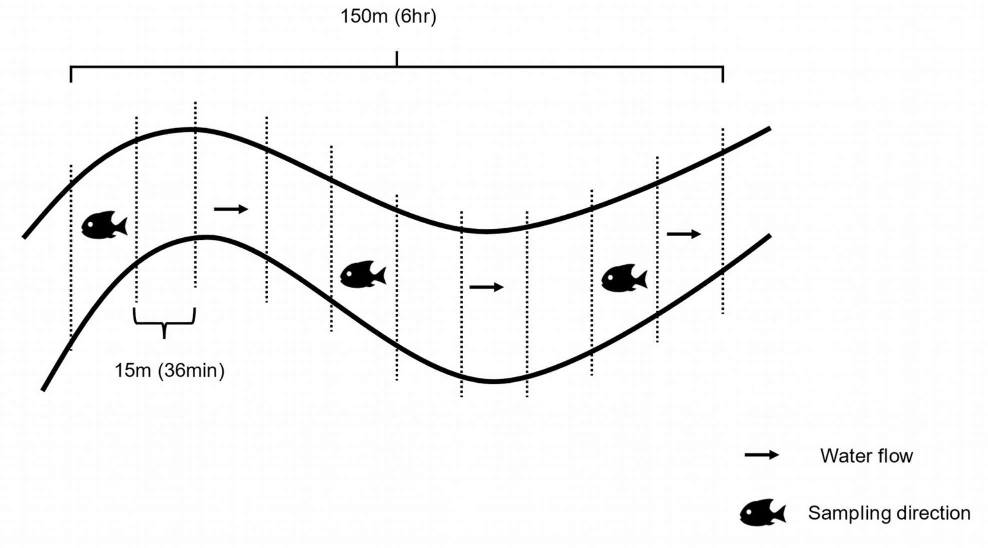
Walther, B. A., Cotgreave, P., Price, R. D., Gregory, R. D., & Clayton, D. H. (1995). Sampling effort and parasite species richness. *Parasitology today*, *11*(8), 306-310. https://doi.org/10.1016/0169-4758(95)80047-6

Xuan Mao, C., Colwell, R. K., & Chang, J. (2005). Estimating the species accumulation curve using mixtures. *Biometrics*, *61*(2), 433-441. https://doi.org/10.1111/j.1541-0420.2005.00316.x

# Attachment



**Attachment 1.** Tanguro farm area, located in Querência (MT) and the sample sites.



**Attachment 2.** Scheme of the sampling design adopted for the collection of fish in the streams. The sampling effort for the entire 150m stretch was six hours, divided into 36min for each 15m plot (adapted from Freitas et al., 2019).

## Conclusão geral

Como resultados, o mínimo amostral necessário para se obter uma amostragem representativa das assembleias de peixes nos riachos amostrados foi de ~45 m, tanto para estimativas de riqueza quanto para composição de espécies. Para riqueza, a média de número de parcelas de 15 m necessárias para se obter estimativas mais precisas foi de 3.07± 0.79, para composição foi de 2.98 (± 0.39). Em relação à variação do mínimo amostral em função do nível de cobertura florestal, não obtivemos resultados significativos. Ou seja, o mínimo amostral para áreas mais impactadas (i.e., com maior supressão vegetal) é o mesmo que para áreas menos impactadas.

Concluímos, portanto, que o tamanho amostral mínimo para se obter estimativas de riqueza e composição de espécies mais precisas de assembleias de peixes de riacho é de aproximadamente 45 metros. Reforçamos, porém, que este resultado foi obtido para uma área específica, e que deve-se ter cautela ao implementar este protocolo em áreas diferentes das analisadas aqui. Além disso, nós sugerimos que seja implementado a esse mínimo um “máximo” amostral, que seria a média multiplicada por duas vezes o erro padrão, para que as estimativas sejam ainda mais precisas. Fazendo isto, há maior segurança nas estimativas, não amostrando pouco ou ultrapassando um limite onde implementar mais amostras traria um desperdício de recursos. Também concluímos que o mínimo de esforço amostral parece não variar em função da cobertura vegetal, porém, é importante incluir os diferentes impactos presentes nos locais de estudos, para que os resultados obtidos com as amostragens estejam representando a realidade, e não um ruído de uma amostragem enviesada.