

# Riparian forest cover impact on aquatic diversity identified by metabarcoding

*A small-scale experimental study in eastern Amazonia*

Rita Cabilan Just Olsen



Thesis submitted for the degree of  
Master of Science in Biology  
60 credits

Institute of Biosciences  
Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

Winter 2022



# **Riparian forest cover impact on aquatic diversity identified by metabarcoding**

*A small-scale experimental study in eastern Amazonia*

Rita Cabilan Just Olsen

ritaolsen@msn.com

## **Supervisors:**

Hugo de Boer

hugo.deboer@nhm.uio.no

Jonathan Ready

jonathan.ready@gmail.com

Marcele Laux

marlaux@outlook.com

Master in Bioscience: Biodiversity and systematics

Department of Biosciences

University of Oslo

© Rita Cabilan Just Olsen

2022

Riparian forest cover impact on aquatic diversity identified by metabarcoding

Rita Cabilan Just Olsen

<http://www.duo.uio.no/>

Printed: Reprosentralen, University of Oslo

# Acknowledgments

I would like to thank my forever supporting supervisors, Hugo de Boer, Jonathan Ready, and Marcele Laux. Thank you for the words of encouragement, feedback, and guidance. I am incredibly grateful for this experience, being given the opportunity to work in Brazil was a dream come true, and all of you made this possible. Thank you for not giving up on me during the hardships of the pandemic. I would also like to thank Quentin, who played a big supporting role in assisting with statistics, proofreading, and motivation.

Thank you to the people at the Natural History Museum in Oslo, it was great to be a part of such a fun and educational environment. Having lunch and gatherings with you was always something I was looking forward to. To the fantastic people at the DNA lab, Jarl Andreas, Lisbeth, and Audun, who helped and guided me through the laboratory work for this project. You made me realize I enjoy lab work, gave me confidence in my skills, and made me always feel welcome.

A special thanks to the people at the GIBI lab for helping me during field and laboratory work in Brazil. I will always be grateful for the long days spent in the field, exploring the eastern Amazonian forests, experiencing so many new species, and most of all, gaining new friends. Cintia, Fabricio, Alan, Jessica, Ángela and Clarisse, you are the best.

I am forever grateful for my friends, you make me enjoy life, both during and outside of my studies. I wish everybody could have as good friends as me.

To my big family, I am grateful for each and every one of you. Thank you for always supporting and believing in me. To my dad, I would not be where I am today without you, for which I am forever grateful.

And last but not least, Sondre, there is so much to thank you for. The endless support and conversations about our projects have been invaluable and critical for this thesis. Experiencing life with you is great, whether it is catching butterflies, discovering new hobbies, or just everyday life.



# Abstract

Fragments of riparian forests in agricultural landscapes, due to their exposure to abiotic and anthropogenic factors, experience a higher rate of deforestation which leads to a net loss of species in the area. To reduce the impact on biodiversity and freshwater ecosystem services, monitoring and knowledge of how deforestation affects the biodiversity in these forest fragments are necessary. This can ensure that policies are being followed and prevent further damage to this important ecosystem. Environmental DNA (eDNA) metabarcoding is a promising tool in biodiversity assessments and monitoring. It is effective, reliable, and less time-consuming compared to conventional methods. This exploratory study aims to assess the aquatic biodiversity of arthropods and fish in a stream of a fragmented riparian forest using eDNA metabarcoding and to investigate the differences in abundance between the forested and deforested areas.

To do this, eDNA samples of bulk aquatic arthropods and water samples were obtained using two types of nets for the arthropods and Sterivex filters for the free-floating DNA from fish. Samples were collected in September and October 2019 in Paragominas, Pará state, Brazil. PCR amplification, of the extracted DNA, was done using two metabarcoding primer sets. A COI primer for arthropod detection from the bulk samples and a 12S primer was used for detecting fish taxa from the water eDNA samples. We identified ten distinct orders for arthropods in 102 Amplicon Sequence Variants (ASVs) from the bulk samples and six distinct orders for fish in 62 ASVs were detected in the eDNA water samples.

Here we show that DNA metabarcoding can be used to detect the aquatic biodiversity of arthropod orders and fish families. The results show that the canopy cover affected the total abundance for both datasets and that the differences between the forested and deforested areas for arthropods were statistically significant, however, we found no evidence for this in the fish community. The estimated richness and relative abundance suggest that there was a difference in the arthropod community composition between sampling points and the opposite was found for the fish community composition. eDNA metabarcoding can provide invaluable information on how canopy cover affects the aquatic biodiversity found in riparian forest fragments.

*Keywords:* Metabarcoding, eDNA, Freshwater ecosystems, Forest fragmentation, Brazil, Fish, Arthropods





# Table of contents

Riparian forest cover impact on aquatic diversity identified by metabarcoding.....	III
Acknowledgments.....	V
Abstract.....	VII
Table of contents.....	IX
1 Introduction.....	1
1.1 Water eDNA and bulk aquatic arthropod metabarcoding.....	2
1.2 DNA markers.....	5
2 Aims of the study.....	7
3 Materials and methods.....	9
3.1 Study site.....	9
3.2 Sample collection.....	11
3.3 DNA extraction, amplification, and sequencing.....	13
3.3.1 DNA extraction.....	13
3.3.2 DNA amplification.....	14
3.3.3 Library preparation and Sequencing.....	15
3.4 Data analysis.....	17
3.4.1 Bioinformatics.....	17
3.4.2 Statistical analysis.....	18
3.4.3 Stream community analysis.....	18
4 Results.....	21
4.1 Alpha diversity.....	21
4.2 Relative abundance.....	23
4.3 Canopy cover.....	26
5 Discussion.....	28
5.1 DNA metabarcoding method.....	28
5.2 Ecological patterns.....	30
5.2.1 Richness and relative abundance.....	30
5.2.2 Differences between forest exterior and interior.....	31
5.3 Perspectives.....	33

6	Conclusion.....	34
	References .....	35
	Appendix A – Summary of primers .....	44
	Appendix B – Lab protocols .....	46
	Appendix C – Scripts used in R Studio.....	52
	Scripts for arthropods bulk samples .....	52
	Scripts for water eDNA samples .....	56

# 1 Introduction

Deforestation affects biological diversity in many ways. Amongst them are habitat destruction, distance to continuous forest, and edge effect within the boundary zone between deforested areas and forest (Collinge, 1996; Fearnside, 2005). The forest edge is more exposed to a range of biological and physical effects, invasive species, and hunters and this leads to a net loss of species in the area (Skole and Tucker, 1993; Fahrig, 2003). Consequences of fragmentation may have long-term effects on populations via alterations of ecological processes such as predation, pollination, feeding habits, and territorial behaviours, in addition to immediate local extinction (Andren and Angelstam, 1988; Fahrig, 2003). Species extinction greatly impacts human health and sustainability. There is an international political agreement to halt the current biodiversity loss, even though the knowledge on biodiversity is limited and numerous taxa and geographical regions remain un-described (Díaz *et al.*, 2006; McCluskey *et al.*, 2017). Species loss as a result of fragmentation in tropical rainforests is especially high due to large species diversity and complex ecological structure (Turner, 1996).

To prevent further increases in the deforestation rate, the Brazilian government has established numerous policies, laws, and regulations concerning the recovery of degraded areas (Nunes *et al.*, 2015; Gastauer *et al.*, 2019; da Cruz *et al.*, 2022). Environmental Liabilities oblige companies to repair the sum of the environmental damage caused by them (da Cruz *et al.*, 2022). Forest restoration with a focus on Riparian forests has been prioritized in environmental policies due to their ecological importance in ensuring processes for freshwater ecosystem functioning, connecting forest fragments, maintaining biodiversity, and protecting against anthropogenic pressures (Firmiano *et al.*, 2021; da Cruz *et al.*, 2022). A study expanding over 22 years shows that restriction in land-use change in riparian zones did not protect the riparian forest more effectively compared to non-riparian forests, however evidence of higher deforestation rate in riparian areas was found (Nunes *et al.*, 2015).

Traditional diversity monitoring has heavily relied on the physical identification of species that have been mass collected, but the performance of these techniques has in some cases been proven challenging, for example, phenotypic plasticity and species with very similar appearance in juvenile stages in closely related species are difficult to detect (Thomsen and Willerslev, 2015). Furthermore, some traditional monitoring techniques are destructive for the species or ecosystem under study and taxonomic expertise is required for morphological

identification, which is often lacking (Basset *et al.*, 2012; Watts *et al.*, 2019). This created a demand for alternative approaches (Thomsen and Willerslev, 2015).

The primary source of Neotropical biodiversity is Amazonia (Antonelli *et al.*, 2018), which includes many unidentified organismal groups. Identification of biodiversity is not possible if the species are new to science and not yet formally described. Sequencing diversity enables us to assess the molecular diversity of taxa, without knowing exactly each species from a specimen, which offers a powerful way to survey biodiversity (Taberlet *et al.*, 2012; Ji *et al.*, 2013). Climate change and anthropogenic pressures are increasingly threatening the aquatic biodiversity (Lambertini, 2020; Beermann *et al.*, 2021), and the speed at which we are describing unknown biodiversity in Amazonia is eclipsed by the speed at which forest is being destroyed and lost forever (Fearnside, 2005; Nessimian *et al.*, 2008). This means that in order to ever get a quantitative assessment of Amazonian biodiversity before it is lost is by taking molecular snapshots through eDNA methods (Taberlet *et al.*, 2012). Therefore, this study will test the DNA metabarcoding approach, using water environmental DNA (eDNA) and bulk samples, for use in aquatic biodiversity assessment of riparian forest fragments.

## **1.1 Water eDNA and bulk aquatic arthropod metabarcoding**

Environmental DNA metabarcoding has been recognized as a powerful tool in biodiversity monitoring and is already used to study freshwater ecosystems (Lacoursière-Roussel, Rosabal and Bernatchez, 2016; Olds *et al.*, 2016; Taberlet *et al.*, 2018; Bylemans *et al.*, 2019; Turunen *et al.*, 2019; Johnsen *et al.*, 2020; Beermann *et al.*, 2021; Wang *et al.*, 2021). eDNA metabarcoding is a method used in molecular ecology and environmental genomics to identify species in a community using a short, standardised region of genetic material. It involves collection of environmental samples, which are samples collected from the environment, such as water, sediment, soil, air, or bulk samples that contain a mix of different species. Then amplifying a specific region of DNA from the samples using Polymerase Chain Reaction (PCR), followed by High-Throughput Sequencing (also known as Next Generation Sequencing or NGS) to identify the species present in the sample (Figure 1.1). The resulting sequencing data is then compared to a reference database of known species to identify the species from the environmental sample (Taberlet *et al.*, 2018; Sawaya *et al.*, 2019). Analysis of the results can provide information about species diversity, ecological processes and population dynamics of

complex communities found in ecosystems (Thomsen *et al.*, 2012; Ji *et al.*, 2013; Thomsen and Willerslev, 2015; Valentini *et al.*, 2016; Taberlet *et al.*, 2018).

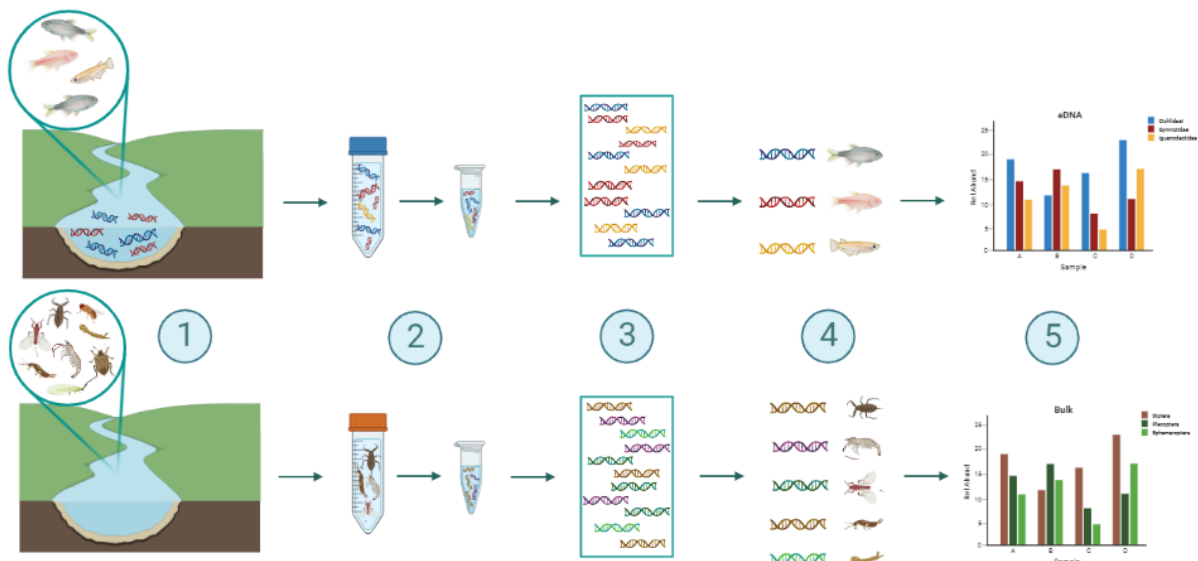


Figure 1. 1. An example of Bulk and water eDNA metabarcoding workflow. 1: Sampling from freshwater. 2: Extraction and amplification. 3: Library preparation and High-throughput sequencing. 4: Bioinformatics and taxonomic assignment. 5. Statistical analysis. Created with BioRender.

## Why study fish?

There are several reasons why it is important to study fish. Studying fish can provide insight into the functioning of aquatic ecosystems, which are critical for the health of the planet and the well-being of human communities (Forio and Goethals, 2020; Lambertini, 2020). Fish are key components of these ecosystems, as they occur in a wide range of trophic levels and play a role in the balance of the ecosystem (Harris, 1995; Olds *et al.*, 2016). Additionally, they are an important food source for humans and are vital in many cultures and economies (Lynch *et al.*, 2016; Liu, Bailey and Davidsen, 2019). Furthermore, studying fish can provide information about the health of aquatic environments, as changes in their populations can be early warning signs of environmental degradation and climate change (Dudgeon *et al.*, 2006; Thomsen *et al.*, 2012; Lambertini, 2020). Freshwater fish can efficiently be detected with environmental DNA (eDNA) compared to traditional approaches, which often rely on invasive methods (Valentini *et al.*, 2016; Taberlet *et al.*, 2018). The free floating eDNA derived from skin cells, faeces, mucus or urine, can easily be collected by water filtration (Taberlet *et al.*, 2012).

## Why study arthropods?

Arthropods are important components of freshwater ecosystems and play a variety of roles in these environments. In these ecosystems, they serve as both predator and prey, and their presence can provide valuable information about the health of the ecosystem (Sundar *et al.*, 2020). For example, the abundance of certain arthropod species can be an indication of water quality (Ab Hamid and Salmah, 2011; Hamid and Md Rawi, 2014). Additionally, arthropods are an important food source for many aquatic species, including fish, and studying them can provide insight into the food web of the ecosystem (Cooper, Walde and Peckarsky, 1990). Overall, studying arthropods in aquatic environments can provide valuable information about the functioning of these ecosystems, the interaction between different species, and responses to stressors introduced by increased anthropogenic pressure (Birk *et al.*, 2012; Mondy *et al.*, 2012; Daam and Rico, 2018; Beermann *et al.*, 2021). eDNA metabarcoding can be used as a tool for macroinvertebrate monitoring, often by analysis of bulk samples, where the genomic information is gained from a mix of different taxa (Elbrecht and Leese, 2017; Fernández *et al.*, 2018; Taberlet *et al.*, 2018; Watts *et al.*, 2019). Arthropod bulk samples are environmental samples that contain many individuals of invertebrates collected with a trap such as Malaise, nets, pitfalls or others (Ji *et al.*, 2013; Taberlet *et al.*, 2018). eDNA metabarcoding of bulk samples allows for the identification of a large number of species in the sample, as it is based on the analysis of genetic material rather than physical characteristics, which makes it less time-consuming compared to conventional methods (Ji *et al.*, 2013). Studies show that eDNA bulk samples are as efficient, if not more, in detection of taxonomic information and richness (alpha diversity) estimations (Yu *et al.*, 2012; Ji *et al.*, 2022).

## 1.2 DNA markers

In metabarcoding analyses, the choice of primer can have a significant impact on the end results and is important because it determines which specific region of DNA will be amplified, and therefore which organism will be detected in the final analysis (Deiner *et al.*, 2017; Taberlet *et al.*, 2018).

Several factors need to be considered when choosing a primer set for a metabarcoding analysis. First, the primer should be specific to the organisms (or groups) that are targeted for detection (Elbrecht and Leese, 2017). Second, the primer should have good amplification efficiency, meaning that it can readily amplify the target DNA without introducing errors or biases (Braukmann *et al.*, 2019; Hajibabaei *et al.*, 2019). Finally, the primer should be compatible with the PCR conditions that will be used, such as the temperature and the type of PCR enzyme and have access to an adequate number of reference sequences (Taberlet *et al.*, 2018). With these considerations in mind, we chose the COI marker for amplification of aquatic arthropods in the bulk samples and a marker that amplifies the 12S gene in fish from the DNA extracted from the water eDNA samples. Both markers have standardized protocols, sufficient sequence reference databases and are validated for freshwater ecosystem biodiversity assessments (Elbrecht and Leese, 2017; Taberlet *et al.*, 2018). The COI primer pair BF2 (forward) and BR2 (reverse) targets cytochrome c oxidase I (COI) and have been proven to be consistent and dependable when studying freshwater macroinvertebrates compared to previously tested markers (Elbrecht and Leese, 2017; Elbrecht *et al.*, 2017). The Tele03 primer pair targets the 12S mitochondrial rDNA for detection of teleost fishes and has excellent taxonomic resolution (Taberlet *et al.*, 2018). Henceforth, the decision of two different DNA metabarcoding markers for each sample type were made.





## 2 Aims of the study

The use of eDNA metabarcoding will allow for the efficient and accurate assessment of aquatic biodiversity in a fragmented riparian forest. By targeting specific genetic markers, such as the COI and the 12S mitochondrial rDNA genes, we will be able to identify a wide range of aquatic organisms, including arthropods and fish. We expect to find higher diversity and abundance of aquatic organisms in intact areas of the riparian forest, compared to areas that have been disturbed by human activities such as agriculture. This information will be useful for understanding the impacts of fragmentation on aquatic ecosystems and for conservation. To shed light in this topic we want to investigate:

- I. If the metabarcoding method is a viable option for making these assessments?
- II. Does the community composition change between the sampling points?
- III. Does the relative abundance of taxa differ between the forest exterior and interior?
- IV. How do the patterns of diversity and relative abundance for fish compare to those of macroinvertebrates?



# 3 Materials and methods

## 3.1 Study site

The study was conducted in the municipality of Paragominas, State of Pará in the eastern Brazilian Amazon (2°47'20.65"S, 47°31'29.91"W) (Figure 3.1.1), known as an old agricultural frontier and including since 2007 an active bauxite mining area (Osis, Laurent and Pocard-Chapuis, 2019) The municipality of Paragominas has been experiencing rapid land use change since its foundation in 1965, including large increases in deforestation rates (from 12% in 1984 to 45% in 2020) because of conversion to pasture and agriculture (da Cruz *et al.*, 2022). In recent decades, environmental policies have been focusing on the restoration and/or conservation of riparian forests (Riparian zones) (Fremier *et al.*, 2015; de la Fuente *et al.*, 2018).

In this region, the rainfall is seasonal with an average annual precipitation of 1.761 mm, from 1991 to 2021, and the average temperature is about 26.5 °C (<https://pt.climate-data.org/americas-do-sul/brasil/para/paragominas-714961/>). The vegetation consists of evergreen, lowland rainforest with a variety in elevation where wide plateaus of clay soil (Belterra clay) are absent of water courses, and bottom valleys are covered by loamy sands and have a dense network of streams (Verissimo *et al.*, 1992; Laurent *et al.*, 2017). The soil is mostly composed of latossols, also known as Ferrasols, which is a type of soil that is deeply weathered and usually have a red or yellow coloration in the humid tropics (Laurent *et al.*, 2017; Schlesinger and Bernhardt, 2020).

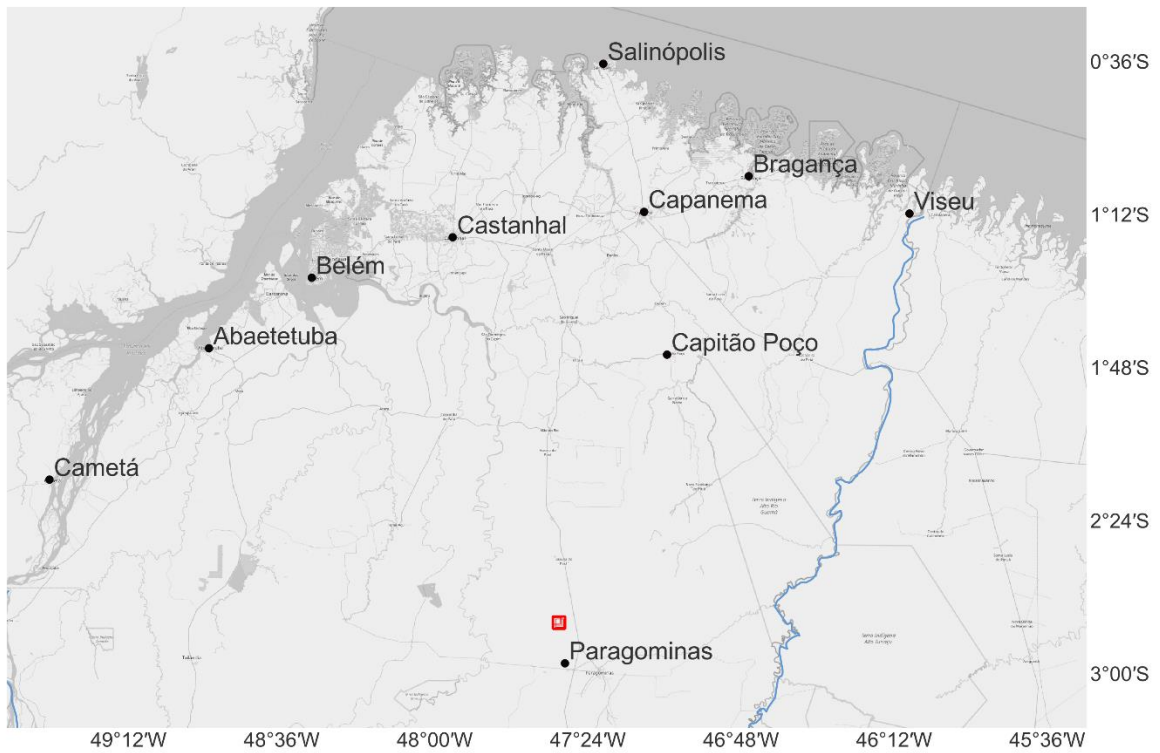
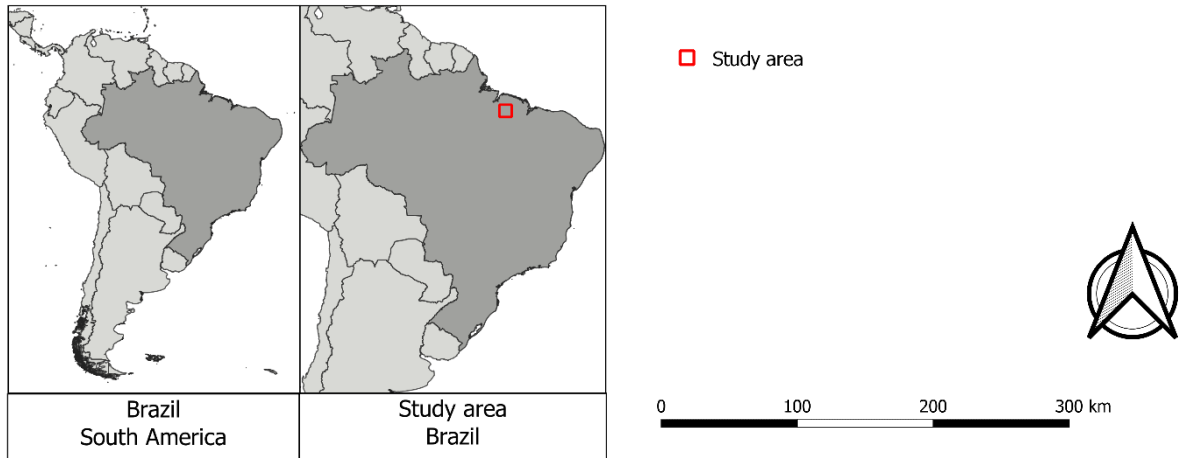


Figure 3.1. 1. Map of study area in Paragominas, State of Pará, Eastern Brazil, SA. Study site marked by the red square.

## 3.2 Sample collection

The sampling period occurred between September and October 2019. Both water eDNA and aquatic arthropod bulk samples were collected from five points (Figure 3.2.1) located in a stream, running through a riparian forest fragment, spanning over approximately 2 km. The sampling points were between 350 – 700 meters from each other. Collection of samples started in Point 1 and continued upstream, to Point 5, to avoid downstream contamination. Points 1 and 5 represent areas with degraded riparian forest with low canopy cover, while points 2 - 4 represent areas with predominantly intact riparian forest with high canopy cover.

Aquatic arthropod bulk samples were collected using a fine mesh seine net (Figure 3.2.2; C&D) and a sieve (peneira feijão), in order to maximise the amount of microhabitats and number of species sampled on each site. To standardize sampling effort between locations, both the seine net and sieve were used five times at each point from recognizable microhabitats (i.e. one sample equals one net haul and one sieving from one microhabitat). Microhabitat types included macrophyte beds, leaf litter, margin with roots, sand bank, and algae. The bulk samples were placed in 50 mL Falcon tubes and preserved in 96% ethanol until extraction.

eDNA was collected by filtering 1 litre of water through Sterivex (Millipore) 0.45 µm filters (Figure 3.2.2; B), manually with the support of a 60 mL luer-lock syringe. Three different filters were used at the water surface level at each collection point to sample from the streams left, middle, and right sides. The 15 eDNA samples were preserved with lysis buffers (Longmire's solution) and placed in a cooling box for transportation and stored in a freezer at -20C until extraction.

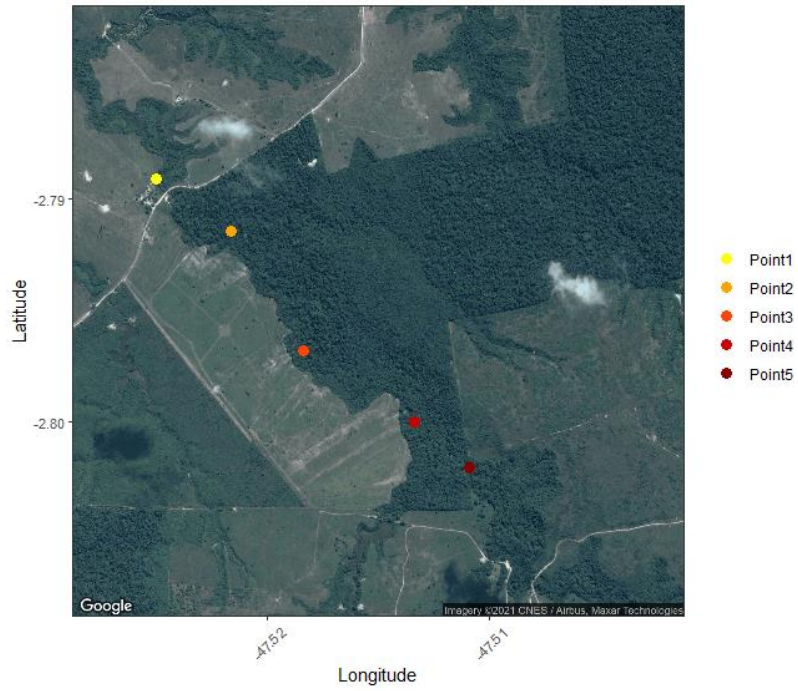


Figure 3.2. 1. Study site with sampling locations in the stream. Sample points from 1-5 upstream. Point 1 and Point 5 represent areas with lower canopy cover. Points 2-4 represent areas with intact riparian forest.



Figure 3.2. 2. Materials used to collect samples on site of sampling location. A: Samples collected in the field. B: Sterivex filter. C: Arthropoda bulk samples collected from net haul. D: Net haul from litter bank.

## **3.3 DNA extraction, amplification, and sequencing**

### **3.3.1 DNA extraction**

All samples were extracted at the laboratories of Instituto Tecnológico Vale and GIBI in Belém, state of Pará, Brazil.

For all 49 bulk arthropod samples, we homogenized the samples using a blender and obtained four technical replicates (Figure 3.3.1: Top) (i.e. four, 0.5 ml each, subsamples from the liquefied original sample). The blender cup was decontaminated using a 20 % bleach bath and subsequent washes with ultrapure water between samples. Negative controls were included between each sample to confirm that the decontamination process used on the liquefier was effective. This was done by activating the blender with only ultrapure water in it (after the decontamination washes) and taking a sample from this blended water. The downstream analysis treated each sub-sample separately, yielding 245 samples. DNA was extracted from the samples by using a protocol following a CTAB and CIAA (Chloroform Isoamyl Alcohol) method (Doyle and Doyle, 1987).

DNA extraction from the 15 eDNA samples were performed using a modified Qiagen DNeasy PowerWater Sterivex Kit (Qiagen, UK), one extraction control was included thus yielding 16 samples. We did not have access to a QIAvac Vacuum Systems (Qiagen, UK) so the modification was using a centrifuge instead of a vacuum machine, putting a small volume in Eppendorf tubes. This process was repeated until all reagents passed through the membrane. Since the DNA passed through Sterivex membrane we decided not to include any technical replicates for the eDNA samples (Figure 3.3.1: Bottom).

All extracted samples were transported, following required legislation (Brazilian SISBIO and SISGEN systems as a part of shared partnership projects (CAPES-SIU, BRC 16/19), to Norway for amplification, sequencing, and analysis.

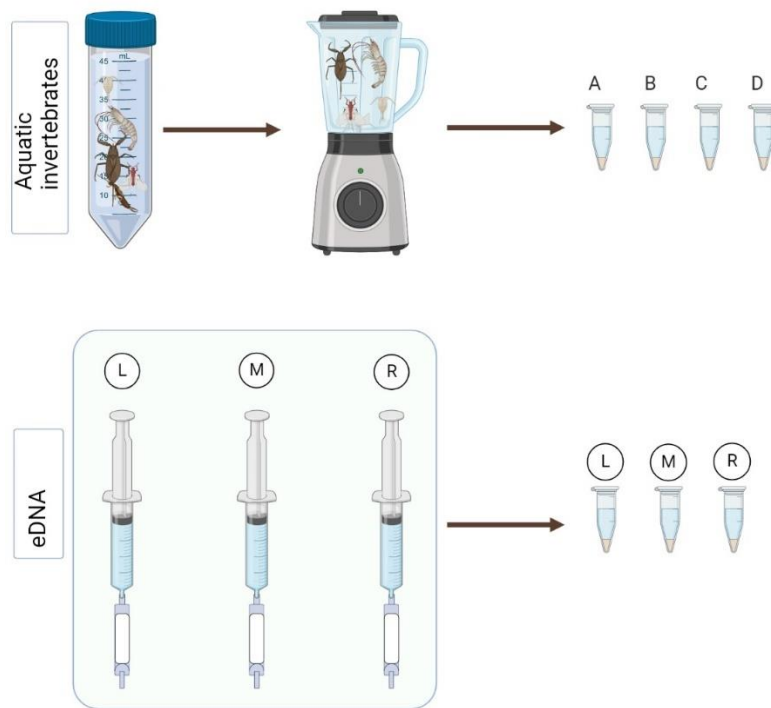


Figure 3.3.1. Extraction of DNA from samples. Top: bulk aquatic arthropod samples and replicates. Four sub-samples from each sample. Bottom: eDNA from Sterivex water filtrates. Created with BioRender.

### 3.3.2 DNA amplification

Polymerase Chain Reaction (PCR) was used to amplify DNA from the 248 bulk aquatic arthropod samples (49 x 4 = 196 bulk samples + 49 extraction controls + 3 PCR controls) and from the 17 water eDNA samples (15 eDNA samples, one extraction control and one PCR negative control). PCR negatives were included for each PCR run. Amplification of aquatic arthropod DNA, extracted from the bulk samples was performed using the primer pair BF2 (5' – GCHCCHGAYATRGCHTTYCC – 3') and BR2 (5' – TCDGGRTGNCCRAARAAYCA – 3') amplifying a 421 bp region of the COI gene (Elbrecht and Leese, 2017). The Tele03 primer were used to amplify the water eDNA samples, Tele03\_F (5' – GTCGGTAAAACTCGTGCCAGC – 3') and Tele03\_R (5' – CATAGTGGGGTATCTAATCCCAGTTTG – 3'), targeting the 12S mitochondrial rDNA gene, amplifying a 163-185 bp gene region of teleost fish DNA (Taberlet *et al.*, 2018). Our aim was to amplify arthropod DNA, from bulk samples, with the BF2/BR2 primer and fish DNA, from the water samples, with the Tele03 primer. However we also performed a PCR test of the



BF2/BR2 primer on the water samples, to examine whether we could detect arthropods in the water samples as well.

During Pre-PCR amplification, a Nanodrop Microvolume Spectrophotometer (Thermo Fisher Scientific) was used to measure nucleic acid concentration and purity ratios in a subset of the DNA extracts from the bulk aquatic invertebrates. 2µl of extracted DNA was added to let the software calculate the purity ratios, which can give indications of contaminants and DNA quantity. A significant difference in 260/280 ratios can indicate contaminants such as proteins and phenol and if there is a lower 260/230 purity ratio detected compared to what is expected, this can indicate if the used isolation technique requires further optimization. From the results, we detected a high difference in DNA concentration between samples and therefore decided to standardize them with a 1:5 dilution, which made it easier for the PCR polymerase to synthesize the denatured DNA strand. To measure the concentration of DNA extracted from the eDNA water samples, we used a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and based on the obtained results, we continued with amplification without further dilution.

The amplification strategy used was a single-step PCR with tagged primers. Primers included indexes so that each tag combination was specific to a sample (Fadrosh *et al.*, 2014). For the BF2/BR2 primer, we had 24 unique indexes and 16 reverse. The Tele03 had eight forward and two reverse indexes. The combination in the primer panels gave each sample a specific tag combination making it possible to trace the sequence reads back to its original sample in later bioinformatic analysis, to account for potential tag-jumps and avoid samples with falsely assigned sequences. The PCR reagents consisted of 2 X Q5, a high-fidelity polymerase, master mix (New England Biolabs), 5 X Q5 enhancer (New England Biolabs), 10 µM of indexed forward and reverse primer, 1.5 µl template DNA and nuclease-free water in a total volume of 25 µl. The thermal profile followed the PCR parameters seen in Appendix B (Lab protocols). Visualization of amplified PCR products were done on 1.5% agarose gels using GelRed loading dye with a FastRuler Low Range DNA ladder (Thermo Fisher Scientific), indicating good amplification of targeted lengths and negative controls appeared to be negative.

### **3.3.3 Library preparation and Sequencing**

Normalizing the amplified samples before sequencing was done to avoid dominant sequences in the data, secure equal concentrations of DNA for all samples, and to be certain that samples with less DNA concentration are not dominated by samples with higher concentrations

(Novoradovskaya *et al.*, 2006). The normalization is based on the relative quantity of the samples, calculated from the chosen standard which depends on the expected amplicon length. In this case, the standard was 400 bp for samples amplified with BF2/BR2 and 200 bp for samples amplified with the Tele03 marker. To calculate the DNA concentration in each sample, we analysed the agarose gels with the Image Lab Software 6.0.1 (Bio-Rad Laboratory), using the DNA ladder, which then was used to determine the desired volume for each sample. Markers with about the same amplicon lengths were pooled together using an automated liquid handler, the Biomek 4000 robot (Beckman Coulter). Measurements of DNA concentration in the normalized pool were done with a Qubit 2.0 fluorometer with the dsDNA High Sensitivity (HS) Assay Kit (see protocol in appendix B). Based on the results from the Qubit (Thermo Fisher Scientific), we diluted the assay tubes containing samples amplified with BF2/BR2 to 1:2. Before cleaning and size selection we used a Fragment Analyzer (Agilent) to perform quality control of our pre-NGS libraries. Pooled samples were mixed based on the length of the amplicons into two libraries. Clean-up of amplified libraries was done by following the AMPure XP protocol (Beckman Coulter) and to require amplicons of the wanted size for NGS libraries we used Blue Pippin (Sage. Science). One last quality check, with the Fragment Analyzer (Agilent), was done before sending them for sequencing by the Norwegian Sequencing Centre in Oslo (<https://www.sequencing.uio.no/>).

Libraries were sequenced on an Illumina MiSeq platform v3 (Illumina Inc). Rapid and accurate sequencing is enabled by Illumina next-generation sequencing (NGS) technologies. They use sequencing by synthesis (SBS) chemistry and clonal amplification to identify and incorporate fluorescently labelled deoxyribonucleotide triphosphate (dNTPs) into the DNA template strand. Each nucleotide has a unique fluorescent signal and when being incorporated, during each sequential cycle, fluorophore emission is used to identify nucleotides and determine the DNA sequence order (Bokulich *et al.*, 2013; Taberlet *et al.*, 2018).

## 3.4 Data analysis

### 3.4.1 Bioinformatics

A bioinformatic pipeline, for metabarcoding with high-throughput sequence data, typically consists of merging paired-end reads, sequence demultiplexing, quality filtering, denoising, and taxonomic assignment (Taberlet *et al.*, 2018).

Here we processed the sequence data for each primer set separately. Firstly, quality control of sequence data was assessed by FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). This enabled awareness of any problems in the data before further analysis. Merging of the raw forward and reverse sequence reads was done using PEAR 0.9.11 (Zhang *et al.*, 2014), with a minimum overlap of 20 bp and a specified 0.001 p-value for implemented statistical testing. Demultiplexing is the process of assigning sequences to the corresponding sample, according to the barcode combination of the amplified fragment. This was done by using the *ngsfilter* command in the OBITools Software (Boyer *et al.*, 2016). Sequences longer than 529 bp in sequence data for COI primer and 290 bp for Tele03 were discarded by the *obigrep* command. *Obiannotate* enabled specific attributes, attached to sequence records, to be removed before further quality filtering.

We used the USEARCH software (Edgar, 2016) to perform sequence quality filtering, dereplication and denoising. During quality filtering we discard reads > 1 of the total expected errors for all bases in the read, remove sequences < 80 bp for 12S data and sequences < 100 for COI data, and filter out all sequences that had one or multiple "N" in the reads. Sequences were dereplicated, grouping all sequences with a 100% identity together, and sequences occurring less than five times in the dataset were removed. Denoising was performed using the unoise3 algorithm (Edgar, 2016), which removes PCR errors and chimeras and returns denoised zero-radius OTUs (zOTUs), also known as Amplicon Sequence Variants (ASVs). The output was then used to generate a table that shows the number of reads per sample per ASV, which was used for downstream statistical analysis.

Taxonomic assignment of sequences was then finally performed using the *blastn* command from the basic local alignment search tool known as BLAST, version BLAST+ 2.7.1 (Altschul *et al.*, 1990; Camacho and Madden, 2022). For this project, we used the reference data from the

complete NCBI nucleotide collection (<https://www.ncbi.nlm.nih.gov/>). The taxonomic assignment was performed in the autumn of 2022.

### 3.4.2 Statistical analysis

Statistical analyses and visualization were done using RStudio (RStudio Team, 2020). To calculate and visualize the number of reads for each sample, and simultaneously identify problematic samples (i.e. sequences with low frequency and PCR errors), we used Phyloseq 1.24.2 (McMurdie and Holmes, 2013) to plot their distribution. Based on these plots' samples with less than 10 total reads and ASVs that were not present in any of the samples were removed. To only obtain taxa of interest, subsets of the datasets were made, for the COI dataset we made a subset of Arthropoda and a subset to the class Actinopterygii for the 12S dataset. Other packages used in the statistical analyses were ggplot2 3.4.0 (Wickham, 2009) and vegan 2.6.4 (Oksanen *et al.*, 2016). To assess the differences in relative abundance of taxa based on sample points and to check for significance of canopy cover we used the *adonis2* function to perform PERMANOVA (permutational multivariate analysis of variance) (Anderson and Walsh, 2013). Assumptions of equal dispersion within points were tested using the *betadisper* function. The *ordinate* function in Phyloseq (McMurdie and Holmes, 2013) was used to calculate Bray-Curtis and Jaccard dissimilarity matrices and used to generate nonmetric multidimensional scaling ordinations (NMDS). The datasets were analysed separately during the statistical analysis.

### 3.4.3 Stream community analysis

For both datasets, we investigated the observed and estimated ASV richness for each sampling point, using the *estimate\_richness* and visualised with *plot\_richness* functions in Phyloseq 1.24.2 (McMurdie and Holmes, 2013), thus performing a standard alpha diversity (Shannon's diversity index) estimation (Jost, 2006). Histograms for observed and estimated richness were made to examine if the data was normally distributed and a Kruskal Wallis non-parametric test was used to determine whether the medians of the groups were different.

To optimise the datasets for our objectives, pre-processing included removal of extraction control and PCR control samples and ASVs that were not present in any of the samples, keeping ASVs with frequency > 5% and removal of low abundance ASVs < 10% of the dataset median.

Following initial data exploration, we discovered a high read abundance of taxa belonging to the order Decapoda. To avoid Decapods to dominate the dataset, we chose to exclude them for further analysis. Number of reads in each sample were standardised to median sequencing depth and then transformed to relative abundance. Since we were not interested in intraspecific variation and focused our analysis at order or family level, we merged ASVs that were taxonomically assigned to the same taxonomic rank by using *tax\_glom* (McMurdie and Holmes, 2013) when exploring the relative abundance of taxonomic composition in the stream, visualised with the *ggplot* function in *ggplot2* (Wickham, 2009).

To study whether the arthropod community composition differed between points, we performed a multivariate analysis based on Bray-Curtis distances and non-parametric multidimensional scaling (NMDS). For detection of any statistical significance in observed differences in the stream community structure we visualised the NMDS ordination and subsequently performed a permutation multivariate analysis of variance, PERMANOVA (Anderson and Walsh, 2013; Bessey *et al.*, 2020; Muri *et al.*, 2020). A heatmap, of the ordination results, was produced for order read abundance in the arthropod samples. Ordination was also performed on the fish dataset, but the fish data had a limited density, number of taxa and samples and could therefore be considered not suitable for that kind of analysis.

To view the difference in community composition between forest exterior and interior we calculated and visualised the data by comparing samples in the corresponding areas and adjusted for the difference in number of sampling points. We performed PERMANOVA (Anderson and Walsh, 2013) to assess the statistical significance difference with Jaccard dissimilarity distance matrices, calculated using the *phyloseq::distance* function in *Phyloseq* 1.24.2 (McMurdie and Holmes, 2013). Assumption of equal dispersion within groups was evaluated using the *betadisper* function in *vegan* 2.6.4 (Oksanen *et al.*, 2016).



## 4 Results

From the initial 956 368 demultiplexed sequence reads of the bulk Arthropoda samples, following filtration and various quality control steps, we obtained 149 384 reads belonging to Arthropods in the 197 samples analysed, resulting in 102 ASVs. Taxa representing 10 distinct orders were detected. Those were Odonata, Decapoda, Hemiptera, Ephemeroptera, Diptera, Coleoptera, Plecoptera, Blattodea, Hymenoptera, and Orthoptera. There was a high abundance of Decapoda in the data ( $\approx 90\%$ ) so this was excluded from further analysis in order to focus on insect diversity patterns. The most prominent orders in the remaining data were Ephemeroptera  $\approx 41\%$ , Hemiptera  $\approx 38\%$ , and Odonata  $\approx 10\%$ .

From the initial 38 875 demultiplexed sequence reads belonging to the water eDNA samples, we obtained 26 453 reads belonging to fishes across the 15 samples analysed, resulting in 62 ASVs. Uniquely distinct orders detected for freshwater fish were Gymnotiformes, Characiformes, Cichliformes, Siluriformes, Synbranchiformes, and Beloniformes, where the number of families detected was 26 taxa. The three most common taxa were Characiformes  $\approx 45\%$ , Gymnotiformes  $\approx 26\%$ , and Cichliformes  $\approx 11\%$ . Only a few reads were detected for the three samples from point 4, this was most likely caused by a period of time when these samples were not stored in cold conditions before extraction. Therefore we chose to exclude the water samples from sampling point 4 in the following analysis.

We also detected a number of bacterial taxa in both datasets, even after filtration. This could indicate contaminants in our samples derived from either extraction, laboratory facilities, or PCR amplification. Although, the use of COI and 12S primers in metabarcoding is known to yield detectable bacterial signals, as these primers are universal and have a tendency to amplify a broad range of genetic material present in the sample (Collins *et al.*, 2019).

### 4.1 Alpha diversity

The Arthropoda and fish communities presented different patterns of alpha-diversity. For Arthropoda ASVs richness (Figure 4.1.1) the diversity gradually decreased downstream, from point 5 down to point 1, although the point with the highest estimated richness was point 3 (forest centre). The arthropods had an observed richness of 850 and a total Shannon of 5.62371.

ASV richness estimation differed less between points for the fish dataset (Figure 4.1.2), however, results indicate that the whole dataset had a high estimation of richness where the observed value was 278 and a total Shannon estimation of 4.090166. From histograms of observed and estimated richness, we found no evidence that either of the datasets was normally distributed, and results from the Kruskal-Wallis test strongly indicated ( $p < .001$ ) that the sampling points had a significant statistical difference between them in the arthropod data but no statistical difference was found for fish.

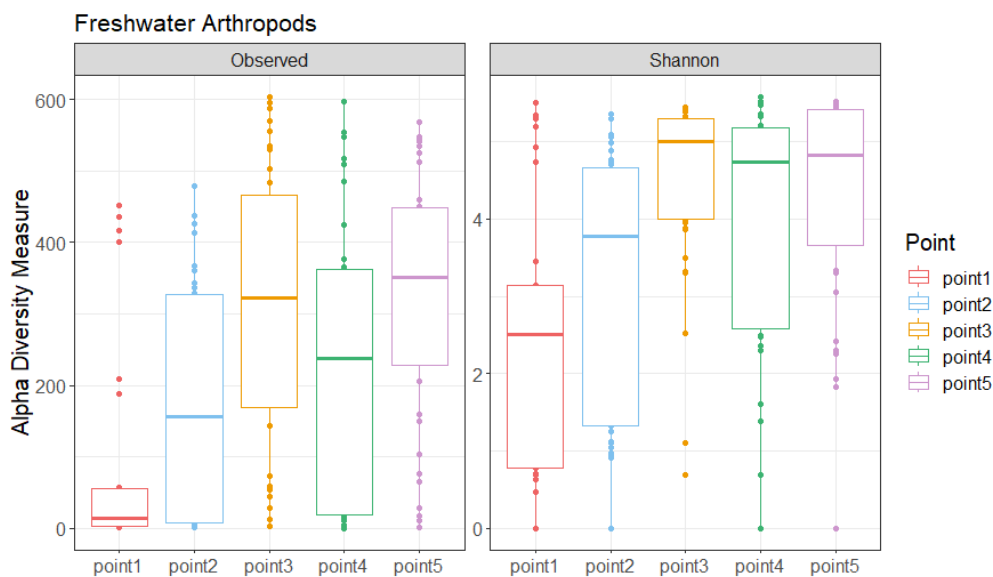


Figure 4.1. 1. Observed and estimated richness for Arthropoda in this study. High estimations for diversity and statistically significant difference between points (1-5). Point 3 had the highest estimated richness.

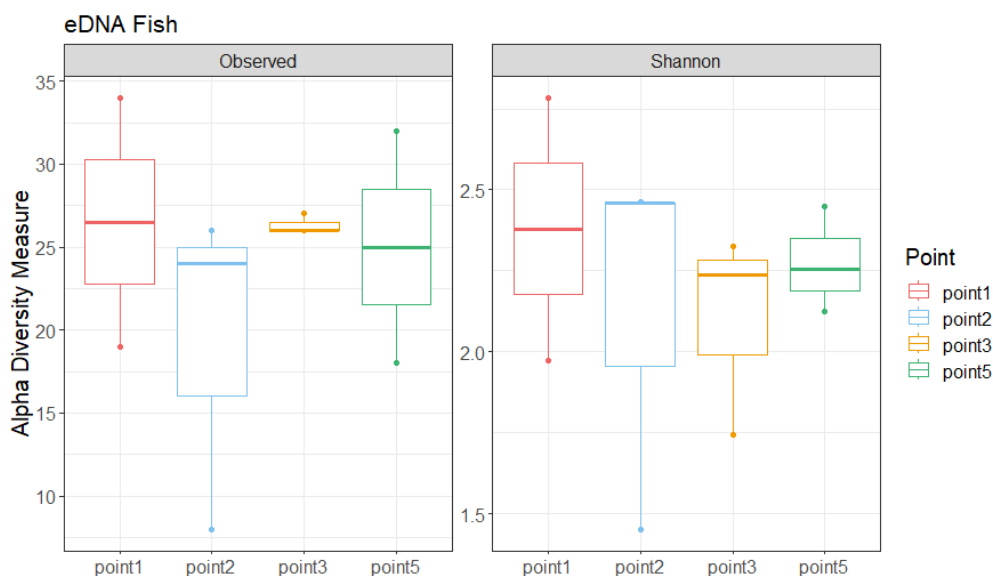


Figure 4.1. 2. Observed and estimated richness for fish in the different points. Where we find high estimations of ASV richness but less difference between sampling points (1-5). Low estimates for point 4 were expected and excluded from the analysis.



## 4.2 Relative abundance

Observed relative ASV abundance for Arthropod samples (Figure 4.2.1) showed that Ephemeroptera was most abundant through a larger portion of the samples and all sampling points. High abundance of Ephemeroptera was also shown in the heatmap which illustrated order abundance across all samples (Figure 4.2.2). Hemiptera abundance was mostly concentrated in the extremities of sampling points (point 1 and 5). The relative arthropod abundance indicated that there was a decrease in diversity downstream from Point 3 as fewer taxa were detected in those points, most common orders found in points 1 and 2 were Ephemeroptera, Odonata, and Hemiptera.

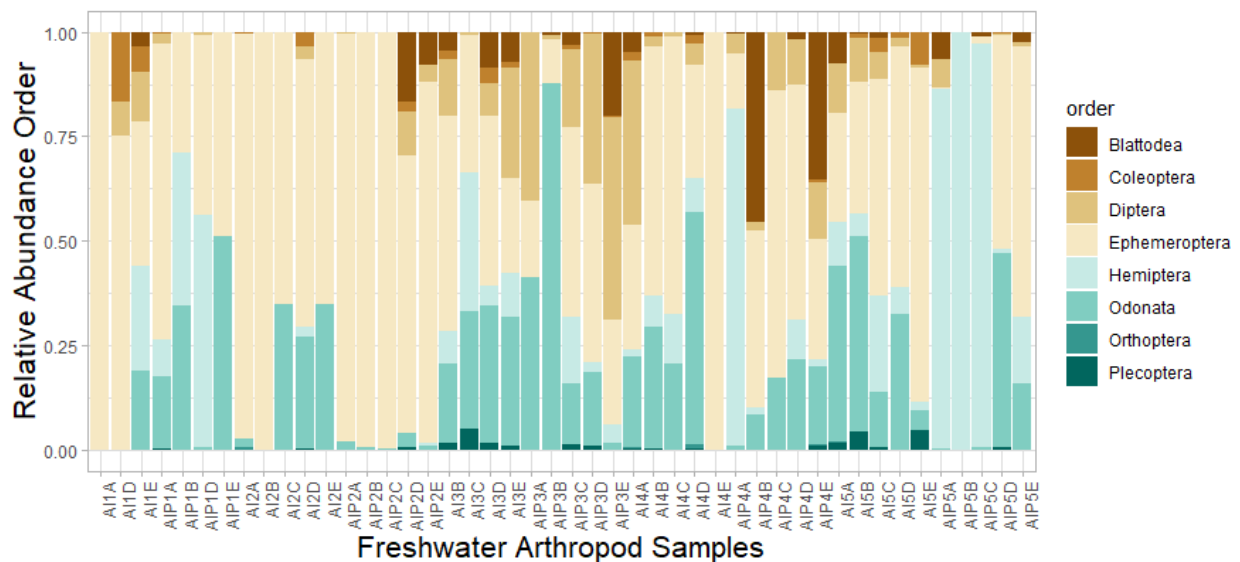


Figure 4.2. 1. Freshwater arthropod relative abundance for all bulk samples. Point 1 = A11A-AIP1E, Point 2 = A12-AIP2D, Point 3 = A13B-AIP3E, Point 4 = A14A-AIP4E, Point 5 = A15B-AIP5E.

We observed a higher abundance of orders in point 3, which reflects the richness estimations from figure 4.1.1. Above that, confidence in the indications of higher richness and abundance in sampling points 3 and 5 was increased by the heatmap (Figure 4.2.2). A relatively low abundance for Plecoptera was found, however they occurred mostly around points 3 and 5.



Figure 4.2. 2. Heatmap for abundance of order across all Arthropod samples. Orders are grouped from top: Plecoptera, Odonata, Hymenoptera, Hemiptera, Ephemeroptera, Diptera, Coleoptera, Blattodea. High abundance of Ephemeroptera in all bulk samples.

Results from PERMANOVA and NMDS ordination indicated that the arthropod community composition was different between sampling points ( $F = 15.92, p < .001$ ) (Figure 4.2.3). However upon further investigation, the test for homogeneity between points was statistically significant ( $Pr(>F) < .001$ ), therefore the PERMANOVA might not be the best fitted model to explain the variance of the data. Ordination of fish community composition was redundant as the dataset was too small in size, furthermore results from PERMANOVA indicated the same.

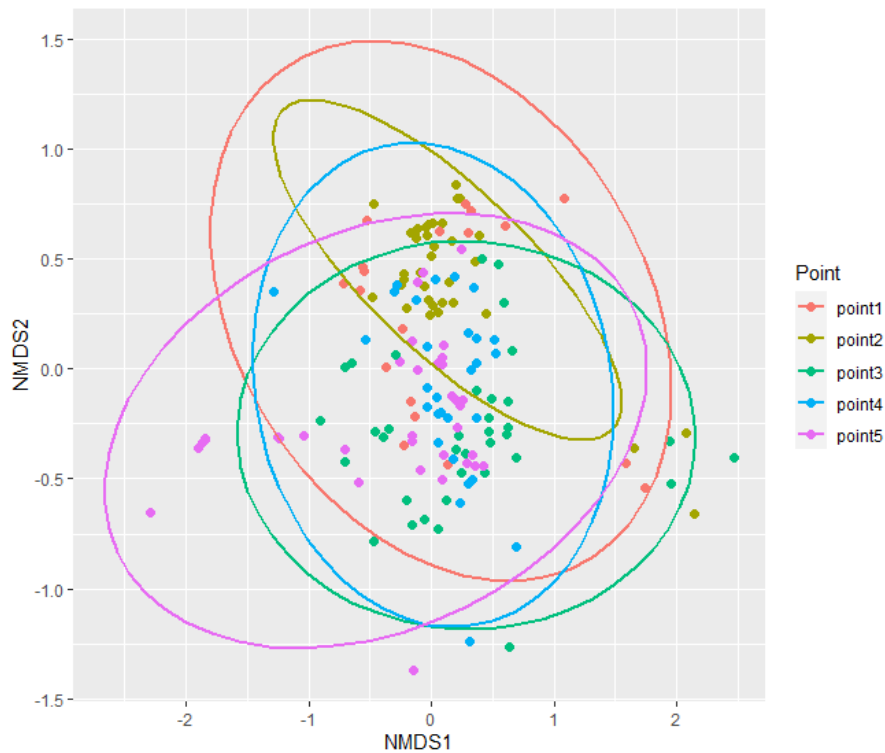


Figure 4.2. 3. Non-parametric multidimensional scaling (NMDS) ordination of Arthropod community in the study site, including 95% confidence ellipses around point centroids. Ordination based on Bray-Curtis dissimilarity.

When exploring relative abundance for fish families (Figure 4.2.4) we found that taxa belonging to Anostomidae dominated samples at all points. Concentrated taxa from Iguanodectidae were found in closer proximity to the exterior of the forest fragment. Relative abundance for fish community did not seem to differ from the different sampling points.

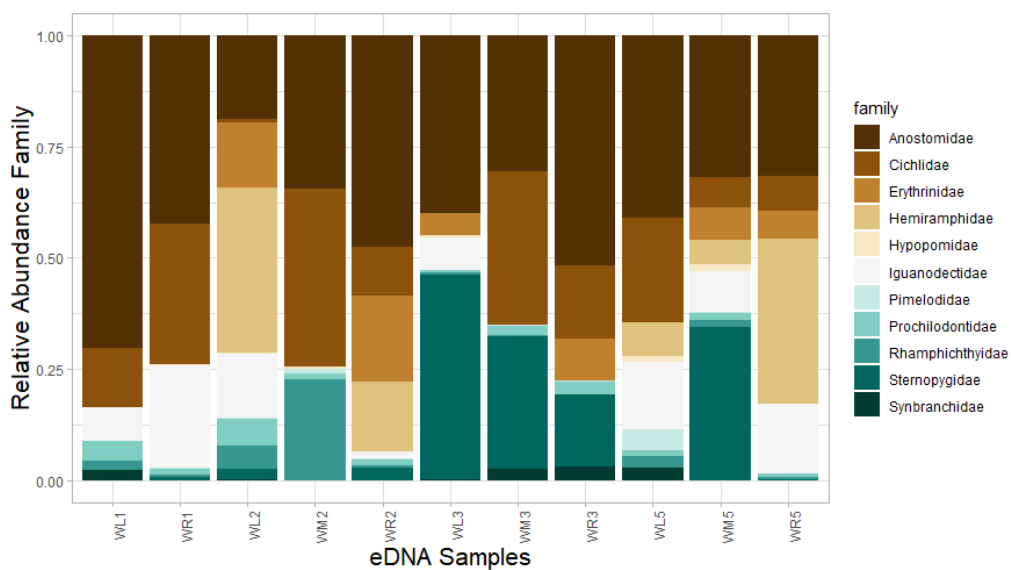


Figure 4.2. 4. Relative abundance for eDNA samples, families of fish taxa for all water samples.

### 4.3 Canopy cover

Our results show that canopy cover influenced abundance for the Arthropods, however no statistical significance supported this for the fish communities (Figure 4.3.1). We compared abundance of sampling points in the deforested and forested areas and PERMANOVA results for arthropods indicated a significant difference ( $F = 5.26$  and  $p < .001$ ), difference in variance was supported by the betadisper test ( $p = 0.1439$ ). The Kruskal-Wallis test determined whether the medians for the two groups (Forested vs Deforested) were different and confirmed the significant difference. The PERMANOVA on the fish community data did not indicate distinct differences between forest interior and exterior ( $F = 0.67$  and  $p = 0.82$ ), although a higher abundance was detected in the forested area.

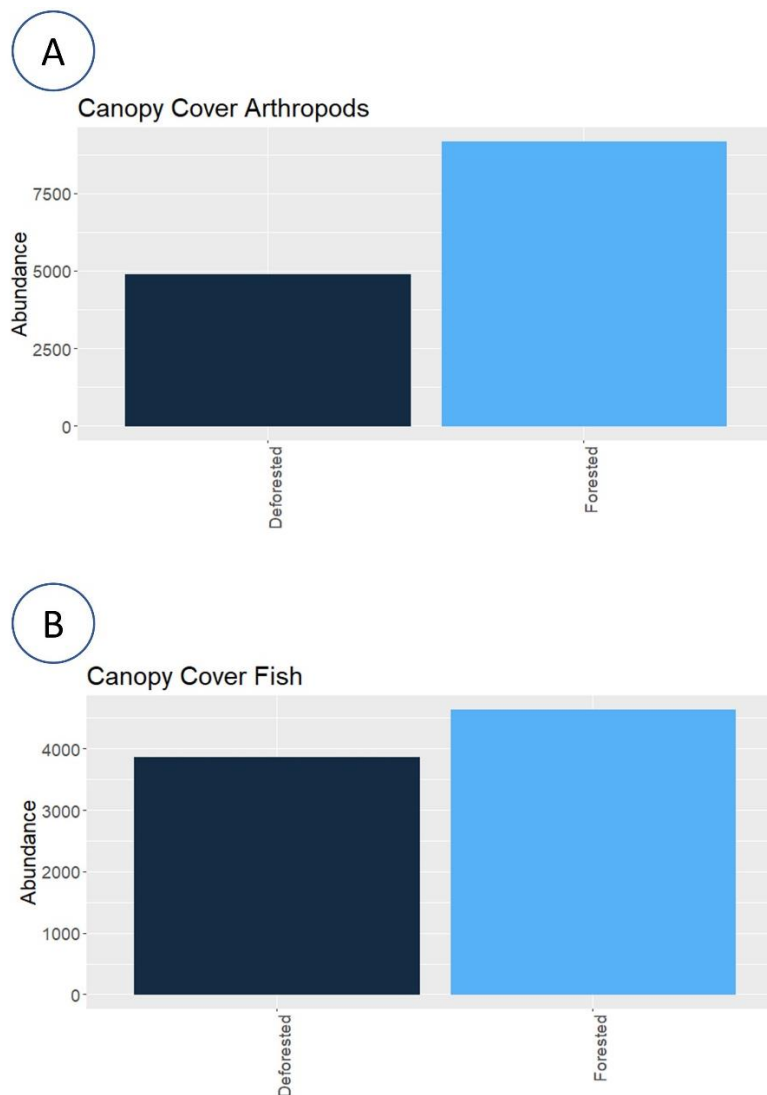


Figure 4.3. 1. Effect of canopy cover. A: Difference in Arthropod abundance between deforested (exterior) and forested (interior) areas. B: Difference in fish abundance between deforested and forested areas.



## 5 Discussion

Here, we aimed to assess the feasibility of DNA-based metabarcoding analysis for arthropod and fish detection, and to assess the effects of canopy cover on their abundance in a headwater stream located in a fragment of riparian forest. This will allow to evaluate if this method can be used for macroinvertebrate and fish monitoring to increase conservation efforts in a degrading landscape where the rate of biodiversity loss, and changes in community composition is potentially higher than the conventional monitoring methods are able to detect (Ji *et al.*, 2013; Valentini *et al.*, 2016; Elbrecht *et al.*, 2017; Taberlet *et al.*, 2018; Watts *et al.*, 2019). This exploratory study aimed to investigate the potential of eDNA metabarcoding in a relatively well known tropical environment as a first step before a future comparison with traditional monitoring methods.

### 5.1 DNA metabarcoding method

The results showed that metabarcoding-based detection, using either freshwater eDNA or bulk arthropod samples, was successful in detecting a substantial number of taxa for both macroinvertebrates and fish. In the bulk arthropods dataset, we identified 10 unique taxa at order level. The fish data contained 26 families in six distinct orders. If these species records were to be obtained by traditional monitoring methods, it would require extensive hours for reliable taxonomic identification by trained experts. Indeed, traditional monitoring techniques mainly relies on the identification of specimens to the species level, and this has proven to be difficult when studying areas in biodiversity hotspots where there is a lack of species records (Ji *et al.*, 2013; Salmah *et al.*, 2013). Although molecular based detection is so far not able to identify all specimens at the species level (due to incomplete reference databases, or primer specificity), studies showed that metabarcoding results can still be taxonomically comprehensive compared to standard biodiversity data (Ji *et al.*, 2013, 2022; Olds *et al.*, 2016; Valentini *et al.*, 2016; Bessey *et al.*, 2020), and that eDNA can be used for species identification and monitoring in tropical systems (Lynggaard *et al.*, 2019; Milan *et al.*, 2020; Baudry *et al.*, 2021; Wang *et al.*, 2021). In line with other metabarcoding studies, when not working on mock community data or comparing to traditional methods, we only considered indices of taxonomic levels for family or order for fish and arthropods respectively (Fernández *et al.*, 2018).

With 12 eDNA freshwater samples (the three samples from sampling point 4 were removed), we were able to detect a substantial number of taxa for fish at family level. It has been recognized that fish releases a high amount of eDNA into the environment, especially in warm waters (Jerde *et al.*, 2011; Thomsen and Willerslev, 2015; Civade *et al.*, 2016; Lacoursière-Roussel *et al.*, 2016). However, many factors can influence the detection probabilities, abiotic factors such as UV-radiation, amount of particles, dissolved oxygen, PH, and water temperature has an effect on the DNA persistence (He *et al.*, 2015; Tsuji, Yamanaka and Minamoto, 2017; Wang *et al.*, 2021). These factors also affect the decay rate of the different states (e.g. intracellular, intraorganellar, dissolved) of the eDNA in the environment (Mauvisseau *et al.*, 2022). For further studies, it would be of interest to record these factors in the area to investigate if they lead to any limitation for the use of molecular monitoring methods in headwaters in fragmented riparian forests. It could also be interesting to evaluate different filtration methods, as there is still no consensus on the amount of water that should be filtered (Taberlet *et al.*, 2018; Coutant *et al.*, 2021). A vampire sampler (Burlke, Germany) could be used to sample from deeper water depths and filter a higher amount of water with less effort, so we could compare our samples collected from the water surface and the amount of water that were filtrated.

We attempted to detect macroinvertebrate diversity in the eDNA water samples using the same COI marker as for the bulk arthropod samples. However, upon exploration of the sequencing results, we were able to obtain a relatively low number of reads from most water samples and several of these samples had too low sequence recovery to be used for further downstream analysis. Some studies show that arthropod detection from eDNA samples can be challenging, due to their external skeleton (Tréguier *et al.*, 2014). However, other studies had greater success in detecting macroinvertebrates in freshwater samples (Thomsen *et al.*, 2012; Fernández *et al.*, 2018; Leese *et al.*, 2021). It is also possible that the choice of COI primers used in this study or inhibition led to these unexpected limitations, and further work will be needed to fully disentangle these complex interactions.

For deeper taxonomic identification to be performed on our datasets, we should consider a comparison between our DNA sequences with local reference databases with more relevant queries (i.e. reference data with only Arthropods found in Brazil). However, this also highlights the need to increase the barcoding efforts in rich or poorly known systems.

## 5.2 Ecological patterns

Our results show significant difference in both estimated richness and relative abundance between sampling points of the study area for the arthropod community and that canopy cover had an influence on the arthropod abundance. However, we could not find evidence of a significant difference in the fish distribution, despite some patterns that could be observed.

### 5.2.1 Richness and relative abundance

Our study found that upstream sampling points, from point 3 and up to point 5, presented a higher estimated richness of arthropods (ASVs), with the highest estimated richness in the center, point 3, of the forest fragment. Higher richness in points 2 - 4 was previously expected, where the canopy cover was denser, but the results show otherwise. It was confirmed that observed and estimated richness for arthropods significantly differed between all sampling points. However, we found no difference between sampling points for the estimated richness of fish (ASVs), although the overall alpha diversity was high. Richness can be sensitive to primer choice and some studies recommend the use of multiple primer sets to recover maximum target taxa richness (Hajibabaei *et al.*, 2019). Here, we only used one primer set for each sample type, and the use of multiple primers could be considered for future biodiversity analysis. Recent work suggests that biodiversity studies should shift their focus from species richness to changes in the community composition instead (Aggemyr *et al.*, 2018; Lynggaard *et al.*, 2019; Wang *et al.*, 2019). This is because changes in alpha diversity caused by the substitution of taxa may not always be detected, and therefore the community composition can be a better indicator of biodiversity changes (Lynggaard *et al.*, 2019). An example of this was discussed in a study where a decrease of Ephemeroptera, Plecoptera, and Trichoptera (EPT) richness was not reflected in the total richness, which was most likely caused by the increased richness of Chironomidae and other groups (Beermann *et al.*, 2021). EPT are often referred to as bioindicators, as they are sensitive to environmental perturbations, and are of special interest for biomonitoring and assessments of stream ecosystem water quality (Bispo and Oliveira, 2007; Zizka *et al.*, 2019; Sundar *et al.*, 2020; Beermann *et al.*, 2021; Leese *et al.*, 2021). EPT found in this study will be discussed later.

However, the use of richness estimates should depend on the objectives of the study. Here, we were mostly interested in differences in estimated richness between sampling points and the effect of canopy cover on all taxa.



The relative sequence abundance of arthropods was higher in sampling points 3-5, where we found all uniquely detected orders (Figure 4.2.1) within these points. No orders were found exclusively in a unique sampling point, and when examining EPT abundance, we were not successful in detecting any Trichoptera. However, the relative abundance of Ephemeroptera was high in almost all samples and Plecoptera was detected in the centre of the fragment and sampling point 5. Substrate embeddedness and canopy cover can influence EPT abundance (Ab Hamid and Salmah, 2011). In a study conducted in Peninsular Malaysia, Ab Hamid and Salmah (2011) found that Ephemeroptera was the most dominant EPT taxa in all the forest streams in the Gunung Jerai Forest Reserve and that their habitats both shaded and unshaded waters, whereas Plecoptera and Trichoptera preferred shaded habitats. Between the sampling points with lower canopy cover (Points 1 and 5), we observed a higher relative abundance for Arthropoda orders in sampling point 5. Sampling point 1 had less canopy cover than the two, was in closer proximity to human constructions and had a finer sediment substrate. Substrate size and fine sediment are important for aquatic insects that require shelter, incubation of eggs, deposition, different food sources, and refuge from predation (Hamid and Md Rawi, 2014; Piggott, Townsend and Matthaei, 2015; Stanford *et al.*, 2020). This could explain the differences in arthropod composition between the deforested areas.

Finally, we found no distinct differences in relative abundance for fish detected in the stream. This can reflect the fact that fish are larger organisms and were able to travel between the five sampling locations. In addition, the fish data were collected by environmental DNA and the distance which eDNA can travel is suggested to be several kilometers (Pont *et al.*, 2018; Villacorta-Rath *et al.*, 2021; Jo and Yamanaka, 2022; Van Driessche *et al.*, 2022). This could mean that the fishes detected in sampling point 1 could originate from eDNA found in point 5, or higher upstream, and could explain the homogeneity found in our fish dataset.

### **5.2.2 Differences between forest exterior and interior**

As expected, we found that canopy cover had an effect on the total abundance of both arthropod and fish communities, and both communities differed between forested and deforested sampling points. Riparian forest with higher integrity supports higher allochthonous insect input to the stream, favourable substrate conditions, temperature regulations, detritus, and light conditions (Décamps, Naiman and McClain, 2009; Sass, 2009; Turunen *et al.*, 2019). The benefits of forested riparian zones should be recognized in freshwater and biodiversity

management in degraded stream habitats, as they support higher biodiversity and can mitigate warming and the ecological degradation of headwater streams (Bowler *et al.*, 2012; Feld *et al.*, 2018; Turunen *et al.*, 2019). The Jaccard dissimilarity strongly indicated that the observed difference in arthropod composition between forested and deforested areas was statistically significant. However, we did not find evidence for this when analysing the fish data. It should be noted that the samples were collected in the dry season and the higher total abundance of fish in the forested areas could be a result of fish seeking shade and cooler water.

Forest fragmentation and habitat reduction of riparian zones lead to diversity loss and reduction of important stream ecosystem processes such as mitigation of input pollutants, organic matter, and macroinvertebrate diversity (Sweeney *et al.*, 2004; Décamps, Naiman and McClain, 2009; Bowler *et al.*, 2012; Fremier *et al.*, 2015; Mello *et al.*, 2018; Firmiano *et al.*, 2021). There has been a significant amount of research on the effect of fragmentation on aquatic biodiversity (Zwick, 1992; Scheffer *et al.*, 2006; Salmah *et al.*, 2013; Fuller, Doyle and Strayer, 2015; Renó, Novo and Escada, 2016). But given the fast and efficient potential of metabarcoding, we see a rise in research efforts on implementing DNA metabarcoding to survey biodiversity in natural communities (Deiner *et al.*, 2017; Fernández *et al.*, 2018; Beermann *et al.*, 2021; Turunen *et al.*, 2021). Despite this, our study was the first to explore the effect of canopy cover on aquatic arthropods and fish in a riparian forest fragment with metabarcoding, in eastern Amazonia.

### 5.3 Perspectives

This study shows the potential use of metabarcoding, both eDNA and metabarcoding analysis of bulk samples, to detect changes in community composition of arthropods and fish in a fragmented riparian habitat. However, more research is needed to strengthen the confidence of the method in this type of study. For instance, we were performing the taxonomic assignment without much knowledge on the species found in the area. For future investigation, the metabarcoding method can be used in addition to traditional taxonomic assignment and expertise. Also, a more restricted reference database should be used, for example, the available data on arthropods found in Brazil from BOLD. However, incorrect taxon assignment can still occur due to incorrect database entries (Weigand *et al.*, 2019). The data obtained in this study could have been affected by different factor, the study area was close to a big cattle pasture, which could introduce contaminants to our samples in the form of allochthonous eDNA, we also detected considerable amounts of sequence reads for bacterial taxa even after various filtration steps. The detected arthropods in our data could be limited due to catchment methods, detection probabilities vary based on the species, habitat, and detection method (Deiner *et al.*, 2017). For instance, our nets/sieves sizes may have been too big, where juveniles or smaller insects could slip through the holes, although we tried to correct for this by including the two methods rather than only one of them. Primer choice can also affect the data outcome and therefore potentially lead to undetected taxa in the samples (Miya *et al.*, 2015; Elbrecht and Leese, 2016; Hajibabaei *et al.*, 2019). Our study used two validated universal primer sets to detect the arthropods and fish in our samples (Elbrecht and Leese, 2016; Taberlet *et al.*, 2018), however, new primers sets for optimal amplification of aquatic invertebrates have since been produced (Leese *et al.*, 2021) and there are other optional universal PCR primers for fish (Miya *et al.*, 2015). An option for future amplification is to curate a local sequence reference database based on the knowledge on the desired study organisms and produce *in-silico* primers (Taberlet *et al.*, 2018). It is important to be aware that biases can be introduced during sampling, extraction, amplification, sequencing, and bioinformatics. There is no perfect way to control this. However, there are tools that can help prevent this as much as possible (Goldberg *et al.*, 2016; Taberlet *et al.*, 2018).

## 6 Conclusion

In conclusion, this study was able to provide relative quantitative information about the aquatic diversity in the small forest fragment located in Paragominas and our results indicates that canopy cover has an effect on abundance of arthropods but not for the fish found there. Conducting remote field work has proven to be challenging, working under hot temperatures affected both the samples collected and the collectors. Considerations for later includes the use of better curated reference databases, collecting samples further away from the edges of the forest fragment, other sample sites for comparison (i.e. fragments of varied sizes and a completely intact riparian forest for reference), and internal extraction control. And, if possible, generate sequence data in the field by using Oxford Nanopore Technologies' MinION™ (Baloglu *et al.*, 2021).

# References

- Ab Hamid, S. and Salmah, M. (2011) 'Influence of substrate embeddedness and canopy cover on the distribution of Ephemeroptera, Plecoptera and Trichoptera (EPT) in tropical rivers', *Aquatic Insects*, 33, pp. 281–292. Available at: <https://doi.org/10.1080/01650424.2011.640940>.
- Aggemyr, E. *et al.* (2018) 'Species richness and composition differ in response to landscape and biogeography', *Landscape Ecology*, 33(12), pp. 2273–2284. Available at: <https://doi.org/10.1007/s10980-018-0742-9>.
- Altschul, S.F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. Available at: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Anderson, M.J. and Walsh, D.C.I. (2013) 'PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing?', *Ecological Monographs*, 83(4), pp. 557–574. Available at: <https://doi.org/10.1890/12-2010.1>.
- Andren, H. and Angelstam, P. (1988) 'Elevated Predation Rates as an Edge Effect in Habitat Islands: Experimental Evidence', *Ecology*, 69(2), pp. 544–547.
- Antonelli, A. *et al.* (2018) 'Amazonia is the primary source of Neotropical biodiversity', *Proceedings of the National Academy of Sciences*, 115(23), pp. 6034–6039. Available at: <https://doi.org/10.1073/pnas.1713819115>.
- Baloğlu, B. *et al.* (2021) 'A workflow for accurate metabarcoding using nanopore MinION sequencing', *Methods in Ecology and Evolution*, 12(5), pp. 794–804. Available at: <https://doi.org/10.1111/2041-210X.13561>.
- Basset, Y. *et al.* (2012) 'Arthropod Diversity in a Tropical Forest', *Science*, 338(6113), pp. 1481–1484. Available at: <https://doi.org/10.1126/science.1226727>.
- Baudry, T. *et al.* (2021) 'Mapping a super-invader in a biodiversity hotspot, an eDNA-based success story', *Ecological Indicators*, 126, p. 107637. Available at: <https://doi.org/10.1016/j.ecolind.2021.107637>.
- Beermann, A.J. *et al.* (2021) 'DNA metabarcoding improves the detection of multiple stressor responses of stream invertebrates to increased salinity, fine sediment deposition and reduced flow velocity', *Science of The Total Environment*, 750, p. 141969. Available at: <https://doi.org/10.1016/j.scitotenv.2020.141969>.
- Bessey, C. *et al.* (2020) 'Maximizing fish detection with eDNA metabarcoding', *Environmental DNA*, 2(4), pp. 493–504. Available at: <https://doi.org/10.1002/edn3.74>.
- Birk, S. *et al.* (2012) 'Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive', *Ecological Indicators*, 18, pp. 31–41. Available at: <https://doi.org/10.1016/j.ecolind.2011.10.009>.

- Bokulich, N.A. *et al.* (2013) ‘Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing’, *Nature Methods*, 10(1), pp. 57–59. Available at: <https://doi.org/10.1038/nmeth.2276>.
- Bowler, D.E. *et al.* (2012) ‘What are the effects of wooded riparian zones on stream temperature?’, *Environmental Evidence*, 1(1), p. 3. Available at: <https://doi.org/10.1186/2047-2382-1-3>.
- Boyer, F. *et al.* (2016) ‘obitools: a unix-inspired software package for DNA metabarcoding’, *Molecular Ecology Resources*, 16(1), pp. 176–182. Available at: <https://doi.org/10.1111/1755-0998.12428>.
- Braukmann, T.W.A. *et al.* (2019) ‘Metabarcoding a diverse arthropod mock community’, *Molecular Ecology Resources*, 19(3), pp. 711–727. Available at: <https://doi.org/10.1111/1755-0998.13008>.
- Bylemans, J. *et al.* (2019) ‘A performance evaluation of targeted eDNA and eDNA metabarcoding analyses for freshwater fishes’, *Environmental DNA*, 1(4), pp. 402–414. Available at: <https://doi.org/10.1002/edn3.41>.
- Camacho, C. and Madden, T. (2022) *BLAST+ Release Notes, BLAST® Help [Internet]*. National Center for Biotechnology Information (US). Available at: <https://www.ncbi.nlm.nih.gov/books/NBK131777/> (Accessed: 30 November 2022).
- Civade, R. *et al.* (2016) ‘Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System’, *PLOS ONE*, 11(6), p. e0157366. Available at: <https://doi.org/10.1371/journal.pone.0157366>.
- Collinge, S.K. (1996) ‘Ecological consequences of habitat fragmentation: implications for landscape architecture and planning’, *Landscape and Urban Planning*, 36(1), pp. 59–77. Available at: [https://doi.org/10.1016/S0169-2046\(96\)00341-6](https://doi.org/10.1016/S0169-2046(96)00341-6).
- Collins, R.A. *et al.* (2019) ‘Non-specific amplification compromises environmental DNA metabarcoding with COI’, *Methods in Ecology and Evolution*, 10(11), pp. 1985–2001. Available at: <https://doi.org/10.1111/2041-210X.13276>.
- Cooper, S., Walde, S. and Peckarsky, B. (1990) ‘Prey Exchange Rates and the Impact of Predators on Prey Populations in Streams’, *Ecology*, 71, pp. 1503–1514. Available at: <https://doi.org/10.2307/1938287>.
- Coutant, O. *et al.* (2021) ‘Detecting fish assemblages with environmental DNA: Does protocol matter? Testing eDNA metabarcoding method robustness’, *Environmental DNA*, 3(3), pp. 619–630. Available at: <https://doi.org/10.1002/edn3.158>.
- da Cruz, D.C. *et al.* (2022) ‘Priority areas for restoration in permanent preservation areas of rural properties in the Brazilian Amazon’, *Land Use Policy*, 115, p. 106030. Available at: <https://doi.org/10.1016/j.landusepol.2022.106030>.
- Daam, M.A. and Rico, A. (2018) ‘Freshwater shrimps as sensitive test species for the risk assessment of pesticides in the tropics’, *Environmental Science and Pollution Research*, 25(14), pp. 13235–13243. Available at: <https://doi.org/10.1007/s11356-016-7451-1>.

Décamps, H., Naiman, R.J. and McClain, M.E. (2009) ‘Riparian Zones’, in G.E. Likens (ed.) *Encyclopedia of Inland Waters*. Oxford: Academic Press, pp. 396–403. Available at: <https://doi.org/10.1016/B978-012370626-3.00053-3>.

Deiner, K. *et al.* (2017) ‘Environmental DNA metabarcoding: Transforming how we survey animal and plant communities’, *Molecular Ecology*, 26(21), pp. 5872–5895. Available at: <https://doi.org/10.1111/mec.14350>.

Díaz, S. *et al.* (2006) ‘Biodiversity Loss Threatens Human Well-Being’, *PLOS Biology*, 4(8), p. e277. Available at: <https://doi.org/10.1371/journal.pbio.0040277>.

*DNeasy PowerWater Sterivex Kit Quick Start Protocol - QIAGEN* (no date). Available at: <https://www.qiagen.com/us/resources/resourcedetail?id=49c37c7c-1872-4989-aaf6-030394aba0cb&lang=en> (Accessed: 11 November 2022).

Doyle, Jeff L. and Doyle, Jane L. (1987) *A rapid DNA isolation procedure from small quantities of fresh leaf tissues*. Available at: <https://www.scienceopen.com/document?vid=1cd6f3da-bc63-466b-9990-270af0960e2a> (Accessed: 11 December 2022).

Dudgeon, D. *et al.* (2006) ‘Freshwater biodiversity: importance, threats, status and conservation challenges’, *Biological Reviews*, 81(2), pp. 163–182. Available at: <https://doi.org/10.1017/S1464793105006950>.

Edgar, R.C. (2016) *UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing*. preprint. Bioinformatics. Available at: <https://doi.org/10.1101/081257>.

Elbrecht, V. *et al.* (2017) ‘Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring’, *Methods in Ecology and Evolution*, 8(10), pp. 1265–1275. Available at: <https://doi.org/10.1111/2041-210X.12789>.

Elbrecht, V. and Leese, F. (2016) ‘Development and validation of DNA metabarcoding COI primers for aquatic invertebrates using the R package “PrimerMiner”’. Available at: <https://doi.org/10.7287/PEERJ.PREPRINTS.2044V1>.

Elbrecht, V. and Leese, F. (2017) ‘Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment’, *Frontiers in Environmental Science*, 5. Available at: <https://doi.org/10.3389/fenvs.2017.00011>.

Fadrosh, D.W. *et al.* (2014) ‘An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform’, *Microbiome*, 2(1), p. 6. Available at: <https://doi.org/10.1186/2049-2618-2-6>.

Fahrig, L. (2003) ‘Effects of Habitat Fragmentation on Biodiversity’, *Annual Review of Ecology, Evolution, and Systematics*, 34, pp. 487–515.

Fearnside, P.M. (2005) ‘Deforestation in Brazilian Amazonia: History, Rates, and Consequences’, *Conservation Biology*, 19(3), pp. 680–688. Available at: <https://doi.org/10.1111/j.1523-1739.2005.00697.x>.

- Feld, C.K. *et al.* (2018) 'Evaluating riparian solutions to multiple stressor problems in river ecosystems — A conceptual study', *Water Research*, 139, pp. 381–394. Available at: <https://doi.org/10.1016/j.watres.2018.04.014>.
- Fernández, S. *et al.* (2018) 'Evaluating freshwater macroinvertebrates from eDNA metabarcoding: A river Nalón case study', *PLOS ONE*, 13(8), p. e0201741. Available at: <https://doi.org/10.1371/journal.pone.0201741>.
- Firmiano, K.R. *et al.* (2021) 'Functional responses of aquatic invertebrates to anthropogenic stressors in riparian zones of Neotropical savanna streams', *Science of The Total Environment*, 753, p. 141865. Available at: <https://doi.org/10.1016/j.scitotenv.2020.141865>.
- Forio, M.A.E. and Goethals, P.L.M. (2020) 'An Integrated Approach of Multi-Community Monitoring and Assessment of Aquatic Ecosystems to Support Sustainable Development', *Sustainability*, 12(14), p. 5603. Available at: <https://doi.org/10.3390/su12145603>.
- Fremier, A.K. *et al.* (2015) 'A riparian conservation network for ecological resilience', *Biological Conservation*, 191, pp. 29–37. Available at: <https://doi.org/10.1016/j.biocon.2015.06.029>.
- de la Fuente, B. *et al.* (2018) 'Natura 2000 sites, public forests and riparian corridors: The connectivity backbone of forest green infrastructure', *Land Use Policy*, 75, pp. 429–441. Available at: <https://doi.org/10.1016/j.landusepol.2018.04.002>.
- Fuller, M.R., Doyle, M.W. and Strayer, D.L. (2015) 'Causes and consequences of habitat fragmentation in river networks', *Annals of the New York Academy of Sciences*, 1355(1), pp. 31–51. Available at: <https://doi.org/10.1111/nyas.12853>.
- Gastauer, M. *et al.* (2019) 'Mine land rehabilitation in Brazil: Goals and techniques in the context of legal requirements', *Ambio*, 48(1), pp. 74–88. Available at: <https://doi.org/10.1007/s13280-018-1053-8>.
- Goldberg, C.S. *et al.* (2016) 'Critical considerations for the application of environmental DNA methods to detect aquatic species', *Methods in Ecology and Evolution*, 7(11), pp. 1299–1307. Available at: <https://doi.org/10.1111/2041-210X.12595>.
- Hajibabaei, M. *et al.* (2019) 'COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems', *PLOS ONE*, 14(9), p. e0220953. Available at: <https://doi.org/10.1371/journal.pone.0220953>.
- Hamid, S.A. and Md Rawi, C.S. (2014) 'Ecology of Ephemeroptera, Plecoptera and Trichoptera (Insecta) in Rivers of the Gunung Jerai Forest Reserve: Diversity and Distribution of Functional Feeding Groups', *Tropical Life Sciences Research*, 25(1), pp. 61–73.
- Harris, J.H. (1995) 'The use of fish in ecological assessments', *Australian Journal of Ecology*, 20(1), pp. 65–80. Available at: <https://doi.org/10.1111/j.1442-9993.1995.tb00523.x>.
- He, X. *et al.* (2015) 'Persistence of mitochondrial DNA markers as fecal indicators in water environments', *Science of The Total Environment*, 533, pp. 383–390. Available at: <https://doi.org/10.1016/j.scitotenv.2015.06.119>.



Jerde, C.L. *et al.* (2011) ‘“Sight-unseen” detection of rare aquatic species using environmental DNA’, *Conservation Letters*, 4(2), pp. 150–157. Available at: <https://doi.org/10.1111/j.1755-263X.2010.00158.x>.

Ji, F. *et al.* (2022) ‘Assessment of benthic invertebrate diversity and river ecological status along an urbanized gradient using environmental DNA metabarcoding and a traditional survey method’, *Science of The Total Environment*, 806, p. 150587. Available at: <https://doi.org/10.1016/j.scitotenv.2021.150587>.

Ji, Y. *et al.* (2013) ‘Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding’, *Ecology Letters*, 16(10), pp. 1245–1257. Available at: <https://doi.org/10.1111/ele.12162>.

Jo, T. and Yamanaka, H. (2022) ‘Meta-analyses of environmental DNA downstream transport and deposition in relation to hydrogeography in riverine environments’, *Freshwater Biology*, 67(8), pp. 1333–1343. Available at: <https://doi.org/10.1111/fwb.13920>.

Johnsen, S.I. *et al.* (2020) ‘Environmental DNA (eDNA) Monitoring of Noble Crayfish *Astacus astacus* in Lentic Environments Offers Reliable Presence-Absence Surveillance – But Fails to Predict Population Density’, *Frontiers in Environmental Science*, 8. Available at: <https://doi.org/10.3389/fenvs.2020.612253>.

Jost, L. (2006) ‘Entropy and diversity’, *Oikos*, 113(2), pp. 363–375. Available at: <https://doi.org/10.1111/j.2006.0030-1299.14714.x>.

Lacoursière-Roussel, A. *et al.* (2016) ‘Quantifying relative fish abundance with eDNA: a promising tool for fisheries management’, *Journal of Applied Ecology*, 53(4), pp. 1148–1157. Available at: <https://doi.org/10.1111/1365-2664.12598>.

Lacoursière-Roussel, A., Rosabal, M. and Bernatchez, L. (2016) ‘Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions’, *Molecular Ecology Resources*, 16(6), pp. 1401–1414. Available at: <https://doi.org/10.1111/1755-0998.12522>.

Lambertini, M. (2020) *Living Planet Report 2020: Bending the Curve of Biodiversity Loss*. Available at: <http://www.deslibris.ca/ID/10104983> (Accessed: 3 December 2020).

Laurent, F. *et al.* (2017) ‘Soil texture derived from topography in North-eastern Amazonia’, *Journal of Maps*, 13(2), pp. 109–115. Available at: <https://doi.org/10.1080/17445647.2016.1266524>.

Leese, F. *et al.* (2021) ‘Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification’, *Environmental DNA*, 3(1), pp. 261–276. Available at: <https://doi.org/10.1002/edn3.177>.

Liu, Y., Bailey, J.L. and Davidsen, J.G. (2019) ‘Social-Cultural Ecosystem Services of Sea Trout Recreational Fishing in Norway’, *Frontiers in Marine Science*, 6. Available at: <https://www.frontiersin.org/articles/10.3389/fmars.2019.00178> (Accessed: 12 December 2022).

- Lynch, A.J. *et al.* (2016) ‘The social, economic, and environmental importance of inland fish and fisheries’, *Environmental Reviews*, 24(2), pp. 115–121. Available at: <https://doi.org/10.1139/er-2015-0064>.
- Lynggaard, C. *et al.* (2019) ‘Vertebrate diversity revealed by metabarcoding of bulk arthropod samples from tropical forests’, *Environmental DNA*, 1(4), pp. 329–341. Available at: <https://doi.org/10.1002/edn3.34>.
- Mauvisseau, Q. *et al.* (2022) ‘The Multiple States of Environmental DNA and What Is Known about Their Persistence in Aquatic Environments’, *Environmental Science & Technology*, 56(9), pp. 5322–5333. Available at: <https://doi.org/10.1021/acs.est.1c07638>.
- McCluskey, K. *et al.* (2017) ‘The U.S. Culture Collection Network Responding to the Requirements of the Nagoya Protocol on Access and Benefit Sharing’, *mBio*, 8(4), pp. e00982-17. Available at: <https://doi.org/10.1128/mBio.00982-17>.
- McMurdie, P.J. and Holmes, S. (2013) ‘phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data’, *PLOS ONE*, 8(4), p. e61217. Available at: <https://doi.org/10.1371/journal.pone.0061217>.
- Mello, K. de *et al.* (2018) ‘Effects of land use and land cover on water quality of low-order streams in Southeastern Brazil: Watershed versus riparian zone’, *CATENA*, 167, pp. 130–138. Available at: <https://doi.org/10.1016/j.catena.2018.04.027>.
- Milan, D.T. *et al.* (2020) ‘New 12S metabarcoding primers for enhanced Neotropical freshwater fish biodiversity assessment’, *Scientific Reports*, 10(1), p. 17966. Available at: <https://doi.org/10.1038/s41598-020-74902-3>.
- Miya, M. *et al.* (2015) ‘MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species’, *Royal Society Open Science*, 2(7), p. 150088. Available at: <https://doi.org/10.1098/rsos.150088>.
- Mondy, C.P. *et al.* (2012) ‘A new macroinvertebrate-based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands: A taxonomical and trait approach’, *Ecological Indicators*, 18, pp. 452–467. Available at: <https://doi.org/10.1016/j.ecolind.2011.12.013>.
- Muri, C.D. *et al.* (2020) ‘Read counts from environmental DNA (eDNA) metabarcoding reflect fish abundance and biomass in drained ponds’, *Metabarcoding and Metagenomics*, 4, p. e56959. Available at: <https://doi.org/10.3897/mbmg.4.56959>.
- Nessimian, J.L. *et al.* (2008) ‘Land use, habitat integrity, and aquatic insect assemblages in Central Amazonian streams’, *Hydrobiologia*, 614(1), pp. 117–131. Available at: <https://doi.org/10.1007/s10750-008-9441-x>.
- Novoradovskaya, N. *et al.* (2006) ‘Normalization of samples for amplification reactions’. Available at: <https://patents.google.com/patent/US20060286558/en> (Accessed: 12 December 2022).

- Nunes, S.S. *et al.* (2015) ‘A 22 year assessment of deforestation and restoration in riparian forests in the eastern Brazilian Amazon’, *Environmental Conservation*, 42(3), pp. 193–203. Available at: <https://doi.org/10.1017/S0376892914000356>.
- Oksanen, J. *et al.* (2016) ‘vegan: Community Ecology Package. Ordination methods, diversity analysis and other functions for community and vegetation ecologists’, *CIFOR-ICRAF*, 30 June. Available at: [https://www.cifor-icraf.org/knowledge/publication/\\_\\_39499/](https://www.cifor-icraf.org/knowledge/publication/__39499/) (Accessed: 3 December 2022).
- Olds, B.P. *et al.* (2016) ‘Estimating species richness using environmental DNA’, *Ecology and Evolution*, 6(12), pp. 4214–4226. Available at: <https://doi.org/10.1002/ece3.2186>.
- Osis, R., Laurent, F. and Pocard-Chapuis, R. (2019) ‘Spatial determinants and future land use scenarios of Paragominas municipality, an old agricultural frontier in Amazonia’, *Journal of Land Use Science*, 14(3), pp. 258–279. Available at: <https://doi.org/10.1080/1747423X.2019.1643422>.
- Piggott, J.J., Townsend, C.R. and Matthaei, C.D. (2015) ‘Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics’, *Global Change Biology*, 21(5), pp. 1887–1906. Available at: <https://doi.org/10.1111/gcb.12861>.
- Pont, D. *et al.* (2018) ‘Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation’, *Scientific Reports*, 8(1), p. 10361. Available at: <https://doi.org/10.1038/s41598-018-28424-8>.
- Renó, V., Novo, E. and Escada, M. (2016) ‘Forest Fragmentation in the Lower Amazon Floodplain: Implications for Biodiversity and Ecosystem Service Provision to Riverine Populations’, *Remote Sensing*, 8(11), p. 886. Available at: <https://doi.org/10.3390/rs8110886>.
- RStudio Team (2020) ‘RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.’
- Salmah, M. *et al.* (2013) ‘Local effects of forest fragmentation on diversity of aquatic insects in tropical forest streams: Implications for biological conservation’, *Aquatic Ecology*, 74, pp. 75–85. Available at: <https://doi.org/10.1007/s10452-012-9426-8>.
- Sass, G.G. (2009) ‘Coarse Woody Debris in Lakes and Streams’, in G.E. Likens (ed.) *Encyclopedia of Inland Waters*. Oxford: Academic Press, pp. 60–69. Available at: <https://doi.org/10.1016/B978-012370626-3.00221-0>.
- Sawaya, N.A. *et al.* (2019) ‘Assessing eukaryotic biodiversity in the Florida Keys National Marine Sanctuary through environmental DNA metabarcoding’, *Ecology and Evolution*, 9(3), pp. 1029–1040. Available at: <https://doi.org/10.1002/ece3.4742>.
- Scheffer, M. *et al.* (2006) ‘Small habitat size and isolation can promote species richness: second-order effects on biodiversity in shallow lakes and ponds’, *Oikos*, 112(1), pp. 227–231. Available at: <https://doi.org/10.1111/j.0030-1299.2006.14145.x>.
- Schlesinger, W.H. and Bernhardt, E.S. (2020) *Biogeochemistry: an analysis of global change*. Fourth edition. London: Academic Press.

Skole, D. and Tucker, C. (1993) 'Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988', *Science*, 260(5116), pp. 1905–1910. Available at: <https://doi.org/10.1126/science.260.5116.1905>.

Stanford, B. *et al.* (2020) 'In-stream habitat and macroinvertebrate responses to riparian corridor length in rangeland streams', *Restoration Ecology*, 28(1), pp. 173–184. Available at: <https://doi.org/10.1111/rec.13029>.

Sundar, S. *et al.* (2020) 'Conservation of freshwater macroinvertebrate biodiversity in tropical regions', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 30(6), pp. 1238–1250. Available at: <https://doi.org/10.1002/aqc.3326>.

Sweeney, B.W. *et al.* (2004) 'Riparian deforestation, stream narrowing, and loss of stream ecosystem services', *Proceedings of the National Academy of Sciences*, 101(39), pp. 14132–14137. Available at: <https://doi.org/10.1073/pnas.0405895101>.

Taberlet, P. *et al.* (2012) 'Environmental DNA', *Molecular Ecology*, 21(8), pp. 1789–1793. Available at: <https://doi.org/10.1111/j.1365-294X.2012.05542.x>.

Taberlet, P. *et al.* (2018) *Environmental DNA: For Biodiversity Research and Monitoring*, *Environmental DNA: For Biodiversity Research and Monitoring*. Available at: <https://doi.org/10.1093/oso/9780198767220.001.0001>.

Thomsen, P.F. *et al.* (2012) 'Monitoring endangered freshwater biodiversity using environmental DNA', *Molecular Ecology*, 21(11), pp. 2565–2573. Available at: <https://doi.org/10.1111/j.1365-294X.2011.05418.x>.

Tréguier, A. *et al.* (2014) 'Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds', *Journal of Applied Ecology*, 51(4), pp. 871–879. Available at: <https://doi.org/10.1111/1365-2664.12262>.

Tsuji, S., Yamanaka, H. and Minamoto, T. (2017) 'Effects of water pH and proteinase K treatment on the yield of environmental DNA from water samples', *Limnology*, 18(1), pp. 1–7. Available at: <https://doi.org/10.1007/s10201-016-0483-x>.

Turner, I.M. (1996) 'Species Loss in Fragments of Tropical Rain Forest: A Review of the Evidence', *Journal of Applied Ecology*, 33(2), pp. 200–209. Available at: <https://doi.org/10.2307/2404743>.

Turunen, J. *et al.* (2019) 'Riparian forests mitigate harmful ecological effects of agricultural diffuse pollution in medium-sized streams', *Science of The Total Environment*, 649, pp. 495–503. Available at: <https://doi.org/10.1016/j.scitotenv.2018.08.427>.

Turunen, J. *et al.* (2021) 'Riparian forests can mitigate warming and ecological degradation of agricultural headwater streams', *Freshwater Biology*, 66(4), pp. 785–798. Available at: <https://doi.org/10.1111/fwb.13678>.

Valentini, A. *et al.* (2016) 'Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding', *Molecular Ecology*, 25(4), pp. 929–942. Available at: <https://doi.org/10.1111/mec.13428>.

Van Driessche, C. *et al.* (2022) ‘Experimental assessment of downstream environmental DNA patterns under variable fish biomass and river discharge rates’, *Environmental DNA*, n/a(n/a). Available at: <https://doi.org/10.1002/edn3.361>.

Verissimo, A. *et al.* (1992) ‘Logging impacts and prospects for sustainable forest management in an old Amazonian frontier: The case of Paragominas’, *Forest Ecology and Management*, 55(1), pp. 169–199. Available at: [https://doi.org/10.1016/0378-1127\(92\)90099-U](https://doi.org/10.1016/0378-1127(92)90099-U).

Villacorta-Rath, C. *et al.* (2021) ‘Long distance (>20 km) downstream detection of endangered stream frogs suggests an important role for eDNA in surveying for remnant amphibian populations’, *PeerJ*, 9, p. e12013. Available at: <https://doi.org/10.7717/peerj.12013>.

Wang, S. *et al.* (2021) ‘Methodology of fish eDNA and its applications in ecology and environment’, *Science of The Total Environment*, 755, p. 142622. Available at: <https://doi.org/10.1016/j.scitotenv.2020.142622>.

Wang, X. *et al.* (2019) ‘The biodiversity benefit of native forests and mixed-species plantations over monoculture plantations’, *Diversity and Distributions*, 25(11), pp. 1721–1735. Available at: <https://doi.org/10.1111/ddi.12972>.

Watts, C. *et al.* (2019) ‘DNA metabarcoding as a tool for invertebrate community monitoring: a case study comparison with conventional techniques’, *Austral Entomology*, 58(3), pp. 675–686. Available at: <https://doi.org/10.1111/aen.12384>.

Weigand, H. *et al.* (2019) ‘DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work’, *Science of The Total Environment*, 678, pp. 499–524. Available at: <https://doi.org/10.1016/j.scitotenv.2019.04.247>.

Wickham, H. (2009) *ggplot2*. New York, NY: Springer New York. Available at: <https://doi.org/10.1007/978-0-387-98141-3>.

Yu, D.W. *et al.* (2012) ‘Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring’, *Methods in Ecology and Evolution*, 3(4), pp. 613–623. Available at: <https://doi.org/10.1111/j.2041-210X.2012.00198.x>.

Zhang, J. *et al.* (2014) ‘PEAR: a fast and accurate Illumina Paired-End reAd mergeR’, *Bioinformatics*, 30(5), pp. 614–620. Available at: <https://doi.org/10.1093/bioinformatics/btt593>.

Zwick, P. (1992) ‘Stream habitat fragmentation — a threat to biodiversity’, *Biodiversity & Conservation*, 1(2), pp. 80–97. Available at: <https://doi.org/10.1007/BF00731036>.

# Appendix A – Summary of primers

## Tele03

Target taxonomic group	Teleostei (Teleost fishes)
Target gene	12S mitochondrial rDNA
Forward primer	GTCGGTAAACTCGTGCCAGC
Reverse primer	CATAGTGGGGTATCTAATCCCAGTTTG
Mean length (bp)	171 bp
Reference	(Miya <i>et al.</i> , 2015; Taberlet <i>et al.</i> , 2018)

## BF2/BR2

Target taxonomic group	Aquatic invertebrates
Target gene	cytochrome c oxidase I (COI)
Forward primer (BF2)	GCHCCHGAYATRGCHTTYCC
Reverse primer (BR2)	TCDGGRTGNCCRAARAAYCA
Mean length (bp)	421 bp
Reference	(Elbrecht and Leese, 2017)



# Appendix B – Lab protocols

## **Bulk aquatic invertebrate extraction protocol**

CTAB and CIAA (Chloroform Isoamyl Alcohol) protocol (Doyle and Doyle, 1987)

Preheat water bath to 65°C.

1. Mixing 650µl of CTAB and 20µl of proteinase K to the samples before incubation (65°C) for four hours.
2. After incubation, add 650µl of CIAA (chloroform isoamyl alcohol) and invert 50 times inside the vacuum chamber.
3. Centrifuge for 10 minutes with 10300 rpm.
4. Collect 500µl of the supernatant and transfer into a new Eppendorf tube.
5. Add 500µl isopropanol (kept on ice) and invert 50 times inside the vacuum chamber.
6. Put samples in the freezer overnight.
7. Centrifuge samples in 10 minutes at 10300 rpm and remove excess liquid.
8. Washing process: add 1000µl of 70% cold ethanol and work on ice. Centrifuge at 10300 rpm and repeat. Discard the alcohol and let samples dry.

## **Water eDNA extraction protocol**

Water samples were extracted by using a modified DNeasy PowerWater Sterivex Kit protocol provided by QIAGEN (*DNeasy PowerWater Sterivex Kit Quick Start Protocol - QIAGEN*).



## PCR parameters

Bulk aquatic invertebrate (COI)		
Temp	Duration	Cycles
98	30 s	1
98	10 s	
60	20 s	29
72	15 s	
72	2 min	1

Water eDNA (Tele03)		
Temp	Duration	Cycles
98	30 s	1
98	10 s	
65	15 s	31
71	10 s	
72	2 min	1

## Qubit dsDNA HS Assay protocol

2µl of the normalized samples were used and calculations were adjusted to number of samples:

1.1 Set up the required number of 0.5-mL tubes for standards and samples. The Qubit® dsDNA HS Assay requires 2 standards. Note: Use only thin-wall, clear, 0.5-mL PCR tubes. Acceptable tubes include Qubit® assay tubes (Cat. no. Q32856) or Axygen® PCR-05-C tubes (part no. 10011-830).

1.2 Label the tube lids. Note: Do not label the side of the tube as this could interfere with the sample read. Label the lid of each standard tube correctly. Calibration of the Qubit® Fluorometer requires the standards to be inserted into the instrument in the right order.

1.3 Prepare the Qubit® working solution by diluting the Qubit® dsDNA HS Reagent 1:200 in Qubit® dsDNA HS Buffer. Use a clean plastic tube each time you prepare a Qubit® working solution. Do not mix the working solution in a glass container. Note: The final volume in each tube must be 200 µL. Each standard tube requires 190 µL of Qubit® working solution, and each sample tube requires anywhere from 180–199 µL. Prepare sufficient Qubit® working solution to accommodate all standards and samples. For example, for 8 samples, prepare enough working solution for the samples and 2 standards: ~200 µL per tube in 10 tubes yields 2 mL of working solution (10 µL of Qubit® reagent plus 1990 µL of Qubit® buffer).

- 1.4 Add 190  $\mu\text{L}$  of Qubit® working solution to each of the tubes used for standards.
- 1.5 Add 10  $\mu\text{L}$  of each Qubit® standard to the appropriate tube, then mix by vortexing 2–3 seconds. Be careful not to create bubbles. Note: Careful pipetting is critical to ensure that exactly 10  $\mu\text{L}$  of each Qubit® standard is added to 190  $\mu\text{L}$  of Qubit® working solution.
- 1.6 Add Qubit® working solution to individual assay tubes so that the final volume in each tube after adding the sample is 200  $\mu\text{L}$ . Note: Your sample can be anywhere from 1–20  $\mu\text{L}$ . Add a corresponding volume of Qubit® working solution to each assay tube: anywhere from 180–199  $\mu\text{L}$ .
- 1.7 Add each sample to the assay tubes containing the correct volume of Qubit® working solution, then mix by vortexing 2–3 seconds. The final volume in each tube should be 200  $\mu\text{L}$ .
- 1.8 Allow all tubes to incubate at room temperature for 2 minutes.

### Fragment Analyzer Automated CE System

#### Protocol for the DNF-488 High Sensitivity Genomic DNA Analysis Kit:

1. Mixed 15ml gel and 1.50 $\mu\text{l}$  dye. Refilled 1X conditioning solution as needed.
2. Fresh 1X inlet buffer tray was placed in the Fragment Analyzer. Replaced 1.0 ml of inlet buffer and storage solution in well.
3. Replaced 200 $\mu\text{l}$  rinse buffer and placed rinse buffer plate in marked drawer location.
4. Mixed 2 $\mu\text{l}$  samples (+ 2 $\mu\text{l}$  MilliQ water) with 22 $\mu\text{l}$  diluent marker thoroughly and added 24 $\mu\text{l}$  blank solution to unused wells.
5. Run software.

#### Reagents required

Genomic DNA separation gel	High sensitivity genomic DNA diluted marker
Intercalating dye	High sensitivity genomic DNA ladder
5X 930 dsDNA inlet buffer	BF-25 Blank solution
5X Capillary condition solution	Capillary storage solution
0.25X TE rinse buffer	

## Gel

We used 24 samples to be analysed for the 12-capillary Fragment Analyzer System.

# of samples to be analysed	Volume of intercalating dye	Volume of gel
24	1.50µl	12 ml

## **Specifications**

Specifications	Description
DNA sizing range	50bp – 40 000bp
gDNA Concentration range	50pg/µl to 5ng/µl input DNA
gDNA Quantification range	± 25%
gDNA Quantification Precision	15% CV
Maximum gDNA concentration	5 ng/µl

## AMPure XP protocol

Library	Ratio	DNA	AMPure
1	1:1	370 $\mu$ l	370 $\mu$ l
2	1:1.8	187 $\mu$ l	330 $\mu$ l

### 6. Clean up of Amplified Library

6.1. Add 1X (lib 1) or 1.8X (lib 2) volume of AMPure XP Reagent to the sample and mix by pipetting up and down.

6.2. Incubate for 5 minutes at room temperature.

6.3. Pulse-spin the tube and place in a magnetic rack for approximately 3 minutes until the beads have collected to the wall of the tube and the solution is clear.

6.4. Carefully remove and discard the supernatant without disturbing the beads.

6.5. Keep the tube on the magnet and add 200  $\mu$ l freshly prepared 80% ethanol. Incubate 30 seconds, and carefully remove and discard the supernatant.

6.6. Repeat Step 6.5 once for a total of two washes. Be sure to remove all visible liquid after the second wash. If necessary, briefly spin the tube/plate, place back on the magnet and remove traces of ethanol with a p10 pipette tip.

6.7. Keeping the tube in the magnetic rack, with the cap open, air dry the beads for up to 5 minutes at room temperature.

Caution: Do not over-dry the beads. This may result in lower recovery of DNA targets. Elute the samples when the beads are still dark brown and glossy looking, but when all visible liquid has evaporated. When the beads turn lighter brown and start to crack, they are too dry.

6.8. Remove the tube from the magnet. Resuspend the beads in 30  $\mu$ l of 0.1X TE and incubate for 2 minutes at room temperature.

6.9. Pulse-spin the tube, return to the magnet until the beads have collected to the wall of the tube and solution is clear.

6.10. Transfer approximately 30  $\mu$ l of supernatant to a fresh tube. Be careful not to transfer any beads.



# Appendix C – Scripts used in R Studio

## Appendix: Arthropods

Rita C. Just Olsen

2022-11-16

### Scripts for arthropods bulk samples

```
#1a. Load the ZOTU table (ZOTU by sample abundance)
bulk_otu_table <- read.table("COI_zotutab_5.txt", header=TRUE, row.names=1)
#Something wrong with colnames, need to edit...A lot to edit 51 columns! EDI
T - fixed, was a # in the file
#1b. convert table to matrix
bulk_otu_table <- as.matrix(bulk_otu_table)

#2a. Load the taxa table
bulk_taxonomy <- read.csv("COI_taxout_forR.txt", sep = "\t", check.names =
FALSE, header = TRUE)
rownames(bulk_taxonomy) <- paste0("Zotu", 1:nrow(bulk_taxonomy))
#2b. convert table to matrix
bulk_taxonomy <- as.matrix(bulk_taxonomy)
#Load metadata
bulk_metadata <- read.table("bulk_metadata_otago_more.txt", header=TRUE, row.names = 1)

#tables conversion to variables according to Phyloseq functions
b_OTU <- otu_table(bulk_otu_table, taxa_are_rows = TRUE)
b_TAX <- tax_table(bulk_taxonomy)
b_META <- sample_data(bulk_metadata)

#Combine the new three data variables into a Phyloseq object, allowing the
concomitant quantitative
#and taxonomical data exploration, along with to any descriptive variable p
resent in metadata file.
bulk_merged <- phyloseq(b_OTU, b_TAX, b_META)

#create a random phylogenetic tree
random_tree = rtree(ntaxa(bulk_merged), rooted=TRUE, tip.label=taxa_names(bulk_merged))
#create the final phyloseq object merging all 4 main components, OTU, TAX,
META and random_tree.
bulk_data <- phyloseq(b_OTU, b_TAX, b_META, random_tree)
```

## Alpha Diversity

### ## BULK FRESHWATER ARTHROPODS

```
#Alpha diversity, calculated with Observed and Shannon indices, on raw data
plot_richness(Bd_Arth, x = "Point", title = "Freshwater Arthropods", color =
"Point", measures = c("Observed", "Shannon")) +
  geom_boxplot(size = 0.65) +
  scale_colour_manual(values = c("indianred2", "skyblue2", "orange2", "dark
seagreen3", "plum3")) +
  theme_bw() +
  theme(plot.title = element_text(size = 16), axis.title = element_text(siz
e = 16)) +
  xlab(NULL)

#Kruskal-wallis chi squared for non-normally distributed data
chi_arth <- kruskal.test(richness_arth$Observed ~ sample_data(Bd_Arth)$Poin
t)
```

## Relative abundance

### *#Relative abundance for arthropod orders*

```
data_ord4 <- Bd_Hex_norm_merged_bulk %>%
  tax_glom(taxrank = "order") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>% #transformation to
relative abundance
  psmelt() %>%
  #filter(Abundance > 0.01) %>% #if you do not want any filter,
just comment the line
  arrange(order)
display.brewer.all()

colourCount = 12
#sample_data(data_ord4)
getPalette = colorRampPalette(brewer.pal(12, "Set3"))
ggplot(data_ord4, aes(x=factor(x = Sample, levels = c("AI1A", "AI1D", "AI1
E", "AIP1A", "AIP1B", "AIP1D", "AIP1E",
"AI2A", "AI2B", "AI2
C", "AI2D", "AI2E", "AIP2A",
"AIP2B", "AIP2C", "AIP
2D", "AIP2E", "AI3B", "AI3C", "AI3D", "AI3E",
"AIP3A", "AIP3B", "AIP3C
", "AIP3D", "AIP3E", "AI4A", "AI4B",
"AI4C", "AI4D", "AI4E",
"AIP4A", "AIP4B", "AIP4C", "AIP4D", "AIP4E",
"AI5A", "AI5B", "AI5C",
"AI5D", "AI5E", "AIP5A", "AIP5B", "AIP5C",
"AIP5D", "AIP5E"))) , y
= Abundance, fill = order)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = getPalette(colourCount)) +
  theme_light() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), axis.title = elem
ent_text(size = 16)) +
```

```
ylab("Relative Abundance Order") +
xlab("Freshwater Arthropod Samples")
```

## Heatmap arthropod taxa orders

```
#Heatmap of freshwater Artropods
h501 <- plot_heatmap(Hex_50_merged, method = "NMDS", distance = "bray",
                    taxa.label = "order", taxa.order = "order",
                    na.value = "white", low="darkseagreen1", high = "darkc
yan",
                    sample.order = c("AIP1A", "AIP1D", "AI1E", "AI2A",
"AIP2A", "AIP2B", "AIP2C", "AI2C", "AI2D",
                    "AIP2D",
, "AI2E", "AIP2E", "AIP3A", "AI3B", "AIP3B", "AI3C", "AIP3C",
                    "AI3D",
"AIP3D", "AI3E", "AIP3E", "AI4A", "AIP4A", "AI4B", "AIP4B", "AI
4C",
                    "AIP4C"
, "AI4D", "AIP4D", "AIP4E", "AI5A", "AIP5A", "AI5B", "AIP5B",
"AI5C",
                    "AIP5C"
, "AI5D", "AIP5D", "AI5E", "AIP5E"))

h501 + theme(text = element_text(size = 15))
```

## Ordination and NMDS plot

```
#Ordination, NMDS
ord.bray <- ordinate(Bd_Hex_norm1, method="NMDS", distance="bray")
plot_ordination(Bd_Hex_norm1, ord.bray, color="Point") + theme(aspect.ratio
=1) +
geom_point(size=2) + stat_ellipse(type = "norm", size=1)
```

## PERMANOVA

```
#Observed difference
set.seed(1)
#jaccard distance
physeq_jac <- phyloseq::distance(Bd_Hex_norm1, method= "jaccard")
#PERMANOVA analysis to test for significant difference between points
arth_perm <- adonis2(physeq_jac ~ sample_data(Bd_Hex_norm1)$Point)
#Betadisper
disp2 = betadisper(physeq_jac, group = data.frame(sample_data(Bd_Hex_norm1)
)$Point)
set.seed(777)
arth_bet <- permutest(disp2, permutations = 1000)
```



## Canopy cover abundance

```
#Barplot of Canopy cover, abundance in forested and deforested points
Bulk_forest <- merge_samples(Bd_Hex_norm1, "Forest_cover")
plot_bar(Bulk_forest, fill = "Forest_cover", title = "Canopy Cover Arthropods") +
  geom_bar(aes(color=Forest_cover, fill=Forest_cover), stat="identity", position="stack") +
  theme(legend.position = 'none', text = element_text(size = 16), axis.text.x = element_text(angle = 90, hjust = 1)) +
  xlab(NULL)

#Forest vs deforest richness
pair_w_arth <- pairwise.wilcox.test(richness_arth$Shannon, sample_data(Bd_Arth)$Forest_cover)
```

## PERMANOVA

```
#PERMANOVA and betadisp
set.seed(1)
#jaccard distance
physeq_jac <- phyloseq::distance(Bd_Hex_norm1, method= "jaccard")
#PERMANOVA analysis to test for significant difference between outer points and centre of forest
arth_jac_perm <- adonis2(physeq_jac ~ sample_data(Bd_Hex_norm1)$Forest_cover)
#Betadisper
disp2 = betadisper(physeq_jac, group = data.frame(sample_data(Bd_Hex_norm1)$Forest_cover))
set.seed(12345)
art_bet_f <- permutest(disp2, permutations = 1000)
```

# Appendix: Fish

Rita C. Just Olsen

2022-11-16

## Scripts for water eDNA samples

```
#1a. Load the ZOTU table (ZOTU by sample abundance)
eDNA_otu_table <- read.table("12S_zotutab_5.txt", row.names=1, header = TRUE)
#1b. convert table to matrix
eDNA_otu <- as.matrix(eDNA_otu_table)
#write.table(wtr_otu_table, file='wtr_OTU_tab.xlsc')
#2a. Load the taxa assignments table (OTU - Lineage - taxa by best hit)
eDNA_taxonomy <- read.csv("12S_taxonomy_fish.txt", sep = "\t", check.names = FALSE, header = TRUE)
rownames(eDNA_taxonomy) <- paste0("Zotu", 1:nrow(eDNA_taxonomy))
#2b. convert table to matrix
eDNA_taxonomy <- as.matrix(eDNA_taxonomy)
#Load metadata
eDNA_metadata <- read.table("12S_metadata_more.txt", header=TRUE, row.names = 1)
#tables conversion to variables according to Phyloseq functions
e_OTU <- otu_table(eDNA_otu, taxa_are_rows = TRUE)
e_TAX <- tax_table(eDNA_taxonomy)
e_META <- sample_data(eDNA_metadata)
#Combine the new three data variables into a Phyloseq object, allowing the concomitant quantitative
#and taxonomical data exploration, along with to any descriptive variable present in metadata file.
eDNA_merged <- phyloseq(e_OTU, e_TAX, e_META)

#create a random phylogenetic tree
random_tree = rtree(ntaxa(eDNA_merged), rooted=TRUE, tip.label=taxa_names(eDNA_merged))

#create the final phyloseq object merging all 4 main components, OTU, TAX, META and random_tree.
eDNA_data <- phyloseq(e_OTU, e_TAX, e_META, random_tree)

#remove blanks - eDNA_data -OTAGO
#####
eDNA_data_c1 <- subset_samples(eDNA_data, SampleName != "blank")
eDNA_data_c1 <- prune_taxa(taxa_sums(eDNA_data_c1) > 0, eDNA_data_c1) #this command removes empty OTUs resulting from the subsetting
#####
#keep OTUs with frequency >5% , for example 5% of 100 samples is 5:
#####
e_data_f1 <- filter_taxa(eDNA_data_c1, function(x){sum(x > 0) > 5}, prune = TRUE)
#####
```

```

#
#remove low abundance OTUs (you define your threshold. I like to use < 10%
of the dataset median)
#####
e_data_f2 = filter_taxa(e_data_f1, function(x) sum(x) > 8, TRUE)
#####
###
#Subset Actinopteri
#####
e_data_Actin <- subset_taxa(e_data_f2, class == "Actinopteri")
e_data_Actin <- prune_taxa(taxa_sums(e_data_Actin) > 0, e_data_Actin)

#Subset without cleaning, for richness estimates
data_Actin <- subset_taxa(eDNA_data, class == "Actinopteri")
data_Actin <- prune_taxa(taxa_sums(data_Actin) > 0, data_Actin)

#Remove samples from point 4
data_Actin <- subset_samples(data_Actin, SampleName != "W4")
data_Actin <- prune_samples(sample_sums(data_Actin) > 0, data_Actin)

#Normalize number of reads in each sample using median sequencing depth.###
#####
total = median(sample_sums(e_data_Actin))
standf = function(e_data_Actin, t=total) round(t * (e_data_Actin / sum(e_da
ta_Actin)))
e_data_trans = transform_sample_counts(e_data_Actin, standf)

```

## Alpha diversity

```

#Estimate alpha diversity
richness_eDNA <- estimate_richness(eDNA_merged, measures=c("Observed", "Sh
annon"))
#Kruskal-wallis chi squared for non-normally distributed data for multiple
groups
kruskal.test(richness_eDNA$Shannon ~ sample_data(eDNA_merged)$Point)

#Estimate total richness with observed and Shannon index to provide info ab
out richness and evenness
e_alpha_diversity <- plot_richness(data_Actin, x = "Point", title = "eDNA Fi
sh", color = "Point", measures = c("Observed", "Shannon")) +
  geom_boxplot(size = 0.70) +
  scale_colour_manual(values = c("indianred2", "skyblue2", "orange2", "medi
umseagreen")) +
  theme_bw() +
  theme(plot.title = element_text(size = 16), text = element_text(size = 15
)) +
  xlab(NULL)

```

## Relative Abundance

```

# Relative abundance eDNA
data_ord_e_trans <- e_data_Actin %>%
  tax_glom(taxrank = "family") %>%
  transform_sample_counts(function(x) x / sum(x) ) %>% #transformation to

```

### *relative abundance*

```
filter_taxa(function(x) mean(x) > 1e-5, TRUE) %>%
psmelt() %>%
  #filter(Abundance > 0.01) %>%           #if you do not want any filter,
just comment the line
  arrange(family)
display.brewer.all()

colourCount = 11
sample_data(data_ord_e_trans)
getPalette = colorRampPalette(brewer.pal(11, "BrBG"))
ggplot(data_ord_e_trans, aes(x = Sample, y = Abundance, fill = family)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = getPalette(colourCount)) +
  theme_light() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), axis.title = elem
ent_text(size = 16))+
  ylab("Relative Abundance Family") +
  xlab("eDNA Samples")

data_ord_e_trans$Sample <- factor(data_ord_e_trans$Sample, levels = c("WL1",
"WR1", "WL2", "WM2", "WR2", "WL3", "WM3", "WR3", "WL5", "WM5", "WR5"))
```

### Count ASV per sample

```
# COUNT ASV PER SAMPLE
genfac = factor(tax_table(e_data_trans)[, "family"])
gentab = apply(otu_table(e_data_trans), MARGIN = 2, function(x) {
  tapply(x, INDEX = genfac, FUN = sum, na.rm = TRUE, simplify = TRUE)
})
```

### Canopy cover abundance

```
#Barplot FORESTED VS DEFORESTED eDNA fishies
eDNA_forest <- merge_samples(e_data_trans, "Forest_cover")
plot_bar(eDNA_forest, fill = "Forest_cover", title = "Canopy Cover Fish") +
  geom_bar(aes(color=Forest_cover, fill=Forest_cover), stat="identity", pos
ition="stack") +
  theme(legend.position = 'none', text = element_text(size = 16), axis.text.
x = element_text(angle = 90, hjust = 1)) +
  xlab(NULL)
```

### PERMANOVA

```
#jaccard distance
physeq_jac_e <- phyloseq::distance(e_data_trans, method= "jaccard")
#PERMANOVA analysis to test whether the points differ significantly from ea
ch other
#USE THIS
perm_e_ <- adonis2(physeq_jac_e ~ sample_data(e_data_trans)$Forest_cover)
#Betadisper
disp_e = betadisper(physeq_jac_e, group = data.frame(sample_data(e_data_tra
ns))$Forest_cover)
set.seed(12345)
disp_perm <- permutest(disp_e, permutations = 1000)
```

