Assessing biodiversity:

Beyond the taxonomic impediment

Camila Duarte Ritter

2018

UNIVERSITY OF GOTHENBURG

Faculty of Science
Department of Biological and Environmental Sciences

Opponent: Peter Heintzman
Examiner: Bengt Oxelman
Supervisors: Alexandre Antonelli & R. Henrik Nilsson

Cover image: Photos from different places in Brazilian Amazonia. From left to right and up to down: terra-firme forest from a tower close to Manaus (credit: Aline Ramos); slam trap in igapós forest in Benjamin Constant; terra-firme forest in Juruá (credit: Yenneie Bredin); a *vitória-régia* in the Amazon river (credit: Maurício Denardin); the meeting of Amazon and Negro rivers (credit: Renata Lemes); Jaú National Park on the Negro river; fungi in a terra-firme forest in Benjamin Constant; campina in Novo Airão; the sunrise in front of Cuieras river. Remaining photos by C. D. Ritter.

Understanding nature is no longer enough,

scientists have a moral duty to protect the subject of their study (unknown author).

This work is dedicated to Amazonia.

Table of Contents

Abstract

Precise and accurate quantification of biological diversity is crucial for many fields of biological research and for understanding ecosystem services, biological interactions, biogeography and designing management strategies for conservation. The majority of current taxonomic knowledge is concentrated on a few groups of macro-organisms, mostly vertebrates and flowering plants, which represent only c. 0.7% and 3%, respectively, of the ~11 million estimated species of eukaryotes. The overwhelming majority of the extant biodiversity is challenging to sample and/or identify, which hinders biodiversity studies. While the world's poorly known, inconspicuous organisms (e.g. fungi, insects, nematodes, and bacteria) are essential to understand the evolution, maintenance, and functioning of biodiversity, this taxonomic impediment is the strongest barrier to quantify the diversity of such groups. Recent molecular (DNA-based) tools, such as DNA metabarcoding, promise to speed up biodiversity quantification by several orders of magnitude at moderate costs. These methodological advances allow researchers to circumvent difficult, time-consuming specimen examination and identification, thereby accelerating biodiversity research in poorly studied groups and diverse ecosystems. In addition, these molecular methods make it possible to detect and identify rare and taxonomically challenging species as well as to quantify the biodiversity in virtually any location, which in turn could aid in conservation design and environmental impact assessment (EIA) studies. This thesis examines the potential and challenges of biodiversity assessments and to recommend methods to identify inconspicuous organisms. It also evaluates the use of such methods to assess the variation of biodiversity across space and ecosystems, and to identify the factors underlying the uneven distribution of biological diversity in Amazonia. The results show that the major diversity patterns observed for macro-organisms in Amazonia do not hold true for all organisms. The results also highlight the complementarity of molecular and traditional taxonomic studies to better assess the biotic-abiotic factors that underpin the community composition and turnover of biological diversity.

Keywords: Amazonia; conservation; genetic diversity; metabarcoding; molecular sequences.

Svensk sammanfattning

En noggrann och korrekt kvantifiering av biologisk mångfald är av avgörande betydelse inom många områden av biologisk forskning och för dess tillämpningar, såsom utformning av strategier för bevarande och förståelse av olika ekosystemtjänster, biologiska interaktioner och biogeografi. Nuvarande taxonomisk kunskap är koncentrerad till några få grupper av makro-organismer, främst ryggradsdjur och blomväxter, vilka endast representerar c:a 0,7 % respektive 3 % av de uppskattade 11 miljoner arterna av eukaryoter. Studier av biodiversitet försvåras av att det är tids-, resurs- och kunskapskrävande att samla in och identifiera den biologiska mångfalden. Kunskap om oansenliga och dåligt kända organismer (t. ex. svampar, insekter, nematoder och bakterier) är ytterst viktig för att vi ska förstå hur den biologiska mångfalden utvecklats, upprätthålls och fungerar. Samtidigt utgör den taxonomiska identifieringen och i viss mån insamlingsaspekten de största hindren för att kartlägga dessa gruppers diversitet på ett effektivt sätt. Nya molekylära (DNA-baserade) verktyg, såsom DNA-streckkodning, har potential att kvantifiera biologisk mångfald mycket snabbare än traditionella metoder, och till en avsevärt lägre kostnad. Metodologiska framsteg som dessa gör det möjligt för forskare att kringgå svåra och tidskrävande provundersökningar och identifieringar, med resultatet att forskningen inom den biologiska mångfalden i dåligt studerade grupper och ekosystem går fortare. De molekylära metoderna möjliggör dessutom upptäckt och identifiering av sällsynta och taxonomiskt svårbestämda arter samt fastställande av biologisk mångfald på praktiskt taget vilken plats som helst, något som i sin tur kan bidra till att mäta miljöförändringar och utföra miljökonsekvensbedömningar. Avhandlingens syfte är att undersöka betydelsen och svårigheterna med att mäta biologisk mångfald och att rekommendera molekylära metoder och tillvägagångssätt för att identifiera oansenliga organismer. Den presenterar vidare hur man kan använda dessa metoder för att beskriva variationen av biologisk mångfald i tid och rum samt för att identifiera de faktorer som förklarar den ojämna fördelningen av biologisk mångfald i Amazonas. Ett generellt resultat är att de biodiversitetsmönster som finns i Amazonas inte gäller för alla organismer. Avhandlingen visar också att molekylära och traditionella taxonomiska metoder kompletterar varandra i undersökningen av vilka biotiska och abiotiska faktorer som kan förklara sammansättningen och variationen av den biologiska mångfalden.

Publications included in this thesis

The thesis is based on the following manuscript and papers, referred to in the text by their Roman numerals.

I – **Ritter, C. D**., McCrate, G., Nilsson, R. H., Fearnside, P. M., Palme, U., and Antonelli, A. (2017). *Environmental impact assessment in Brazilian Amazonia: Challenges and prospects to assess biodiversity.* Biological Conservation, 206, pp. 161-168. DOI: 10.1016/j.biocon.2016.12.031

II – **Ritter, C. D.**, Zizka, A., Barnes, C., Nilsson, R. H., Roger, F. and Antonelli, A. (2018). *Locality or habitat? Exploring predictors of biodiversity in Amazonia*. Ecography. DOI: 10.1111/ecog.03833

III – **Ritter, C. D.,** Zizka, A., Roger, F., Tuomisto, H., Barnes, C., Barnes, C., Nilsson, R. H. and Antonelli, A. (2018). *High-throughput metabarcoding reveals the effect of physicochemical soil properties on soil and litter biodiversity and community turnover across Amazonia.* PeerJ Preprints, 6, p. e27012v1. DOI: 10.7287/peerj.preprints.27012v1

IV – **Ritter, C. D.**, Häggqvist, S., Karlsson, D., Sääksjärvi, I.E., Muasya, M., Nilsson, R. H. and Antonelli, A. (in review). *Biodiversity assessment in the 21st century: An integrative protocol for assessing eukaryotic and prokaryotic diversity from soil and insect traps using high-throughput DNA metabarcoding in various environments*. Genome.

V – **Ritter, C. D.**, Faurby, S., Bennett, D. J., Naka, L., ter Steege, H., Haenel, Q., Nilsson, R. H., Antonelli, A. (manuscript). *DNA metabarcoding shows distinct patterns of diversity for Amazonian macro- and micro-organisms.*

Additional publications not included in this thesis

Abarenkov, K., Adams, R. I., Laszlo, I., Agan, A., Ambrosio, E., Antonelli, A., Bahram, M., Bengtsson-Palme, J., Bok, G., Cangren, P., Coimbra, V., Coleine, C., Gustafsson, C., He, J., Hofmann, T., Kristiansson, E., Larsson, E., Larsson, T., Liu, Y., Martinsson, S., Meyer, W., Panova, M., Pombubpa, N., **Ritter, C. D.**, Ryberg, M., Svantesson, S., Scharn, R., Svensson, O., Töpel, M., Unterseher, M., Visagie, C., Wurzbacher, C., Taylor, A. F. S., Kõljalg, U., Schriml, L. and Nilsson, R. H. (2016). *Annotating public fungal ITS sequences from the built environment according to the MIxS-Built Environment standard - a report from a May 23-24, 2016 workshop (Gothenburg, Sweden).* MycoKeys., v. 16, p. 1 - 15.

Nilsson, R. H., Wurzbacher, C., Bahram, M., R. M. Coimbra, V., Larsson, E., Tedersoo, L., Eriksson, J., **Ritter, C. D.**, Svantesson, S., Sánchez-García, M., Ryberg, M. K., Kristiansson, E. and Abarenkov, K. (2016). *Top 50 most wanted fungi*. MycoKeys, v. 12, p. 29 - 40.

Gubili, C., **Ritter, C. D.**, Motta, J., Farias, I. P., Bates, J., Canton, R., Capurucho, J. M., Cornelius, C., Feldheim, K. A. and Ribas, C. C. (2016). *Isolation and characterization of polymorphic microsatellite DNA Markers from an Amazonian white-sand vegetation specialist bird, Xenopipo atronitens (Aves: Pipridae).* The Wilson Journal of Ornithology, 128(3), 668-672.

Antonelli, A., Ariza, M., Albert, J., Andermann, T., Azevedo, J., Bacon, C., Faurby, S., Guedes, T., Hoorn, C., Lohmann, L., Matos-Maraví, P., **Ritter, C. D.**, Sanmartín, I., Silvestro, D., Tejedor, M., ter Steege, T., Tuomisto, H., Werneck, F., Zizka, A. and Edwards, S. (2017). *Conceptual and empirical advances in Neotropical biodiversity research*. PeerJ Preprints, No. e3074v1.

Lentendu, G., Buosi, P. R. B., Cabral, A. F., Segóvia, B. T., Meira, B. R., Lansac-Tôha, F. M.,

Velho, L. F. M., **Ritter, C. D.** and Dunthorn, M. (2018). *Planktonic Protist Biodiversity and Biogeography in Lakes From Four Brazilian River-Floodplain Systems*. bioRxiv, 337824.

Matos-Maraví, P., **Ritter, C. D.**, Barnes, C. J., Nielsen, M., Olsson, U., Wahlberg, N., Marquina, D., Sääksjärvi, I. and Antonelli, A. (2018). *Biodiversity seen through the perspective of insects: 10 simple rules on methodological choices, common challenges, and experimental design for genomic studies.* PeerJ Preprints, 6, e26661v1.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J**., Ritter, C. D.**, Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V. and Antonelli, A. (in review). *CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases*. In review: Methods in Ecology & Evolution.
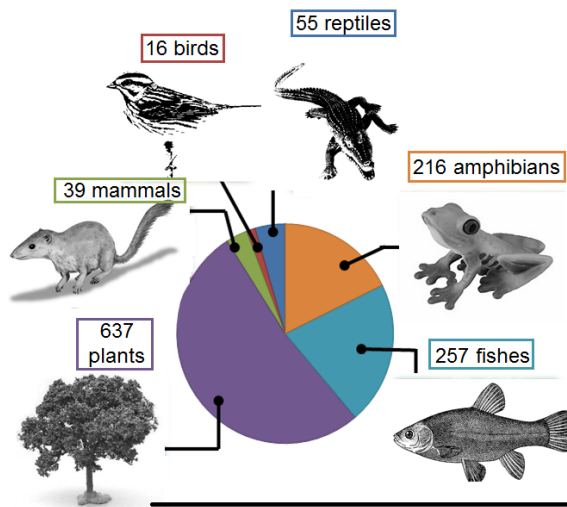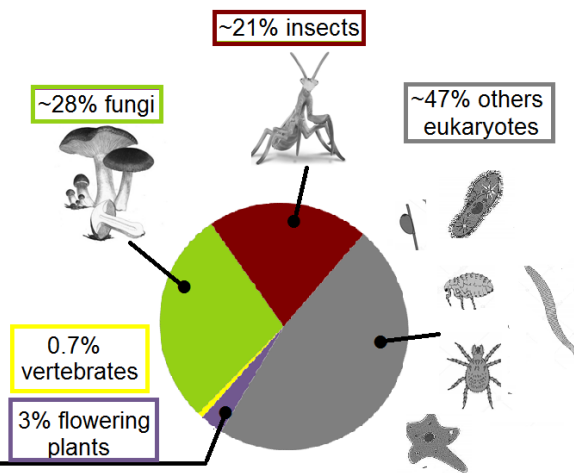
Introduction

Biodiversity assessment

The term biodiversity is relatively recent, formally introduced in the "National Forum on BioDiversity", Washington, DC, 1986 (Wilson & Peter 1988). However, the need and desire to understand biological diversity is much older. A basic and widely used measure to quantify biodiversity is the number of species in a given area or species richness (Magurran 2004). There are currently 1.5 million eukaryote species described, and a recent estimate puts the number at 11 million extant eukaryotes on Earth, although other estimates range from 2 million to 1 trillion (Mora *et al.* 2011, Larsen *et al.* 2017). Impressive as these numbers are, they are dwarfed by the estimated prokaryote richness, estimated between 10^6-10^8 species (Amman *et al.* 1995, Locey & Lennon 2016, Schloss *et al.* 2016).

There is a strong bias in our understanding of biodiversity: due to the ease of sampling and identification in some taxonomic groups, nearly all of their expected diversity has been formally described. This is the case for birds were around 98% of the extant species are thought to be known (Bebber *et al.* 2007, Chapman 2009) and mammals, where around 99% of all extant species are thought to have been formally described (Chapman 2009). However, the overwhelming majority of the extant biodiversity does not belong to these groups. All vertebrates combined represent only 0.7% of the estimated number of eukaryote species, whereas 23-34% are represented by insects (Hamilton *et al.* 2010) and 14-28% by fungi (Fig 1; Hawksworth 2001, Taylor *et al.* 2014).

Figure 1. Comparison of the number of newly described species in Amazonia and the proportion of flowering plants and vertebrates in relation to all described eukaryotes. A) The number of new species described in Amazonia between 1999 and 2009 according to a WWF report (2013). B) Approximate proportion of the most well-studied groups of eukaryotes in relation to all of the Eukarya. The minute region in yellow represents all vertebrates combined (0.7% of all eukaryotes). The grey slice indicates the other eukaryotes (e.g. protists, nematodes and molluscs).

Species identification requires taxonomic expertise, which in turn represents a substantial prior investment of resources and time (Campbell *et al.* 2011). To catalogue all eukaryote species on Earth at the current speed of species discovery and description, some 1,200 years of additional research would be needed (Mora *et al.* 2011). In view of this, the decreasing costs of DNA sequencing combined with new methods for taxonomic assessment may offer an unique opportunity to characterize large sets of organisms or even entire communities using bulk samples or environmental DNA (Fig 2; Gibson *et al.* 2014). These advances came in the wake of DNA barcoding (molecular identification) of species and the existence of public and relatively highly populated reference sequence databases (Hebert *et al.* 2003). For some organisms, no prior taxonomic information is available (*i.e.* they are undescribed), or a complete taxonomic identification is impossible in the absence of sequence

data (as is often the case for bacteria and many fungi). Indeed, molecular data are often the only source of taxonomic information for many groups of microorganisms (Blaxter *et al.* 2005, Tedersoo *et al.* 2014, Vartoukian 2016).
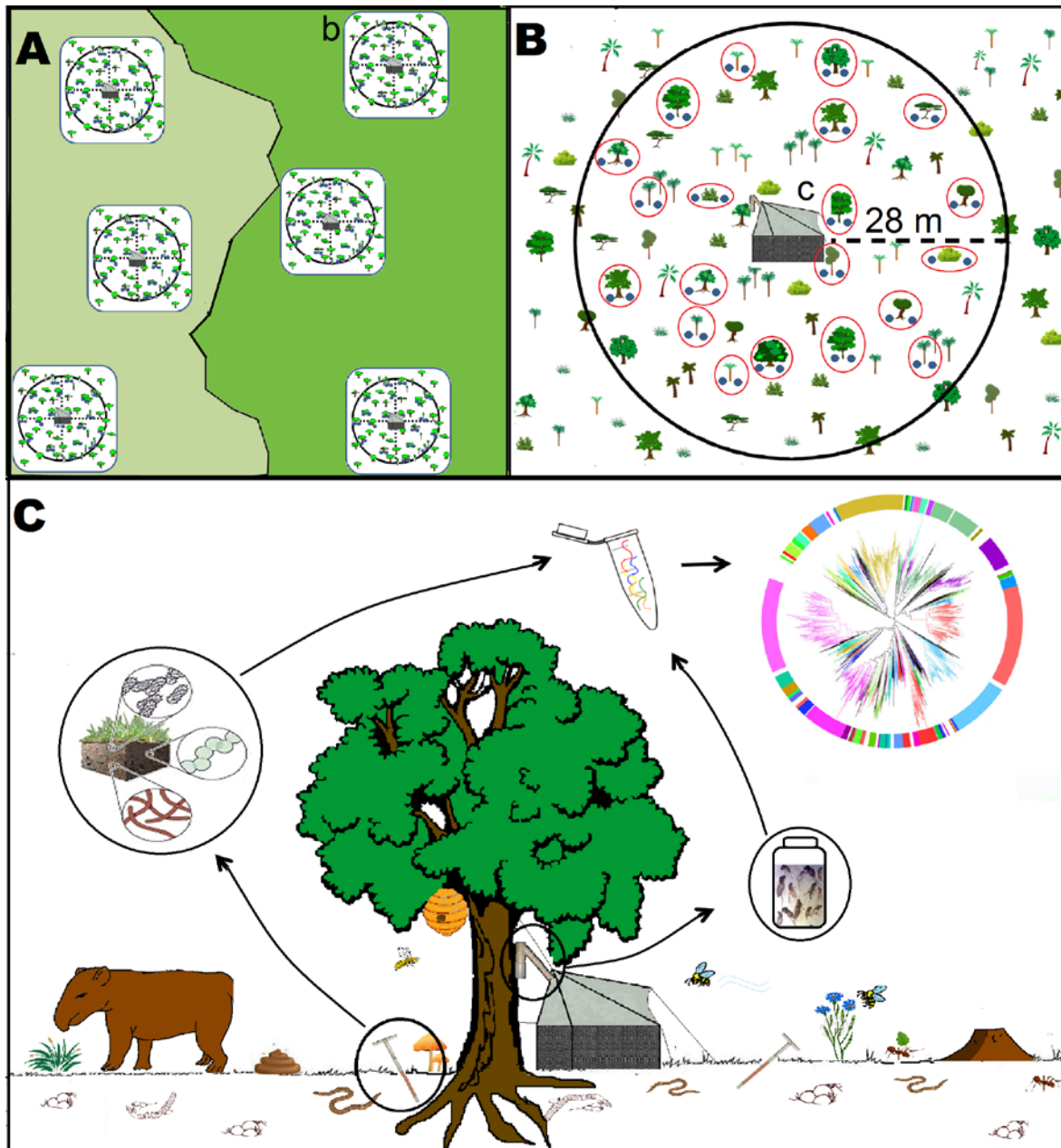
Figure 2. **Sampling strategy to sequence a large proportion of a site's full biological community using environmental DNA (eDNA) and bulk insect samples.** (A) Study design and location of plots in one locality: at least three plots should be set up in each major vegetation type present (illustrated here by different shades of green) in the interest of statistical reproducibility; (B) Scheme for one plot, showing a Malaise trap (used to collect insect samples) in the middle, and twenty trees arbitrarily chosen for soil sampling within a 28-m radius (in red circles); (C) The Malaise trap will capture insects and the eaten plants and parasites in their bodies, while

the soil samples will provide environmental DNA (eDNA) for a large proportion of the habitat's total biodiversity (*e.g.*, soil organisms, roots, leaves, dead animals, and faeces). The soil sampling design is that of Tedersoo *et al.* (2014).

Molecular tools

Molecular data have revolutionized the study of inconspicuous organism groups. Nowadays, DNA sequence are an important source of information on micro-organismal evolution and ecology (Stajich *et al.* 2009, Bik *et al.* 2012, Loman *et al.* 2012, Creer *et al.* 2016). Biodiversity assessments using molecular methods form a powerful tool to understand entire communities and to understand their biotic and abiotic interactions. Even in highly diverse and poorly sampled environments for which reference databases are very sparsely populated, the use of molecular operational taxonomic units (OTUs; Blaxter *et al.* 2005) allows for assessment of genetic diversity and enables comparison among multiple sites (Stahlhut *et al.* 2013). DNA-based studies have the potential to overcome at least some taxonomic limitations and have been identified as a "transformative technology" for the entire field (Baird & Hajibabaei 2012). However, more effort and method development are needed to populate and correct errors in reference sequence databases and to understand the intrinsic differences between taxonomic and molecular biodiversity assessments.

There are many factors to consider when using molecular tools in biodiversity assessments, including DNA extraction, choice of genetic marker (*e.g.* Clarke *et al.* 2014, Elbrecht *et al.* 2016), sequencing method (*e.g.* Liu *et al.* 2013, Schirmer *et al.* 2015), and data analysis procedures (*e.g.* Beng *et al.* 2016). A serious caveat for using these molecular methods for biodiversity assessments is the lack of richly populated taxonomic reference databases, especially for the tropical regions of the globe. Without such reference databases, the recovered sequences cannot be assigned to resolved taxonomic levels. Furthermore, the molecular assessment is hampered by miss-annotated reference sequences, technically

compromised sequences (*e.g.*, chimeras), and reference sequences annotated only at high taxonomic levels (*e.g.* phylum; Kang *et al.* 2010, Nilsson *et al.* 2012). Some recent initiatives try to mitigate these problems, including among others the Global Genomic Observatories (Davies, Field & The Genomic Observatories Network 2012) and the UNITE database for molecular identification of fungi (Abarenkov *et al.* 2016, Kõljalg *et al.* 2013). But even with these efforts it remains challenging to correctly assess many newly generated sequences.

The lack of trustworthy reference sequences has implications for biodiversity assessment. Beyond the obvious problems of erroneous characterization of biodiversity, the use of high (unresolved) taxonomic levels (as a consequence of the partial nature of the information available in reference databases) only provides limited biological information. This is particularly problematic in the context of the current biodiversity crisis, with an increasing loss of biodiversity at the species level (*e.g.* Scholes & Biggs 2005, Barbault 2011, Giam *et al.* 2012), which has been identified as a major challenge to humanity in the next decades (Steffen *et al.* 2015). Correct quantification and measurement of biodiversity are prerequisites for its protection and for efficient conservation of natural ecosystems. However, such quantification is not feasible for many taxonomic groups and most geographical regions using traditional approaches. Considerable sampling and research efforts are necessary, but these are often constrained by time, funding, logistical difficulties, and the availability of taxonomic expertise (Campbell *et al.* 2011, Wheeler 2014).

The difficulty in assessing biodiversity is amplified in mega-diverse and poorly sampled environments such as tropical rainforests (Balmford & Whitten 2003, Giam *et al.* 2012). For instance, hundreds of new species are described each year from the Amazonia rainforest (Fig. 1; *e.g.* Del Hoyo *et al.* 2013, WWF 2013), and the rate of degradation of this biome is alarmingly high (Fig 3; Soares-Filho *et al.* 2006, Malhi *et al.* 2008). This degradation is probably driving many species to extinction before they are even discovered and formally

described (Wearn *et al.* 2012). In this way, even with the limitations of molecular methods, the use of unified protocols to describe diversity patterns and drivers are essential for fast and cost-efficient biodiversity studies (**Paper IV**).
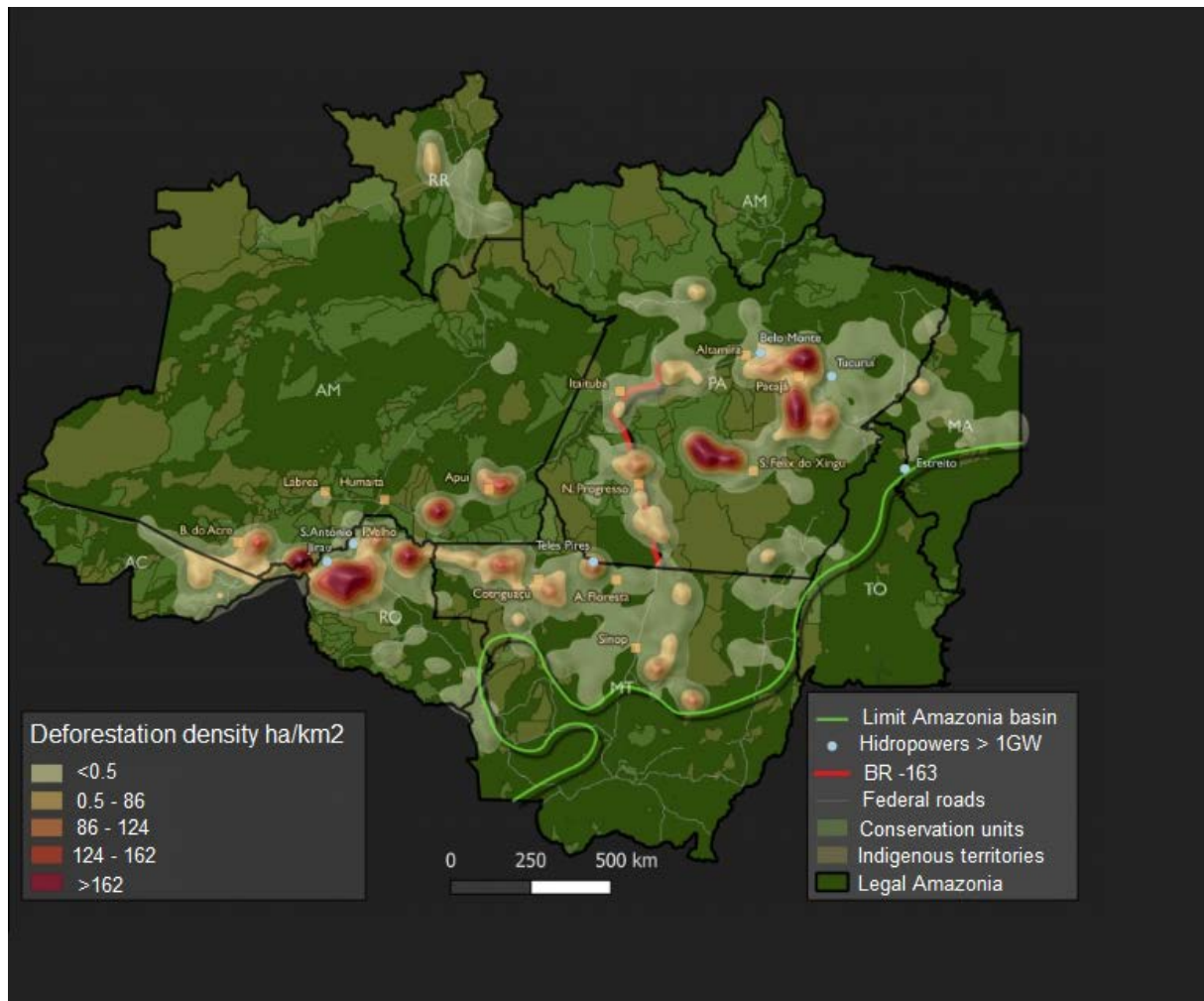
Figure 3. Deforestation density in Brazilian Amazonia during 2015 - 2016. Map modified from *Instituto Socio-Ambiental* – ISA with data generated from: deforestation from INPE/PRODES (2016), conservation units and indigenous territories from ISA (2016), roads and Amazonian limits from IBGE (2012), and hydropower stations from ANEEL (2016). The data were collected between August 2015 and July 2016 and indicated almost 8,000 square kilometres of deforestation, an increase of about 30% in relation to the year before.

Amazonian rainforest

Amazonia is the largest tropical forest in the world. It comprises a tremendously high number of coexisting species, forming a very complex web of life linked together through biotic and abiotic interactions. At the same time Amazonia is considered fragile in the context of current human impact: relatively small alterations may have major impacts, such as the loss of ecosystem functionality (Malhi *et al.* 2008). Review and rigorous evaluation of current methods to protect biodiversity could therefore help to increase the efficiency of conservation.

One such method is the **Environmental Impact Assessment (EIA),** which is a technical and legal system for assisting in environmental management and for supporting sustainable development (Jay *et al.* 2007). Unfortunately, due to economic pressures, EIA requirements are often relaxed in some parts of the world, such as Brazil. It is therefore crucial to evaluate the performance and impact of systems such as the EIA, and outline new tools to improve the assessment and monitoring of diversity for the sustainable development of Amazonia and for the protection of its biodiversity (**Paper I**).

Quantifying biodiversity is crucial, but not enough. To understand how historical and ecological factors influence patterns of diversification and endemism, it is essential to identify the drivers generating and maintaining diversity. Several hypotheses have been suggested to explain the origin and distribution of the Amazonian biota (Leite & Rogers 2013). These hypotheses focus on processes that likely exerted different influences in different regions and periods of time (Ávila-Pires *et al.* 2007, Antonelli *et al.* 2010). Some examples include the zones of endemism between large rivers (Ribas *et al.* 2011); a west to east decrease in tree and animal diversity (Fig, 4; ter Steege *et al.* 2003, Pitman *et al.* 2001, Hoorn *et al.* 2010, Jenkins *et al.* 2015); and the diversity and evolution of vegetation types linked to particular soil characteristics, flooding regime, and nutrient availability. Biodiversity patterns may ultimately be linked to differences in geological formations and historical processes such as basin

formation (Hoorn *et al.* 2010), marine incursions (Webb 1995, Bates 2001, Lovejoy *et al.* 2006), and biological interactions (*e.g.* Chan *et al.* 2008).

However, all these patterns are primarily known from well-studied macro-organismal groups. In order to understand the distribution of all biodiversity and the factors that determine them, additional studies need to consider a much larger proportion of poorly known, inconspicuous but ubiquitous organism groups such as fungi, "protists" (various minute eukaryotes belonging to a range of different eukaryotic groups), and bacteria. These lineages are crucial to nutrient cycling, food-web dynamics, and host-pathogen processes (Stajich *et al.* 2009, Dominati *et al.* 2010, Mahé *et al.* 2017). The use of soil samples and universal genetic markers such as the 16S gene (prokaryotes) and the 18S gene (eukaryotes) would allow us to test general diversity patterns and their drivers (Fig 2; **Paper II**).
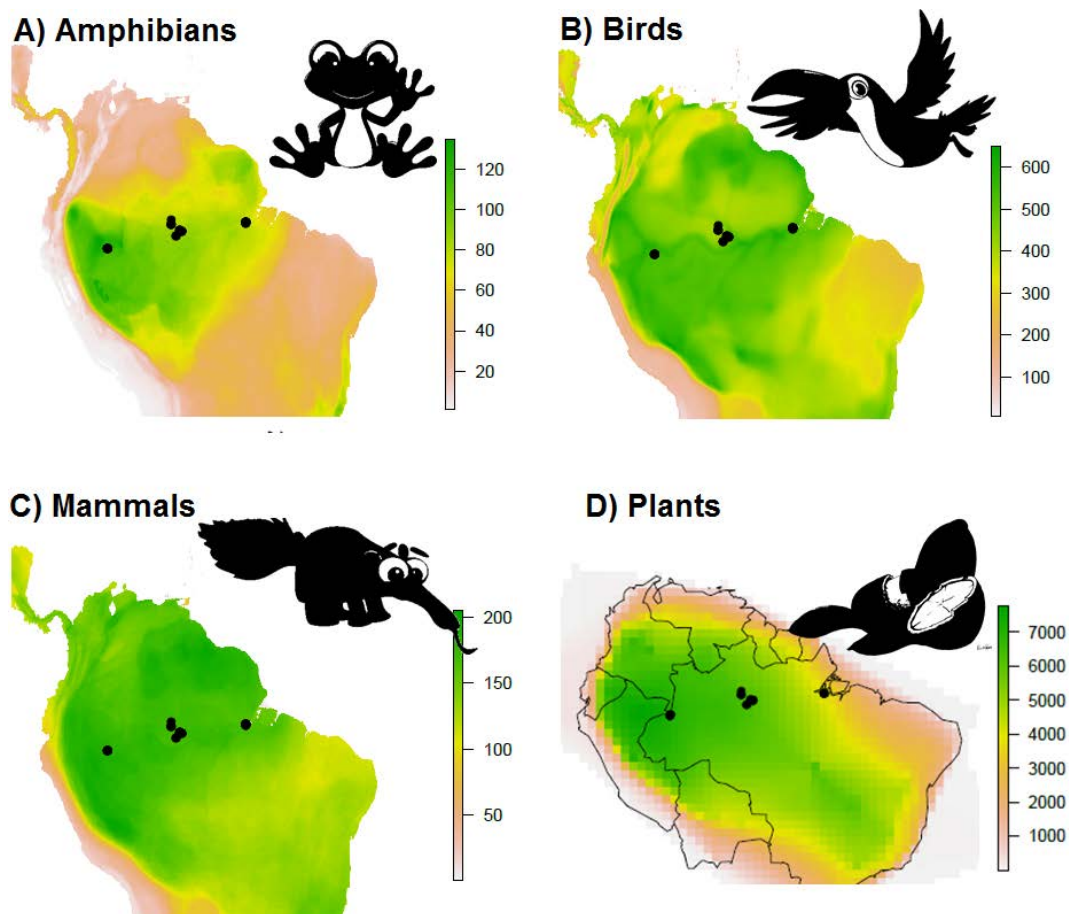
Figure 4. Richness distribution of amphibians, birds, mammals, and vascular plants. A – C) Northern South America's richness patterns of animals as inferred by Jenkins *et al.* (2015), with the sampling in this thesis indicated by black dots. D) Plant richness with data from the Botanical Information and Ecology Network – BIEN (Enquist *et al.* 2016). It is possible to observe some patterns of richness distribution in Amazonia such as a west-to-east richness gradient. The scale bar indicates the number of known species.

Amazonia comprises characteristic vegetation types (here after referred to as environmental types) that are closely linked to soil characteristics, flooding regime, and nutrient availability. Four widespread and important environmental types are terra-firme forests (Fig. 5A, 5B and 5C), várzeas (Fig. 5D, 5E and 5F), igapós (Fig. 5G, 5H and 5I), and naturally open areas (*e.g.* campinas; Fig. 5J, 5K and 5L). These environments support distinct sets of plants and animals, typically associated with mutually different kinds of soil.

Terra-firme forests are considered the most biodiverse environmental type and cover the largest area in Amazonia. They are unflooded and generally characterized by latosols (Falesi 1984). By contrast, várzeas and igapós are seasonally flooded forests. These areas cover approximately 5-7% of the Amazon basin (Peres 1997) and remain submerged during parts of the year, sometimes up to 240 days (Junk *et al.* 1989). Várzeas are flooded by white-water rivers that have their origin in the Andes, from where the floods transport large amounts of nutrient-rich sediments that are deposited in large floodplains. This makes the várzeas fertile areas (Junk *et al.* 2011). Igapós are flooded by black-water rivers that drain the pre-Cambrian Guiana shield, which is characterized by large areas of white sands (podzols). The water of these floods is transparent, with low quantities of suspended matter but with high amounts of humic acids, which give the water a brownish-reddish colour (Junk *et al.* 2011). Finally, the open areas of Amazonia comprise non-forested areas dominated by grasses and shrubs (campinas), and low-canopy forests (campinaranas) (Anderson 1981), which together

cover approximately 1. 6% of the Amazon basin. Open areas resemble islands in a sea of forest and are related to nutrient-impoverished sandy soils (Prance 1996, Fine *et al.* 2005).
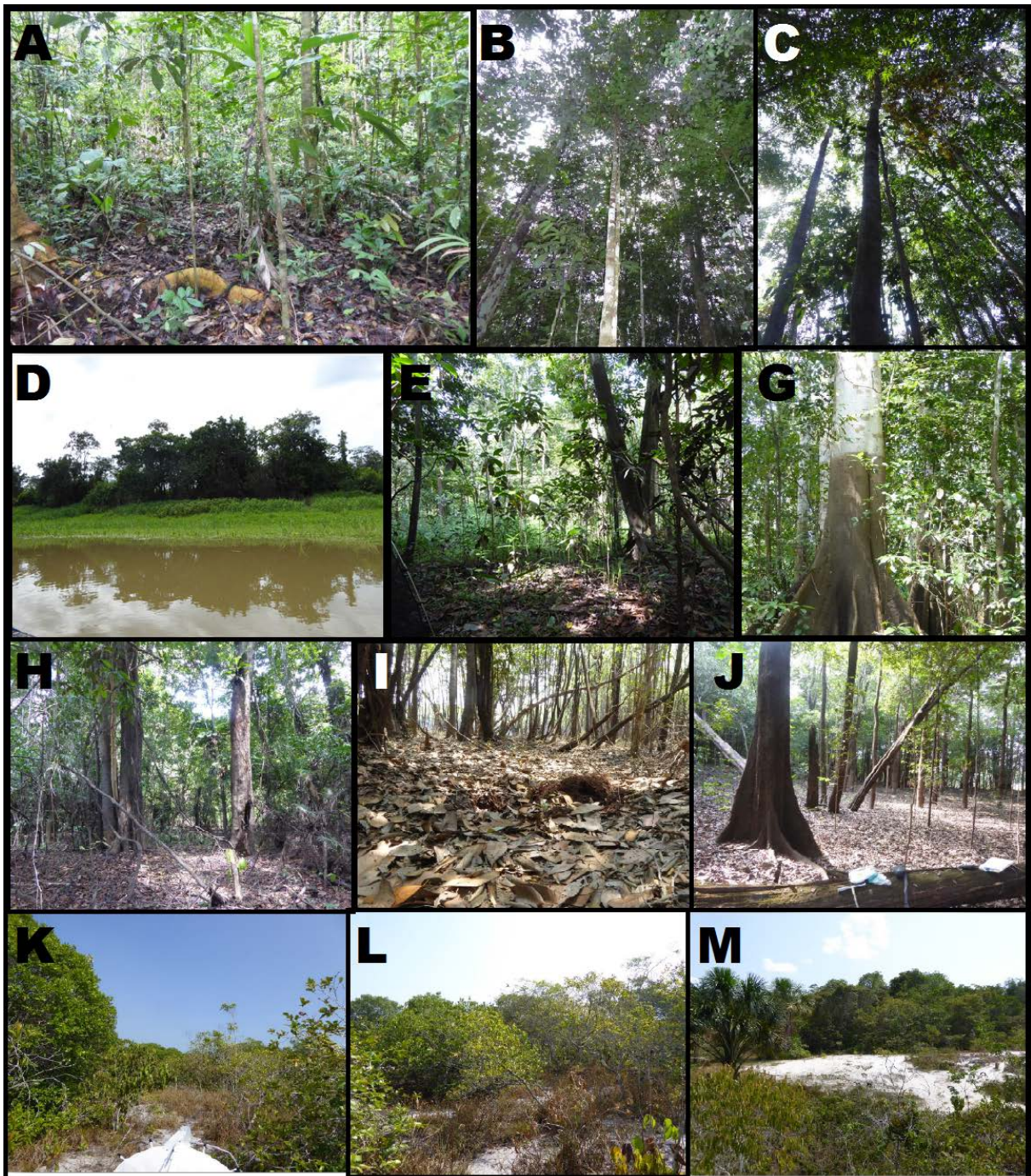
Figure 5. **Major vegetation types in Amazonia.** A - C: Terra-firme (non-flooded forest) is considered the most diverse environmental type in Amazonia. The photos show the high plant density and high canopy (around 40 m); D - F: Várzea forests (seasonally flooded forest) are considered the second most diverse habitat in Amazonia. The photos show the water colour and the high density of plants. F shows a water mark at ca. 15 m height in a várzea forest; G - I: Igapó (seasonally dry water streams). The photos show the low density of plants

and high leaf accumulation in the soils; J - L: Campinas (white-sand areas). Typical features of campinas are sparse, small trees and exposed white sand soils, as can be construed from these photos.

Geology and soil physicochemical characteristics are often considered crucial for biotic dynamics, vegetation, and diversity patterns across Amazonia at local to regional scales (Vogel *et al.* 2009, Laurence *et al.* 2010, Higgins *et al.* 2011). For instance, diversity patterns and community composition have been associated with soil nutrients in plants (Laurence *et al.* 2010). Soil chemistry is associated with bacterial composition (Buckley & Schmidt 2001, Navarrete *et al.* 2013) and phosphorus, is associated with general microbial community composition (Faoro *et al.* 2010). For micro-organism diversity and composition, pH is a very influential factor (*e.g.* Osborne *et al.* 2011, Kuramae *et al.* 2012, Barnes *et al.*, 2016). Even with several studies reporting the importance of soil characteristics on community structure, no unified pattern or scientific consensus has yet been achieved. Understanding the role of soil physicochemical compounds in shaping organism richness and composition of the world's largest rainforest is therefore crucial for identifying the factors underpinning biodiversity patterns (**Paper III**).

Molecular tools such as metabarcoding suffer from a set of limitations. For instance, metabarcoding methods have to compromise between taxonomic coverage and taxonomic resolution. When coupled with universal primers, 18S sequencing may capture the majority of eukaryote organisms. However, 18S is not variable enough to distinguish all eukaryotes at the species level (Hartmann *et al.* 2010, Lindahl *et al.* 2013). Additionally, for 18S the lack of reference sequences for the vast majority of extant eukaryotes is problematic (Guardiola *et al.* 2015). Furthermore, while is an excellent gene to target in environmental metabarcoding studies, no primer pair is truly "universal". There is just too much nucleotide variability in potential primer sites (preventing primer binding; *e.g* Tedersoo *et al.* 2015). The use of other

genetic markers could increase the number of taxonomic groups detected or provide an improved taxonomic resolution for more specific groups. Such markers include the nuclear internal transcribed spacer region (ITS) for plants (Chen *et al.* 2010, Yao *et al.* 2010) and the cytochrome c oxidase subunit I (COI) for metazoans (Hebert *et al.* 2003). With better taxonomic resolution and coverage (*e.g.* use of multiple markers) it is possible to compare the diversity patterns obtained from environmental samples with those of well-studied groups such as birds and trees (**Paper V**).

Objectives

The aim of this thesis is to explore methods to assess biological diversity beyond taxonomic limitations in a diverse ecosystem with little reference data. Furthermore, I use the obteined data, covering a substantial part of the terrestrial biodiversity in the sampled regions, to describe diversity patterns and drivers in the world's largest tropical forest. The thesis consists of two published and two submitted papers as well as one manuscript not yet submitted for publication. These studies have the following objectives:

Paper I – The goal of this study was to discuss the efficiency of Brazilian Environmental Impact Assessments in protecting the biodiversity of Brazilian Amazonia. To this end I assessed three of the most recent and largest infrastructure projects in Amazonia, with a particular focus on whether ecosystem threats and potential environmental impacts were properly assessed in accordance with the EIA principles. I also suggested three cost-effective complementary approaches (remote sensing, reflectance spectroscopy, and DNA metabarcoding) to complement the biodiversity assessment in EIAs.

Paper II – The aim of this manuscript was to test the patterns of biodiversity distribution in Amazonia, using operational taxonomic units (OTUs) derived from 39 soil and litter samples. I examined the data for correlation of eukaryote (18S) and prokaryote (16S) OTU richness and for OTU richness and composition patterns along a longitudinal gradient throughout Amazonia. I furthermore tested whether OTU richness and composition are correlated with Amazonian environmental types related to stress-level and nutrient availability.

Paper III – In this study, I aimed at comparing the relation between OTU richness, effective number of OTUs (abundance-based analysis), and community composition in different soils substrates: litter; the organic matter composed by animal debris, leaves, roots, and other organisms, and the mineral soil. Additionally, I aim to test whether OTU diversity (richness and abundance-based measures) and community turnover correlate with physical and chemical soil properties.

Paper IV – In this paper my goal was to test the potential use of bulk insect samples to assess the diversity of prokaryotes and eukaryotes that are associated with the sampled insects. I also aimed to assess to what extent soil diversity data are complementary to insect sample data in the context of producing more comprehensive biodiversity estimates of any particular site. I finally proposed a universal standardised protocol for metabarcoding studies - one that should be easy to apply in most terrestrial environments. To assess this, we tested our protocol in three countries (Brazil, South Africa, and Sweden) in different habitats.

Paper V – In this manuscript my goal was to compare the richness of environmental DNA data from insects, litter, and soil for three markers (16S – targeting prokaryotes; 18S targeting eukaryotes; and COI targeting metazoans) with the species richness of the some of most well-studied organisms in Amazonia: trees and birds, as assessed by traditional taxonomic studies. I analysed data from four localities in Amazonia in different environments (terra-firme, várzeas, igapós, and campinas). Additionally, I assessed whether similar patterns of diversity in Amazonia could be recovered with our richness data both from OTUs and from taxonomic based species richness.

Methods

Sampling areas (Paper II, III, and V) – Four localities were sampled in Brazilian Amazonia (Fig. 6). We chose these localities to maximize geographic distance and cover all major ecosystem types (terra firme, várzeas, igapós, and campinas). These localities are:

1 - Benjamin Constant – a municipality on the triple border of Brazil, Colombia, and Peru, situated approximately 1,100 km west of Manaus at the upper Solimões River (4°22′58″ S, 70°1′51″ W). This is a very difficult region to access (boat access only), with low population density and relatively low rates of deforestation. The region is situated in the south margin of the Amazonas river, comprising large areas of várzea forests, terra-firme forests, and some igapós forests from black water lakes;

2 - PARNA Jaú – this is a national park with an area of the 2,272,000 ha located in the lower Negro river (1°51'0.00"S, 61°36'60.00"W), 200 km northwest of Manaus, AM. There is a marked seasonality in the water level of Negro river. Generally, the low water season extends from September to November and the high water from May to August (Fundação Vitória Amazônica 1998). About 70% of the forest area is covered by terra-firme forest, which is characterized by large trees in the order of 25 m in height (Borges *et al.* 2001). There is considerable heterogeneity in local plant communities in the terra-firme forests, something that is related to soil mosaics in the region (Fundação Vitória Amazônica 1998). This heterogeneity may also in part be due to human disturbance (Ferreira & Prance 1999). About 12% is covered by igapó forests. We also collected samples from campinas in the municipality of Novo Airão, which is situated in the west margin of the Negro river and for the purpose of this thesis was considered the same locality;

3 - Reserva da Campina – situated 60 km north of Manaus, AM (2°35'30.26"S, 60°01'48.79"W). The reserve comprises approximately 900 ha, of which 6. 5 ha is stunted

heath forest (campinas) and tall heath forest (THF). The campina area (2. 6 ha) is composed of a mosaic of shrub islands surrounded by white bare sandy soil. The canopy height is about 4-7 m for SHF with sparse trees (Luizão *et al.* 2007); and **Reserva do Cuieras** – which covers 22.7 hectares and is located about 70 km north of Manaus, AM ($2^o36'32.67"$S, $60^o12'33.48"$W). The vegetation is a mosaic of evergreen forest with a canopy height of about 35 - 40 m, with emergent trees over 45 m tall, varying to open areas and igapós forest. Analyses of the Igarape Asu catchment (Cuieiras Reserve area) indicated that valley forest environments (igapó and Campinarana) cover 43% of the area, whereas slope and plateau (terra-firme) forests occupy 26% and 31%, respectively (Zanchi *et al.* 2002). These two reserves are situated in the east part of Negro river and for the purpose of this thesis were considered the same locality;

4 - FLONA Caxiuanã – 371,000 hectares of rainforest, located 350 km west of Belém, PA (1°44'07"S, 51°27'47"W), in the lower Amazon region of northern Brazil. About 85% of the forest area is covered by terra-firme forest and about 10% by várzea and igapó forests. This reserve also has some campinas (Behling & Costa 2000).
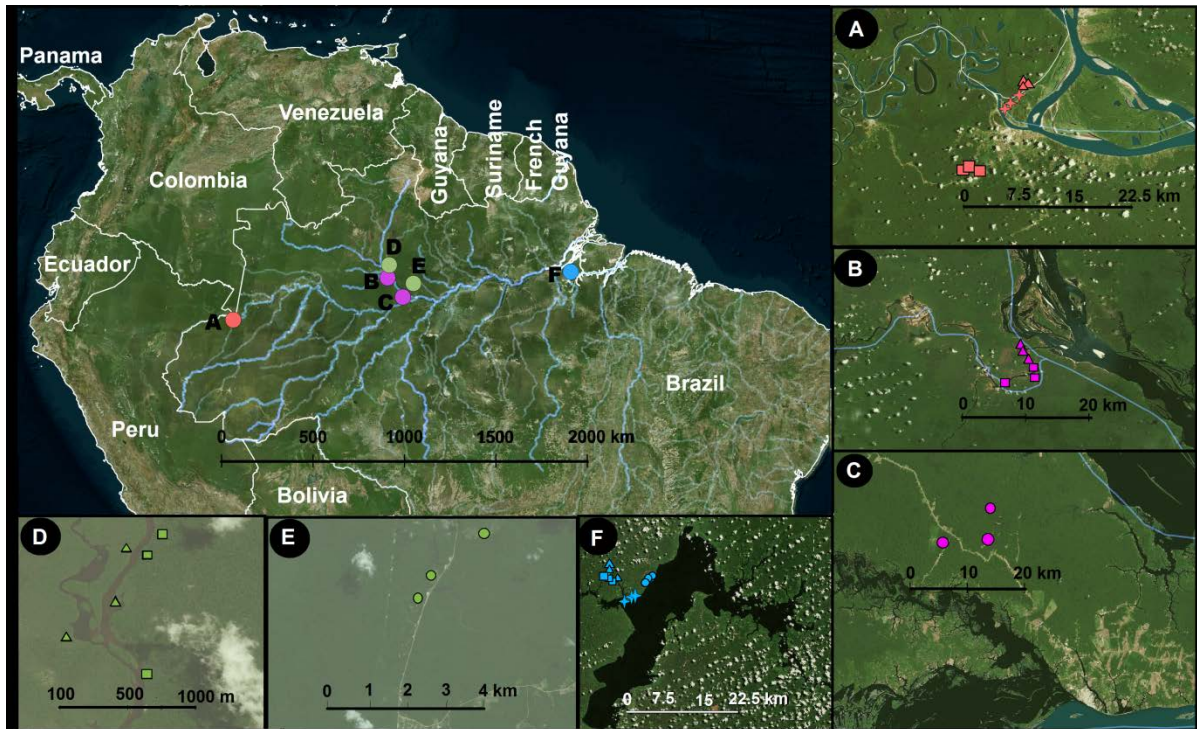
Figure 6. Study area and sampling locations across Amazonia. Inset panels show details of each locality. A: Benjamin Constant; B: Jaú; C: Jaú, naturally open areas; D: Cuieras; E: Cuieras, naturally open areas; and F: Caxiuanã. The symbols in A–F represent different vegetation types characterised by different soil properties: circles = open areas; triangles = forest seasonally flooded by black water rivers; squares = unflooded forest; and crosses = forest seasonally flooded by white water rivers. The sampling strategy was designed to cover a wide longitudinal range in Amazonia. The map was constructed using QGIS (2012). Legend from **Paper III**.

Sampling areas (Paper IV) – In order to test the efficiency of all protocols, in paper IV we used a subsample of Brazilian samples, including all samples from Benjamin Constant (Fig. 7). Additionally, we sampled South Africa and Sweden as described below.

5 - South Africa – Seven plots were sampled with Malaise traps (Fig. 1F) kept open for 24 h. The plots were set up in south-western South Africa (33.51°S, 18.48°E), in a dry coastal area during August 2016. The Cape landscapes are dominated by the Cape Folded Belt (Deacon *et al.* 1992), and the Cape Floristic Region is considered a global biodiversity hotspot (Myers *et al.* 2000). The mean temperature in the sampled period ranges from 10°C to 18°C, and the mean precipitation is 24 mm/month (https://weatherspark. com);

6 - Sweden – Insects were collected with Malaise traps (Fig. 1E) in three forest fragments, three agricultural farms, and three pasture farms. The Malaise traps were deployed for a period of seven days in nine sampling localities. The sampling was done in June 2016. The samples were collected around the city of Skövde, an area characterized by relatively high biodiversity compared to other mainland regions in Sweden, and usually nutrient-rich soils derived from calcareous bedrock (58. 44° N, 13. 66° E). The temperature in the sampled period ranges from 10°C to 21°C, and the mean precipitation is 52 mm/month (https://weatherspark. com);

Legend: IG ● TF ● VZ ● PC1 ● PC2 ● PC3 ● ARA ● FOR ● PAS
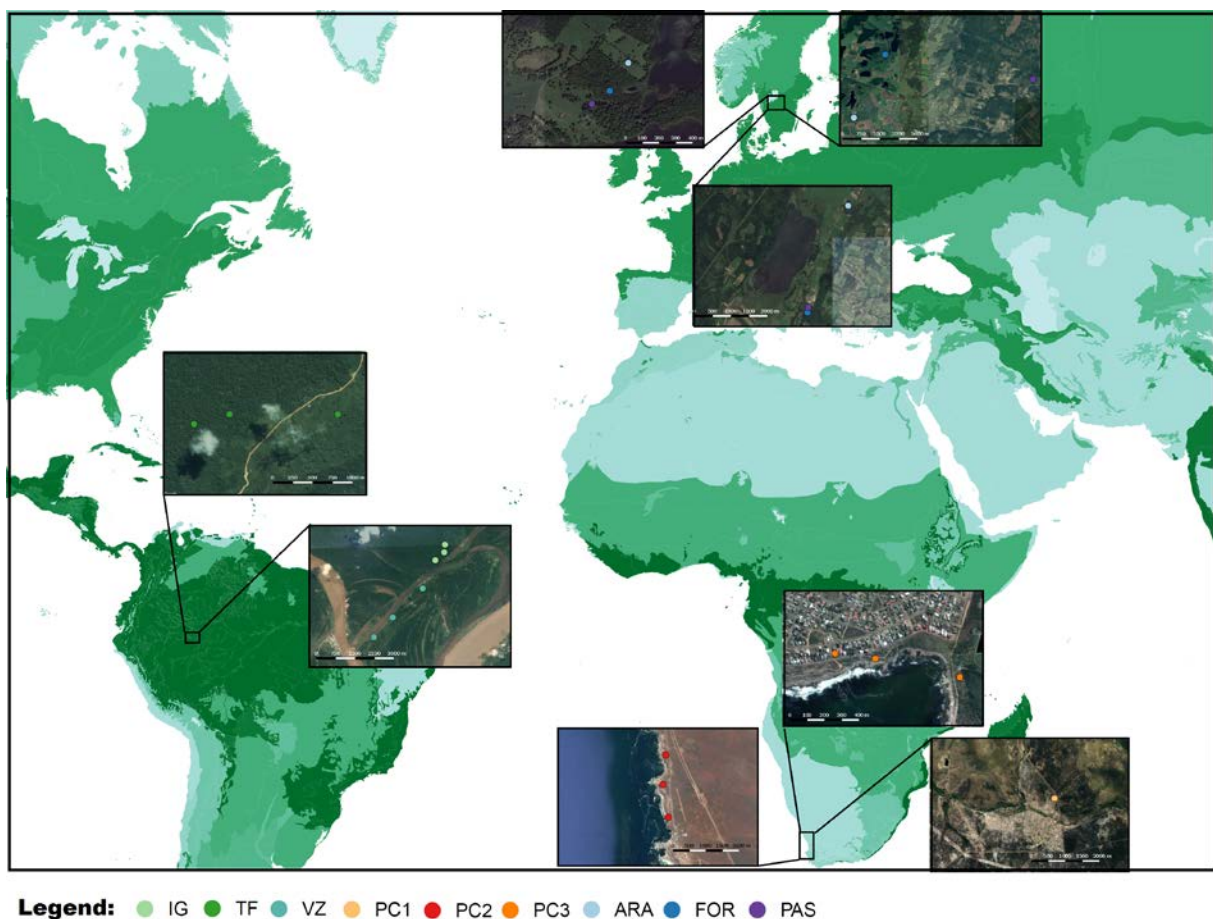
Figure 7. Map of sampling localities (Paper IV). The sampling localities used for this study cover a wide geographical range comprising the Amazonian rainforest in Brazil, the western coast of South Africa, and central-south Sweden. In each locality, different environments were sampled to test the usefulness of our method

in a wide range of habitats. Habitats are: TF = terra-firmes, VZ = várzeas, and IG = igapós in Brazil, PC1 = dry area, PC2 = costal area, and PC3 = costal area more urbanized in South Africa, and ARA = arable farms, FOR = forest fragments, and PAS = pasture farms in Sweden. The green gradient represents biomes described in Olson *et al.* (2001), ranging from densely forested areas (dark green) to open areas (light green). Legend from **Paper IV**.

Sampling strategy/Sample collection of soil and litter (Paper II, III, IV and V) – Three plots were sampled in each major vegetation type present in each locality (3–4 depending on the locality). We sampled soils by adapting the approach of Tedersoo *et al.* (2014) in order to minimize information loss while keeping comparability between the present and other studies. First, 20 trees were randomly selected within 28 m of the Malaise traps (Fig. 2). The collectors used gloves (changed between each sampled tree) and masks all the time to reduce the risk of sample contamination (Fig. 8A). We collected 40 litter and soil cores taken in two opposite directions from those trees and subsequently pooled soil and litter samples to obtain one soil and one litter sample for each plot. The litter consisted of all organic material above the mineral soil and varies from 0 to 50 cm of thickness (Fig. 8B). We collected litter with gloves that were changed between sampling of each tree. Thereafter, we collected soil in the same places, with the soil samples taken from the top 5 cm of the mineral soil using a metal probe with 2.5 cm of diameter (Fig. 8C). The probe was sterilized with fire after collecting soil from both sides of each tree to reduce the risk of cross-contamination among samples. The samples were stored in a plastic bag with the same weight of sterilized white silica gel, 1 – 4 mm thick, pre-treated for two minutes of microwave heating (800 W) and 15 min of UV light. All plots were tagged with GPS coordinates. All dry samples were processed at the University of Gothenburg, Sweden.

Sampling strategy/Sample collection insects (Papers IV and V) – Arthropods, mainly flying insects, were collected with Slam traps (Brazil) or Malaise traps (South Africa and Sweden). Both are tent-like traps made of fine mesh-netting, widely used in entomological studies aimed to capturing strong-flying insects (*e.g.* wasps, mosquitos, and

butterflies) that typically fly upwards after hitting a fine-scale net. Those insects are ultimately trapped in a bottle filled with ethanol at 96% concentration. These two types of traps differ mainly in shape, with Slam traps resembling an igloo (dome-shaped, Fig. 8D) and Malaise traps resembling a Canadian tent (higher on one end, Fig. 8E and F). We sampled for one day in Brazil and South Africa and for seven days in Sweden. The difference in sampling time was due to logistic limitations, and although this should affect the empirical diversity estimates, it should not compromise the evaluation of these methodologies.

Physicochemical soil analyses – We determined the physicochemical soil properties of each plot (from three soil subsamples totalling 117 samples). The pH was measured in water (soil:water ratio 1:2.5). The exchangeable concentrations were measured for sodium (Na), potassium (K), and phosphorus (P) using Mehlich-1 extraction (unit mg/dm^3) and for calcium (Ca) and magnesium (Mg) using KCl (1 mol/L) extraction (unit $cmolc/dm^3$). The sum of all exchangeable bases (SB, which comprises K^+, Ca^{+2}, Mg^{+2}, and Na^+; unit $cmolc/dm^3$) was then calculated. We also estimated exchangeable aluminium (Al and H+Al; unit $cmolc/dm^3$) extracted with calcium acetate (0.5 mol/L at pH 7.0), aluminium saturation index (m; unit %), and Base Saturation Index (V; unit %). The effective cation exchange capacity (t) as well as the cation exchange capacity (T) were measured at pH 7 0 (unit $cmolc/dm^3$). The organic matter (M.O) was quantified (unit g/kg), and the C (organic carbon) was quantified using organic matter (M.O) = C (organic carbon) x 1,724 - Walkley-Black (unit g/kg). Soil texture was characterized as the percentage of fine (0.05 – 0.2 mm), coarse (0.2 – 2 mm), and total sand (0.05 – 2 mm) as well as the silt (0.002 – 0.05 mm) and clay (< 0.002 mm) fraction of the soil weight. We did not quantify nitrogen levels due to the highly volatile nature of nitrogen; its concentration changes quickly during sample storage due to the activity of soil bacteria and other microbes, and freezing the samples in our remote sampling localities was not feasible. All analyses were commissioned from EMBRAPA Ocidental (Brazil) and

followed the protocol described in Donagema *et al.* (2011). Afterwards, we used the mean of the three soil samples from the same plot to obtain a representative value for the measurement of each variable for each plot.
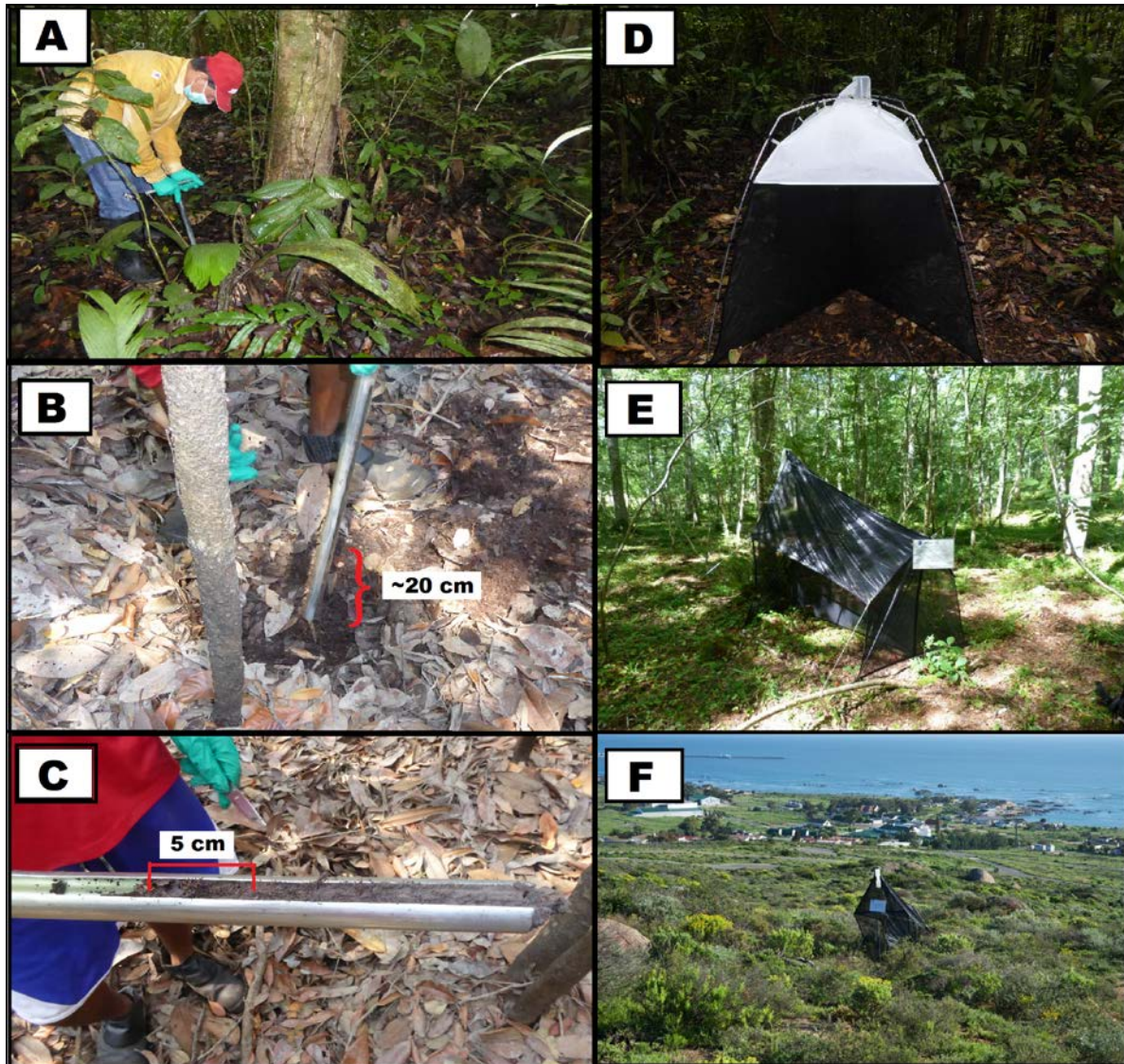
Figure 8. Photos from field sampling. A) Soil collection in a terra-firme in Bejamin Constant; B) Litter in an igapó in Cuieras Reserve. In this plot the litter depth was around 20 cm; C) Soil collection with the probe, we used the top five centimetres of soil; D) Slam trap used in Brazil; E) Malaise trap in a forest fragment in Sweden (photo credit: Karl Mauritsson); F) the Malaise trap used in the coastal area of South Africa (photo credit: Tobias Anderman).

DNA extraction of soil and litter – Ten grams (dry weight) of soil samples and 15 ml of the litter samples (corresponding to 3 - 10 grams of dry weight litter, depending on texture and composition of each samples) were processed for total DNA extraction using the PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories), according to the manufacturer's instructions. Each DNA sample was then concentrated and washed, following the manufacturer's instructions. We added 0.2 ml of 5M NaCl and inverted 3-5 times to mix. We then added 10.4 ml of 100% cold ethanol and inverted 3-5 times to mix. We centrifuged at 2500 x g for 30 minutes at room temperature, decanted all liquid, and washed the DNA pellet with 70% cold ethanol. Residual ethanol was removed at ambient air temperature overnight, and the precipitated DNA was re-suspended in 1 ml sterile 10mM Tris buffer. The concentration of DNA was checked using a Nanodrop 2000c UV-Vis Spectrophotometer (Thermo Scientific) before and after the concentration and washing steps.

DNA extraction from the insect samples – We first tested the efficiency of five DNA extraction protocols on five insect samples obtained from Sweden (Paper IV). Since we found no significant differences among the protocols we decided to perform all subsequent analyses using a non-destructive protocol, namely the one described in Aljanabi and Martinez (1997), for insect sample extractions. The samples were immersed in 15 ml of salt buffer (0. 4 M NaCl, 10 mM Tris-HCl pH = 8, and 2mM EDTA pH = 8) using a vortex mixer for 1 min. Then 1. 5 ml of 20% SDS and 20 µl of 20mg/ml Proteinase K were added. Whenever the insects were not completely covered by the buffer, we added additional buffer with the same proportion of reagents until all insects were covered. The samples were incubated at 60°C overnight. After that, we transferred 15 ml of clear lysis solution into a new tube, and the insects were transferred into 99% ethanol for preservation. Then, 11.25 ml of 6 M NaCl was added to the lysis solution, and the samples were vortexed for 30 s at maximum speed. The samples were then centrifuged at 10,000 g for 30 min, and 25 ml of the supernatant were

transferred to a new tube. An equal volume of isopropanol was added. The samples were incubated at -20°C for 1h and then centrifuged at 10,000 g for 20 min at 4°C. Pellets were washed with 2 ml of 70% ethanol, dried in room temperature, and re-suspended in 300 µl of sterile dH2O.

PCR Amplification – Ribosomal small subunit (SSU) 18S rRNA soil and litter: We targeted the V7 region of the 18S rRNA gene using the forward and reverse primers (5'-TTTGTCTGSTTAATTSCG-3') and (5'-TCACAGACCTGTTATTGC-3') designed by Guardiola *et al.* (2015) to yield 100-110 base long fragments. Amplification was performed in a total volume of 25 µl and consisted of: 0.25 µl of AmpliTaq1 Gold DNA polymerase, 5U/µl, 2.5 µl Pfu polymerase buffer 10x, 0.5 µl dNTP (final concentration of each dNTP 200µmol; all above mentioned reagents are from Promega®, Sweden), 0.25 µl of 50 mol of forward and reverse primers, 20.25 µl of nuclease free water, and 1 µl of DNA template. The PCR started with an initial denaturation step of 2 min at 95°C and then 30 cycles of denaturation at 95°C for 1 min, hybridization at 50°C for 45 s, and elongation at 72°C for 1 min, followed by a final elongation at 72°C for 10 min and finishing at 4°C. Each sample was amplified three times and pooled to reduce biases of amplification efficiency variation on different species and stochastic effects of amplification (Carew *et al.* 2013, Edgar 2013, Piñol *et al.* 2015). The quality of the amplification was checked in UV light using GelRed[TM] stain (1%; Biotium, USA) on a 2% agarose gel. All samples were purified using the QIAquick® PCR purification kit. Dual PCR amplifications were performed for Illumina MiSeq sequencing (Illumina, USA), using fusion primers as described in Bourlat *et al.* (2016). For indexing, we used the Nextera XT DNA index kit (Illumina, USA) according to the manufacturer's instructions. We checked the quality of the PCR products on a 2% agarose gel. We then made a size selection using magnetic beads and a magnetic stand, adopting the ratio 0.9:1 for beads/PCR product. We checked the DNA concentration in a Qubit 30® fluorimeter (Invitrogen, Sweden). We

assessed the quality and size selection of the PCR products with a 2200 Agilent 2200 TapeStation® (Agilent, USA). We normalized and pooled the PCR product (with the same concentration) following the Illumina protocol. The samples were sequenced at SciLifeLab (Stockholm, Sweden) using an Illumina MiSeq 2x250 machine. **18S insects:** We used the same primers as described above, however the amplification and sequencing was done by Macrogen (Republic of Korea) following standard protocols using the Illumina MiSeq 2x250 platform.

Ribosomal small subunit (SSU) 16S rRNA: We targeted the V3-V4 region (~460 bases) of the 16S rRNA gene using the forward primer (5'- CCTACGGGNGGCWGCAG - 3') and reverse primer (5'- GACTACHVGGGTATCTAATCC - 3') from Klindworth *et al.* (2013). **Cytochrome c oxidase subunit mitochondrial gene (COI):** We amplified a COI region of ~313 bases using an internal forward primer (5' - GGWACWGGWTGAACWGTWTAYCCYCC - 3', Leray *et al.* 2013) and the CO1 degenerate reverse primer (5' – TAAACTTCAGGGTGACCAAARAAYCA - 3', Folmer *et al.* 1994).

Sequence analyses and taxonomic assessment – We used the USEARCH/UPARSE v9. 0. 2132 Illumina paired reads pipeline (Edgar 2013) to filter sequence reads for quality, de-replicate and sort reads by abundance, infer OTUs, and remove singletons OTUs. We filtered the sequences to discard chimeras, and we clustered sequences into OTUs at a minimum similarity of 97% using a "greedy" algorithm that performs chimera filtering and OTU clustering simultaneously (Edgar 2013). We used SILVAngs 1. 3 (Quast *et al.* 2012) for assessment of the taxonomic composition of the OTUs, using a representative sequence from each OTU as query sequence. For the 16S and 18S data, we used the SINA v1. 2. 10 reference data for ARB SVN (revision 21008, Pruesse *et al.* 2012) for both markers. For COI we used

the sequences available in GenBank (Benson *et al.* 2005) as reference and blast with the "blastn" tool.

Results and discussion

In general, our results show that the spatial diversity patterns of macro-organisms observed in Amazonia are not the same for all organism groups. The incongruence between our environmental metabarcoding data and macro-organismal diversity patterns highlight the importance of complementary studies to arrive at general diversity conclusions. This is extremely important from the vantage point of conservation and to understand biotic-abiotic interactions that maintain ecosystem functionality. Below I discuss the main results of each paper of this thesis.

Paper I - My review of the Environmental Impact Assessment (EIA) of three large infrastructure projects showed that the current EIA system in Brazil is very general and imprecise in its requirements. Several reasons contribute to making biodiversity assessments based on EIAs inadequate, notably the limited spatial and temporal scopes. Due to the lack of sufficient biological characterization of multiple distinct environments, EIAs often fail to deliver a proper description of the species composition as well as a characterization of the abiotic environment, such as the quality of the soil and water. More should be done not only to assess and document biodiversity but also to provide suitable baselines for comparison between largely pristine and more altered environments. Similarly, more efficient mitigation strategies must be proposed. Many aspects should be considered to improve the quality of EIAs in Brazil: improving the training of researchers and practitioners; making the formal requirements of EIAs more specific and biologically sound; providing an improved definition of the terms of reference (the document that sets the minimal assessment necessary in each project); and requiring a more thorough inventory of the species in the areas expected to be directly and indirectly affected by the infrastructure projects. With these improvements, future studies are likely to become more analytical, less descriptive, and of enhanced usefulness.

Paper II - In this study I showed that the currently accepted diversity patterns in Amazonia do not hold true for diversity in general, suggesting a decoupling of biodiversity patterns between macro- and micro-organisms. It is urgent to characterize the diversity patterns also for micro-organism communities, which comprise the vast majority of the global biodiversity. After centuries of studies, our understanding of biodiversity patterns is still heavily biased towards a handful of charismatic organism groups. The assessment of tropical diversity is enhanced by the use of 16S and 18S markers, as explored alongside traditional inventories by taxonomists. However, many constraints and complications remain, notably the compromise between taxonomic coverage and taxonomic resolution, and the lack of well-populated reference sequence databases. But even with incomplete reference databases, the use of OTUs gives a reasonable community composition proxy. We found that the biodiversity patterns are characterized more strongly by habitats types than by the west to east gradient. The community composition is more similar between flooded forests (igapós and várzeas) than between other environmental types. The campinas and terra-firme also have a high number of shared OTUs, highlighting that the inundation stress could be a filter for biotic communities.

Paper III – In this study, I found that soil and litter richness, the effective number of OTUs (abundance-based analysis), and the community composition are related in prokaryotes but not in eukaryotes. These differences could be observed at the OTU level but not at the phylum and order levels. I also found that the dominant bacterial groups in our dataset differed from those reported as globally dominant in a recent study (Delgado-Baquerizo *et al.* 2018). Additionally, the physicochemical soil variables, notably organic carbon and pH, could predict, to a certain extent, soil and litter diversity across Amazonia. pH had a positive correlation with OTU diversity, which was expected given the generally low pH in Amazonian soils, varying between 3.65 to 5.14. I found a negative correlation between soil

organic carbon content and prokaryotic and eukaryotic OTU diversity. This result was unexpected since diversity is usually related to biomass in soils and litter. However, soil/litter diversity is also related to decomposition rates. With rapid decomposition, the bulk of the biomass may be bound to aboveground stocks. Soils are crucial for carbon cycling in terrestrial ecosystems, and our results suggest that a better understanding of the relationship between diversity (above and belowground) and carbon cycling may help modelling carbon deposition and biodiversity patterns. Furthermore, I found a significant effect of soil organic carbon content on community composition.

Paper IV – In this study I found a significant difference between the diversity of prokaryote and eukaryote organisms registered in insect and soil samples. This stresses the advantages of including insect samples to complement biodiversity assessments in soil metabarcoding studies. To achieve this goal we presented a combined protocol using Malaise traps and soil samples. I tested the protocol in a wide range of environments, from hot and humid rainforests, to species-rich Mediterranean meadows and natural and anthropogenic habitats in the temporal zone. All localities showed a strong difference in community composition between the two sorts of samples (insects and soils). I suggested the use of non-destructive DNA extraction in insect samples to preserve specimens for, *e.g.*, subsequent taxonomic studies, which are fundamental to the discovery and description of new species. Adopting a massive and standardized sampling scheme would allow fast and cost-effective estimations of global biodiversity and complement traditional biodiversity inventories.

Paper V – In this study, I found no significant relationship between OTU and taxonomic richness. Furthermore, I did not recover the same spatial patterns of diversity between macro- and micro-organisms for any taxonomic group. This suggests that at small spatial scales, the biological characteristics of each taxonomic group and peculiarities of each environment may be more important than the general diversity patterns for explaining

differences in richness distribution. In particular, I observed higher-than-expected tree richness in terra-firme environmental types from the Cuieras locality. Furthermore, I found a higher-than-expected biological richness in metabarcoding data from campinas, which are assumed to have lower levels of macro-organismic diversity than the other environmental types (ter Steege & Hammond 2001, Fine *et al.* 2010, Borges & Almeida 2001, Draper *et al.* 2018). Previous studies already reported on the importance of campinas for beta-diversity (*e.g.* Draper *et al.* 2018), and my data suggest that these environments could be potential hotpots of microbial diversity.

Conclusions and outlook

The papers in this thesis identified some of the major challenges of biodiversity assessments and presented prospects and tools to describe biodiversity beyond the taxonomic impediment. The identified issues include difficulties in describing some organism groups; comparing biodiversity among localities with many undescribed species; and scarcely populated reference sequence databases. These studies focused on inconspicuous organisms, many of which are difficult to examine using traditional methods, and on the assessment and comparison of biodiversity in megadiverse, yet poorly studied, Amazonian ecosystems.

This thesis contributes to mapping Amazonian diversity in the context not only of charismatic macroscopic organisms, but of the majority of terrestrial organisms. With metabarcoding data, it was possible to test if the general diversity patterns for macro-organisms could be generalized to more extensive biodiversity data, as collected in this thesis. I hope that this thesis will advance the biodiversity assessment debate in terms of Amazonian diversity patterns and drivers. With the development and implementation of the molecular-based methods presented here, it was possible to quantify biodiversity without the need to spend decades to obtain a sizable batch of taxonomic information. This is of extreme importance considering the current biodiversity crisis and the socio-political instability in the countries that comprise Amazonia. We need to know and understand biodiversity to use our knowledge in areas such as ecology, conservation biology, landscape planning, and evolution.

Is important to highlight that, in general, molecular data derived from "environmental sequencing" should be seen as complementary to, rather than as competing with, traditional taxonomic studies. Indeed, a confluence of both lines of evidence is highly warranted, as it will be necessary to overcome their respective shortcomings. Currently, parts of the taxonomic community take little (positive) note of environmental sequencing studies and *vice*

versa (Ryberg & Nilsson 2018). It is particularly important that results of environmental sampling are made available and communicated to taxonomists, because this could encourage work with lineages untreated in a taxonomic framework. This interaction is likely to create "taxonomy feedback loops" which would accelerate species and lineage discovery, and potentially add prominent branches to the tree of life (Nilsson *et al.* 2016, Tedersoo *et al.* 2017). Thereby such an integrative approach will also improve the efficiency of conservation strategies, as more effort could be directed to "true" comprehensive hot-spots of biodiversity, including taxonomic richness and unique lineages in groups "hidden" to approaches solely based on either taxonomy or environmental DNA. Conversely, taxonomists are in a position to increase the resolution in ecological studies by providing ample, richly annotated DNA sequence data of legacy species as well as all newly described species for use in molecular characterization of biological communities.

Paper contributions

Paper I – *Environmental impact assessment in Brazilian Amazonia: Challenges and prospects to assess biodiversity.* Camila Duarte Ritter (CDR) conceived of this study together with the co-authors and wrote the manuscript with contributions from all co-authors.

Paper II – *Locality or habitat? Exploring predictors of biodiversity in Amazonia.* CDR conceived of this study together with the co-authors, carried out all the field and lab work, performed all the analyses together with co-authors, and led the writing with contributions from all co-authors.

Paper III – *High-throughput metabarcoding reveals the effect of physicochemical soil properties on soil and litter biodiversity and community turnover across Amazonia.* CDR conceived of this study together with the co-authors, carried out all the field and lab work, performed all the analyses together with co-authors, and led the writing with contributions from all co-authors.

Paper IV – *Biodiversity assessment in the 21st century: An integrative protocol for assessing eukaryotic and prokaryotic diversity from soil and insect traps using high-throughput DNA metabarcoding in various environments*. CDR conceived of this study together with the co-authors, carried out all the field and lab work, performed all the analyses together with co-authors, and led the writing with contributions from all co-authors.

Paper V – *DNA metabarcoding shows distinct patterns of diversity for Amazonian macro- and micro organisms.* CDR conceived of this study together with the co-authors, carried out all the field and lab work, performed all the analyses together with co-authors, and led the writing with contributions from all co-authors.

References

Abarenkov, K., Adams, R. I., Laszlo, I., Agan, A., Ambrosio, E., Antonelli, A., Bahram, M., Bengtsson-Palme, J., Bok, G., Cangren, P., Coimbra, V., Coleine, C., Gustafsson, C., He, J., Hofmann, T., Kristiansson, R., Larsson, E., Larsson, T., Liu, Y., Martinsson, S., Meyer, W., Panova, M., Pombubpa, N., Ritter, C., Ryberg, M., Svantesson, S., Scharn, R., Svensson, O., Töpel, M., Unterseher, M., Visagie, C., Wurzbacher, C., Taylor, A. F. S., Kõljalg, U., Schriml, L. and Nilsson, R. H. 2016. Annotating public fungal ITS sequences from the built environment according to the MIxS-Built Environment standard–a report from a May 23-24, 2016 workshop (Gothenburg, Sweden). *MycoKeys,* 16, p. 1-15.

Agência Nacional de Energia Elétrica – ANEEL. Potencial hidrelétrico brasileiro. Em: Energia hidráulica. Brasília: DF, s. d. Available from: http://www. aneel. gov. br/aplicacoes/atlas/energia_hidraulica/4_3. htm. Accessed in: Dezember, 2016.

Aljanabi, S. M. and Martinez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research, 25*(22), pp. 4692-4693.

Amann, R. I., Ludwig, W. and Schleifer, K. H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, *59*(1), pp. 143-169.

Anderson, A. B. 1981. White-sand vegetation of Brazilian Amazonia. *Biotropica*, pp. 199-210.

Antonelli, A., Quijada-Mascareñas, A., Crawford, A. J., Bates, J. M., Velazco, P. M. and Wüster, W. 2010. Molecular studies and phylogeography of Amazonian tetrapods and their relation to geological and climatic models. *Amazonia, landscape and species evolution: a look into the past*, pp. 386-404.

Ávila-Pires, T. C. S., Hoogmoed, M. S. and Vitt, L. J. 2007. Herpetofauna da Amazônia. *Herpetologia no Brasil II. Belo Horizonte: Sociedade Brasileira de Herpetologia*, pp. 13-43.

Baird, D. J. and Hajibabaei, M. 2012. Biomonitoring 2. 0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, 21(8), pp. 2039-2044.

Balmford, A. and Whitten, T. 2003. Who should pay for tropical conservation, and how could the costs be met?. *Oryx*, *37*(02), pp. 238-250.

Barbault, R. 2011. 2010: A new beginning for biodiversity? *Comptes rendus biologies* (334),pp. 483-488.

Barnes, C. J., Maldonado, C., Frøslev, T. G., Antonelli, A. and Rønsted, N. 2016. Unexpectedly high beta-diversity of root-associated fungal communities in the Bolivian Andes. *Frontiers in microbiology*, *7*, p. 1377.

Bates, J. M. 2001. Avian diversification in Amazonia: evidence for historical complexity and a vicariance model for a basic diversification pattern. *Diversidade biológica e cultural da Amazônia*, pp. 119-137.

Bebber, D. P., Marriott, F. H., Gaston, K. J., Harris, S. A. and Scotland, R. W. 2007. Predicting unknown species numbers using discovery curves. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*(1618), pp. 1651-1658.

Behling, H. and da Costa, M. L. 2000. Holocene environmental changes from the Rio Curuá record in the Caxiuanã region, eastern Amazon Basin. *Quaternary Research*, *53*(3), pp. 369-377.

Beng, K. C., Tomlinson, K. W., Shen, X. H., Surget-Groba, Y., Hughes, A. C., Corlett, R. T. and Slik, J. F. 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific reports*, *6*.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Wheeler, D. L. 2005. GenBank. *Nucleic acids research*, *33*(S1), pp. D34-D38.

Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R. and Thomas, W. K. 2012. Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in ecology & evolution*, *27*(4), pp. 233-243.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. and Abebe, E. 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical transactions of the royal society B: biological sciences*, 360(1462), pp. 1935-1943.

Borges, S. H. and de Almeida, R. A. M. 2013. Birds of the Jaú National Park and adjacent areas, Brazilian Amazon: new species records with reanalysis of a previous checklist. *Revista Brasileira de Ornitologia-Brazilian Journal of Ornithology*, *19*(44), p. 27.

Borges, S. H., Cohn-Haft, M., Carvalhaes, A. P., Henriques, L. M., Pacheco, J. F. and Whittaker, A. 2001. Birds of Jaú National Park, Brazilian Amazon: species check-list, biogeography and conservation. *Ornitologia Neotropical*, *12*(2), pp. 109-140.

Bourlat, S. J., Haenel, Q., Finnman, J., & Leray, M. (2016). Preparation of amplicon libraries for metabarcoding of marine eukaryotes using Illumina MiSeq: the dual-PCR method. *Marine Genomics: Methods and Protocols, (1452),* pp. 197-207.

Buckley, D. H. and Schmidt, T. M. 2001. Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil. *FEMS microbiology ecology*, *35*(1), pp. 105-112.

Campbell, G., Kuehl, H., Diarrassouba, A., N'Goran, P. K. and Boesch, C. 2011. Long-term research sites as refugia for threatened and over-harvested species. *Biology letters*, 7(5), pp. 723-726.

Carew, M. E., Pettigrove, V. J., Metzeling, L. and Hoffmann, A. A. 2013. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology*, *10*(1), pp. 10-45.

Chan, K. Y., Van Zwieten, L., Meszaros, I., Downie, A. and Joseph, S. 2008. Agronomic values of greenwaste biochar as a soil amendment. *Soil research*, 45(8), pp. 629-634.

Chapman, A. D. 2009. Numbers of living species in Australia and the world. Report for the *Australian Biological Resources Study Canberra*, Australia, pp. 1-78. Available from: https://www. environment. gov. au/system/files/pages/2ee3f4a1-f130-465b-9c7a-79373680a067/files/nlsaw-2nd-complete. pdf (accessed 20/07/2018).

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X. and Luo, K. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PloS one*, *5*(1), p. e8613.

Clarke, L. J., Soubrier, J., Weyrich, L. S. and Cooper, A. 2014. Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular ecology resources*, *14*(6)*,* pp. 1160-1170.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C. and Bik, H. M. 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, *7*(9), pp. 1008-1018.

Davies, N., Meyer, C., Gilbert, J. A., Amaral-Zettler, L., Deck, J., Bicak, M., Rocca-Serra, P., Assunta-Sansone, S., Willis, K. and Field, D. 2012. A call for an international network of genomic observatories (GOs). *GigaScience*, 1(1), p. 5.

Deacon, H. J., Jury, M. R., Ellis, F. 1992. Selective regime and time. In: Cowling, R. M. (Ed.), *The Ecology of Fynbos: Nutrients, Fire and Diversity.* Oxford University Press, Cape Town, pp. 6–22.

Del Hoyo, J. A. Elliott, A. Sargatal, J. and Christie, D. A. (eds.). 2013. Handbook of the Birds of the World, Special Volume, New Species and Global Index". *Lynx Edicions, Barcelona*, pp. 812.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K. and Fierer, N. 2018. A global atlas of the dominant bacteria found in soil. *Science*, *359*(6373), pp. 320-325.

Dominati, E., Patterson, M. and Mackay, A. 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. *Ecological economics*, *69*(9)*,* pp. 1858-1868.

Donagema, G. K., De Campos, D. B., Calderano, S. B., Teixeira, W. G. and Viana, J. M. 2011. Manual de métodos de análise de solo. *Embrapa Solos-Documentos (INFOTECA-E).*

Draper, F. C., Honorio Coronado, E. N., Roucoux, K. H., Lawson, I. T., A. Pitman, N. C., A. Fine, P. V., Phillips, O. L., Torres Montenegro, L. A., Valderrama Sandoval, E., Mesones, I. and García‑Villacorta, R. 2018. Peatland forests are the least diverse tree communities documented in Amazonia, but contribute to high regional beta-diversity. *Ecography*.

Edgar, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), pp. 996-998.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J. N., Coissac, E., Boyer, F. and Leese, F. 2016. Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ*, 4, p. e1966.

Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M. and Thiers, B. M. 2016. Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity (No. e2615v1). *PeerJ Preprints.*

Falesi, I. C. 1984. Estado atual de conhecimento dos solos da Amazônia Brasileira. In: *Simpósio do Trópico Umido, Belém, 1984. Embrapa- CPATU, Documentos, 36*: pp. 168-191.

Faoro, H., Alves, A. C., Souza, E. M., Rigo, L. U., Cruz, L. M., Al-Janabi, S. M., Monteiro, R. A., Baura, V. A. and Pedrosa, F. O. 2010. Influence of soil properties on the diversity of bacteria in the Southern Brazilian Atlantic Forest. *Applied and Environmental Microbiology*, *76*(14), pp. 4744-4749.

Ferreira, L. V. and Prance, G. T. 1999. Ecosystem recovery in terra firme forests after cutting and burning: a comparison on species richness, floristic composition and forest structure in the Jaú National Park, Amazonia. *Botanical Journal of the Linnean Society*, *130*(2), pp. 97-110.

Fine, P. A., Daly, D. C. and Cameron, K. M. 2005. The contribution of edaphic heterogeneiyt to the evolution and diversity of burseracear trees in the western Amazon. *Evolution*, *59*(7), pp. 1464-1478.

Fine, P. V., García-Villacorta, R., Pitman, N. C., Mesones, I. and Kembel, S. W. 2010. A floristic study of the white-sand forests of Peru. *Annals of the Missouri Botanical Garden*, pp. 283-305.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, pp. 294-299.

Fundação Vitória Amazônica - FVA. 1998. Plano de manejo do Parque Nacional do Jaú. *Fundação Vitória Amazônica/Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Manaus.*

Giam, X., Scheffers, B. R., Sodhi, N. S., Wilcove, D. S., Ceballos, G. and Ehrlich, P. R. 2012. Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity. *Proceedings of the royal society of London B: biological sciences,* 279(1726), pp. 67-76.

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., Hallwachs, W. and Hajibabaei, M. 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, 111(22), pp. 8007-8012.

Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, O. S. and Turon, X. 2015. Deep-sea, deep-sequencing: metabarcoding extracellular DNA from sediments of marine canyons. *PloS one*, *10*(10), p. e0139633.

Hamilton, A. J., Basset, Y., Benke, K. K., Grimbacher, P. S., Miller, S. E., Novotný, V., Samuelson, G. A., Stork, N. E., Weiblen, G. D. and Yen, J. D. 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. *The american naturalist*, 176(1), pp. 90-95.

Hartmann, M., Howes, C. G., Abarenkov, K., Mohn, W. W. and Nilsson, R. H. 2010. V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16 S/18 S) ribosomal RNA gene sequences. *Journal of Microbiological Methods*, *83*(2), pp. 250-253.

Hawksworth, D. L. 2001. The magnitude of fungal diversity: the 1. 5 million species estimate revisited. *Mycological research*, 105(12), pp. 1422-1432.

Hebert, P. D., Cywinska, A. and Ball, S. L. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1512), pp. 313-321.

Higgins, M. A., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O. L., Vásquez, R. and Räsänen, M. 2011. Geological control of floristic composition in Amazonian forests. *Journal of Biogeography*, *38*(11), pp. 2136-2149.

Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., Sanmartín, I., Sanchez-Meseguer, A., Anderson, C. L., Figueiredo, J. P. and Jaramillo, C. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science,* 330(6006), pp. 927-931.

Instituto Brasileiro de Biogeografia e Estatística – IBGE. 2012. Available from: https://mapas. ibg*e.g*ov. br/

Instituto Nacional de Pesquisas Espaciais – INPE/PROBES. 2016. Catálogo de imagens de satélite. Available from: http://www. dpi. inpe. br/prodesdigital/prodes. php

Instituto Socio-Ambiental – ISA. 2016. Unidades de conservação. Available from: https://uc. socioambiental. org/

Jay, S., Jones, C., Slinn, P. and Wood, C. 2007. Environmental impact assessment: Retrospect and prospect. *Environmental impact assessment review*, 27(4), pp. 287-300.

Jenkins, C. N., Alves, M. A. S., Uezu, A. and Vale, M. M. 2015. Patterns of vertebrate diversity and protection in Brazil. *PloS one*, *10*(12), p. e0145064.

Junk, W. J., Bayley, P. B., & Sparks, R. E. 1989. The flood pulse concept in river-floodplain systems. *Canadian Special Publication of Fisheries and Aquatic Sciences*, *106*(1), pp. 110-127.

Junk, W. J., Piedade, M. T. F., Schöngart, J., Cohn-Haft, M., Adeney, J. M., & Wittmann, F. 2011. A classification of major naturally-occurring Amazonian lowland wetlands. *Wetlands*, *31*(4), pp. 623-640.

Kang, S., Mansfield, M. A., Park, B., Geiser, D. M., Ivors, K. L., Coffey, M. D., Grünwald, N. J., Martin, F. N., Lévesque, C. A. and Blair, J. E. 2010. The promise and pitfalls of sequence-based identification of plant-pathogenic fungi and oomycetes. *Phytopathology,* 100(8), pp. 732-737.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. O. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic acids research*, *41*(1), pp. e1-e1.

Kuramae, E. E., Yergeau, E., Wong, L. C., Pijl, A. S., Van Veen, J. A. and Kowalchuk, G. A. 2012. Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiology Ecology*, *79*(1), pp. 12-24.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M. and Douglas, B. 2013. Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, 22(21), pp. 5271-5277.

Larsen, B. B., Miller, E. C., Rhodes, M. K. and Wiens, J. J. 2017. Inordinate fondness multiplied and redistributed: the number of species on earth and the new pie of life. *The Quarterly Review of Biology*, *92*(3), pp. 229-265.

Laurance, S. G., Laurance, W. F., Andrade, A., Fearnside, P. M., Harms, K. E., Vicentini, A. and Luizão, R. C. 2010. Influence of soils and topography on Amazonian tree diversity: a landscape-scale study. *Journal of Vegetation Science*, *21*(1), pp. 96-106.

Leite, R. N. and Rogers, D. S. 2013. Revisiting Amazonian phylogeography: insights into diversification hypotheses and novel perspectives. *Organisms diversity & evolution*, 13(4), pp. 639-664.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T. and Machida, R. J. 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in zoology*, *10*(1), p. 34.

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J. and Kauserud, H. 2013. Fungal community

analysis by high-throughput sequencing of amplified markers–a user's guide. *New Phytologist*, *199*(1), pp. 288-299.

Liu, S., Li, Y., Lu, J., Su, X., Tang, M., Zhang, R., Zhou, L., Zhou, C., Yang, Q., Ji, Y. and Yu, D. W. 2013. SOAPBarcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in ecology and evolution*, 4(12), pp. 1142-1150.

Locey, K. J. and Lennon, J. T. 2016. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, *113*(21), pp. 5970-5975.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. and Pallen, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature biotechnology*, 30(5), pp. 434-439.

Lovejoy, N. R., Albert, J. S. and Crampton, W. G. 2006. Miocene marine incursions and marine/freshwater transitions: Evidence from Neotropical fishes. *Journal of South American Earth Sciences*, *21*(1-2), pp. 5-13.

Luizão, R. C., Luizão, F. J. and Proctor, J. 2007. Fine root growth and nutrient release in decomposing leaf litter in three contrasting vegetation types in central Amazonia. *Plant Ecology*, *192*(2), pp. 225-236.

Magurran, A. E. 2004. Measuring biodiversity. *John Wiley & Sons.*

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S. and Siemensmeyer, T. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nature ecology & evolution*, *1*(4), p. 0091.

Malhi, Y., Roberts, J. T., Betts, R. A., Killeen, T. J., Li, W. and Nobre, C. A. 2008. Climate change, deforestation, and the fate of the Amazon. *Science,* 319(5860), pp. 169-172.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. and Worm, B. 2011. How many species are there on Earth and in the ocean?. *PLoS biology*, 9(8), p. e1001127.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. and Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), p. 853.

Navarrete, A. A., Kuramae, E. E., De Hollander, M., Pijl, A. S., Van Veen, J. A. and Tsai, S. M. 2013. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. *FEMS Microbiology Ecology*, *83*(3), pp. 607-621

Nilsson, R. H., Tedersoo, L., Abarenkov, K., Ryberg, M., Kristiansson, E., Hartmann, M., Schoch, C. L., Nylander, J. A., Bergsten, J., Porter, T. M. and Jumpponen, A. 2012. Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys*, 4, p. 37.

Nilsson, R. H., Wurzbacher, C., Bahram, M., Coimbra, V. R., Larsson, E., Tedersoo, L., Eriksson, J., Ritter, C. D., Svantesson, S., Sánchez-García, M. and Ryberg, M. 2016. Top 50 most wanted fungi. *MycoKeys*, (12), pp. 29-40.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C. and Loucks, C.J. 2001. Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*(11), pp.933-938.

Osborne, C. A., Zwart, A. B., Broadhurst, L. M., Young, A. G. and Richardson, A. E. 2011. The influence of sampling strategies and spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiology Ecology*, *78*(1), pp. 70-79.

Peres, C. A. 1997. Primate community structure at twenty western Amazonian flooded and unflooded forests. *Journal of Tropical Ecology, 13*(03), pp. 381-405.

Pitman, N. C., Terborgh, J. W., Silman, M. R., Núñez, V., Neill, D. A., Cerón, C. E., Palacios, W. A. and Aulestia, M. 2001. Dominance and distribution of tree species in upper Amazonian terra firme forests. *Ecology, 82*(8), pp. 2101-2117.

Piñol, J., Mir, G., Gomez-Polo, P. and Agustí, N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular ecology resources*, *15*(4), pp. 819-830.

Prance G. T. 1996. Islands in Amazonia. *Philosophical Transactions of the Royal Society of London. B, 29*(1341), pp. 823-33.

Pruesse, E., Peplies, J. and Glöckner, F. O. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*(14), pp. 1823-1829.

QGIS Development Team. 2012. QGIS Geographic Information System. Open Source Geospatial Foundation Project.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F. O. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), pp. D590-D596.

Ribas, C. C., Aleixo, A., Nogueira, A. C., Miyaki, C. Y. and Cracraft, J. 2011. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proceedings of the royal society of London B: biological sciences*, p. rspb20111120.

Ryberg, M. and Nilsson, R. H. 2018. New light on names and naming of dark taxa. *MycoKeys*, (30), p. 31.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T. and Quince, C. 2015. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic acids research*, p. gku1341.

Scholes, R. J. and Biggs, R. 2005. A biodiversity intactness index. *Nature*,*434*(7029), pp. 45-49.

Schloss, P. D., Girard, R. A., Martin, T., Edwards, J. and Thrash, J. C. 2016. Status of the archaeal and bacterial census: an update. *MBio*, *7*(3), pp. e00201-16.

Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E,. McDonald, A., Lefebvre, P. and Schlesinger, P. 2006. Modelling conservation in the Amazon basin. *Nature,* 440(7083):520-3.

Stahlhut, J. K., Fernández-Triana, J., Adamowicz, S. J., Buck, M., Goulet, H., Hebert, P. D., Huber, J. T., Merilo, M. T., Sheffield, C. S., Woodcock, T. and Smith, M. A. 2013. DNA barcoding reveals diversity of Hymenoptera and the dominance of parasitoids in a sub-arctic environment. *BMC ecology*, 13(1), p. 2.

Stajich J. E., Berbee M. L., Blackwell M., Hibbett D. S., James T. Y., Spatafora J. W., Taylor J. W. 2009. The fungi. *Current biology,* 19(18): R840-R845.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., de Vries, W., de Wit, C. A. and Folke, C. 2015. Planetary boundaries: Guiding human development on a changing planet. *Science*, 347(6223), p. 1259855.

Taylor, D. L., Hollingsworth, T. N., McFarland, J. W., Lennon, N. J., Nusbaum, C. and Ruess, R. W. 2014. A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological monographs,* 84(1), pp. 3-20.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A. L., Nilsson, R. H., Morgado, L. N., Mayor, J., May, T. W., Majuakim, L., Lodge. J., Lee, S. S., Larsson, K., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T. W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., Kesel, A. D., Dang, T., Chen, X., Buegger, F., Brearley, F. Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K. 2014. Global diversity and geography of soil fungi. *Science*, 346(6213), p. 1256688.

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, R. H., Hildebrand, F. and Bork, P. 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*, (10), pp. 1-43.

Tedersoo, L., Bahram, M., Puusepp, R., Nilsson, R. H. and James, T. Y. 2017. Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome*, 5(1), p. 42.

ter Steege, H. and Hammond, D. S. 2001. Character convergence, diversity, and disturbance in tropical rain forest in Guyana. *Ecology*, *82*(11), pp. 3197-3212.

ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., Van Der Hout, P., Daly, D. C., Silveira, M., Phillips, O., Vasquez, R., Van Andel, T. and Duivenvoorden, J. 2003. A spatial model of tree α-diversity and tree density for the Amazon. *Biodiversity & conservation*, 12(11), pp. 2255-2277.

Vartoukian, S. R. 2016. Cultivation strategies for growth of uncultivated bacteria. *Journal of oral biosciences,* 58(4), pp. 143-149.

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., Van Elsas, J. D., Bailey, M. J., Nalin, R. and Philippot, L. 2009. TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature Reviews Microbiology*, 7, 252.

Wearn, O. R., Reuman, D. C., and Ewers, R. E. 2012. Extinction Debt and Windows of Conservation Opportunity in the Brazilian Amazon. *Science,* pp. 228-232.

Webb, S. D. 1995. Biological implications of the middle Miocene Amazon seaway. *Science,* 269(5222), p. 361.

Wheeler, Q. 2014. Are reports of the death of taxonomy an exaggeration?. *New phytologist*, 201(2), pp. 370-371.

Wilson, E. O. and Peter, F. M. (Eds). 1988. Biodiversity. Washington, DC, USA: *National academic press,* 535 pp.

World Wildlife Fund – WWF. 2013. Amazon New Species List 2010-2013. : 1-17. [http://assets. worldwildlife. org/publications/645/files/original/Amazon_new_species_list_2010-2013_ES. pdf].

Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P. and Chen, S., 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PloS one*, *5*(10), p. e13102.

Zanchi, F.B., Waterloo, M.J., Dolman, A.J., Groenendijk, M., Kesselmeier, J., Kruijt, B., Bolson, M.A., Luizão, F.J. and Manzi, A.O. 2011. Influence of drainage status on soil and water chemistry, litter decomposition and soil respiration in central Amazonian forests on sandy soils. *Ambiente e Agua-An Interdisciplinary Journal of Applied Science*, *6*(1), pp. 6-29.

Acknowledgements

After for years I could finish this thesis thanks to several people that are crucial during this time.

First of all, I want to thank my supervisor **Alex Antonelli** for the opportunity to develop this project, his exciting view of science and all good moments.

My supervisor **Henrik R. Nilsson** for all help with analysis, writing, English and to be a fountain of support all the time. I didn't know if I would be able to keep my mental sanity without you.

Urban Olsson for all good moments, talks and possibility to live in a great place.

Daniele Silvestro to be a true friend, share many beers, cigarettes, funny moments and beyond have patience to read my texts and help me all the time I requested.

Soren Faurby for all funny moments and all help, even more in the last month. Thanks a lot!

Christine Bacon for all dinners, talks and be an amazing woman scientist that is an inspiration.

Bengt Oxelman, my examiner, for South Africa trip, discussions and help me to keep the project in track.

Alex Zizka, I don't think I need to say how important you are, thanks to be part of my life and be one of the best friend I could have.

Toby Anderman, thank for all long talks, comprehension, patience and good shared moments, you are one the best person I meet.

All colleagues and formers to share so may moments and make it better: **Sven, Patrik, Harith, Yann, Allison, Ruud, Angela, Maria, Bernard, Claes, Johannes, Triranta, Sommah** and **Andrei.**

Papers